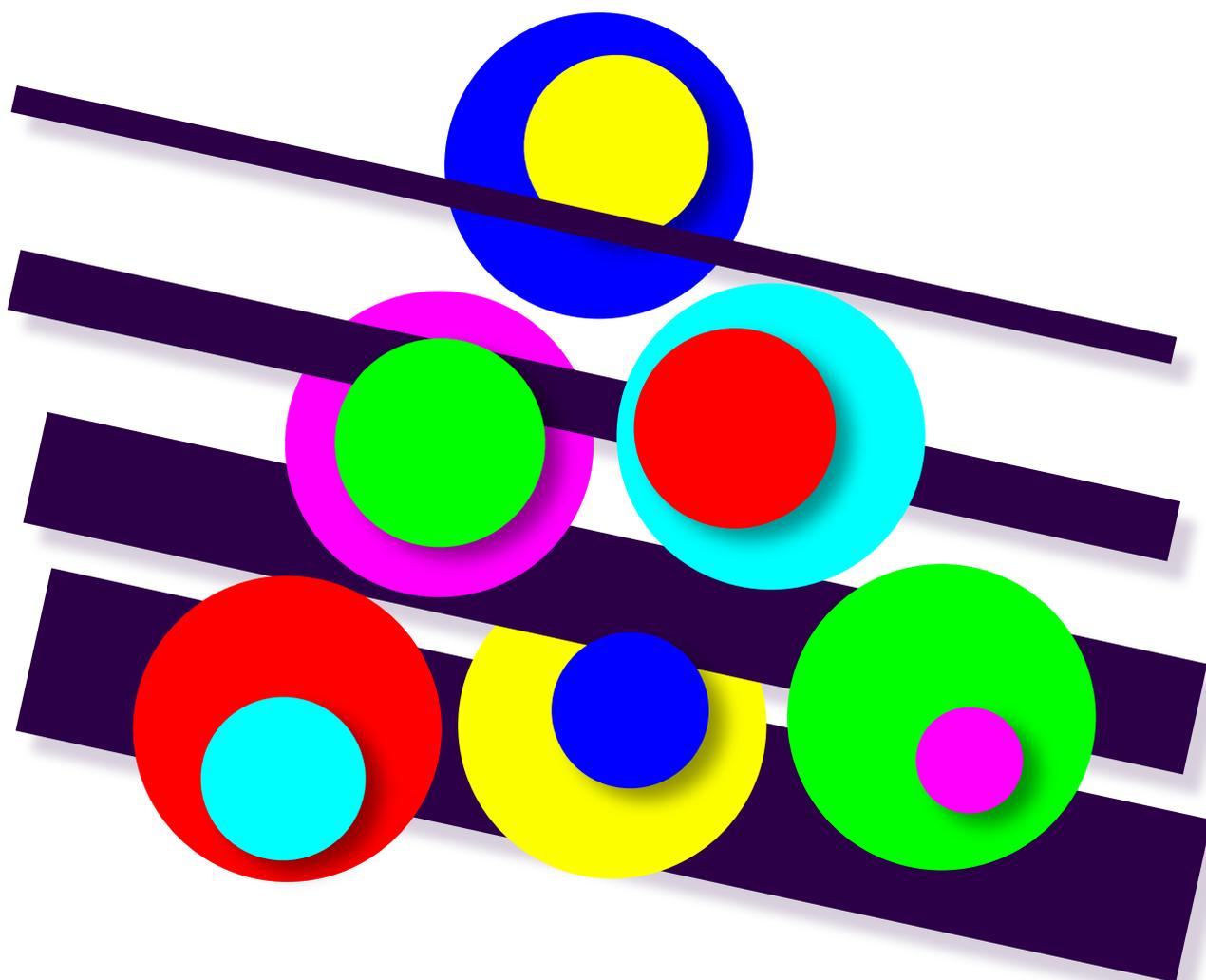


Numero XVII
Anno 2021



Viaggio nella Scienza

Ithaca



Dare i numeri: (ab)usi della Matematica (A)

Ithaca: Viaggio nella Scienza

Una pubblicazione del Dipartimento di Matematica e Fisica “Ennio De Giorgi” dell’Università del Salento.

Registrazione presso il Tribunale di Lecce n. 6 del 30 Aprile 2013.
e-ISSN: 2282-8079

Direttore Responsabile
Luigi Spedicato.

Ideatore
Giampaolo Co’.

Comitato di Redazione
Adriano Barra,
Rocco Chirivì,
Paolo Ciafaloni,
Maria Luisa De Giorgi,
Vincenzo Flaminio,
Luigi Martina,
Giuseppe Maruccio,
Marco Mazzeo,
Francesco Paparella,
Carlo Sempi.

Segreteria di Redazione
Daniela Dell’Anna.

© 2013-2023 Dipartimento di Matematica e Fisica “*Ennio de Giorgi*”.

© 2023 per i singoli articoli dei rispettivi autori.

Il materiale di questa pubblicazione può essere riprodotto nei limiti stabiliti dalla licenza
“*Creative Commons Attribuzione – Condividi allo stesso modo 3.0 Italia*” (CC BY-SA 3.0 IT).

Per il testo della licenza: <http://creativecommons.org/licenses/by-sa/3.0/it/deed.it>

Ithaca: Viaggio nella Scienza

è disponibile sul sito:

<http://ithaca.unisalento.it/>

Scriveteci all’indirizzo:

ithaca@unisalento.it

- 3** In questo numero
- 5** Non sempre 2 è il doppio di 1
Paolo Ciafaloni
- 15** Dinamica e controllo del Covid-19: commenti da due matematici
Bernhelm Booß-Bavnbek, Klaus Krickeberg
- 39** Il trattamento per gruppi
Christian Genest, Christiane Rousseau
- 45** Applicazioni del Formalismo Quantistico alle Scienze Cognitive
Sandro Sozzo
- 63** Big Data, profezie, assicurazione: una prospettiva sociologica
Alberto Cevolini
- 69** Linguaggio nei numeri e numeri nel linguaggio. Linguistica, Matematica e Cryptonalisi
Thomas Christiansen

93 **La Fisica e la Matematica negli Studi di Chopin**
Massimo Blasone

109 **Galileo, la matematica, la natura**

In questo numero

L'uso della matematica per studiare il mondo che ci circonda è l'argomento che lega gli articoli del numero XVII di **Ithaca**. Non si tratta di articoli di indirizzo filosofico o epistemologico sul rapporto tra ciò che osserviamo e il modo di analizzarlo, catalogarlo, sistematizzarlo. Gli articoli presentano esempi di come la Matematica sia utilizzata in differenti campi delle scienze, sia naturali che umane.

Dato il rilevante numero di contributi, abbiamo pensato di presentarli in due parti.

Il primo articolo di questa prima parte del numero XVII di **Ithaca** è una guida alla quantificazione dell'osservazione. Se è vero il luogo comune che "i numeri non mentono", è anche vero che chi li usa può sbagliare ad interpretarli, e questo nella migliore ipotesi, perché può anche mentire citando numeri corretti. L'articolo di Paolo Ciafaloni considera l'ipotesi migliore, e mette in evidenza alcune difficoltà legate al processo che associa numeri ad una osservazione, procedura niente affatto banale tantomeno priva di ambiguità.

In questo momento storico, un (ab)uso dei numeri nell'analisi di dati è, purtroppo, quotidianamente esibito per quanto riguarda lo studio della situazione epidemiologica della crisi pandemica legata al virus SARS-2 / COVID-19. Boß-Bavnbek e Krickeberg discutono di quanto sia importante utilizzare la matematica per studiare e, quindi, controllare la pandemia. In un altro articolo, Genest e Rousseau, basandosi esclusivamente su considerazioni matematico-statistiche, propongono una metodologia efficiente per il tracciamento dei contagi.

L'articolo di Cevolini discute di come le assicurazioni possano utilizzare l'enorme quantità

di dati, disponibile grazie alla rete informatica, per generare probabilità di rischio.

Nel suo articolo, Sozzo mostra come il formalismo matematico usato in Meccanica Quantistica sia adeguato per lo studio delle Scienze Cognitive.

Se l'uso della Matematica in ambito sociologico-economico è un fatto acquisito, risulta più difficile concepire l'applicazione della Matematica nell'ambito delle scienze umane. I contributi che seguono presentano delle applicazioni nell'ambito della Linguistica e della Musica.

Per quanto riguarda la Linguistica, l'articolo di Christiansen mostra come la Matematica sia stata applicata al problema della decrittazione dei messaggi cifrati.

Nel suo articolo, Blasone usa la Matematica, e l'analogia con situazioni ben studiate in Fisica, per analizzare tre Studi al pianoforte di Chopin.

Galileo Galilei, considerato il padre della scienza occidentale moderna, afferma che "... il libro dell'universo ... è scritto in lingua matematica...". Abbiamo chiesto a vari colleghi di commentare, in breve, questa frase. L'ultimo contributo di questa prima parte del numero XVII è una raccolta dei commenti che abbiamo ricevuto.

Speriamo che i vari articoli provochino interesse per l'argomento trattato, sia nel suo aspetto più generale che nel dettaglio della specifica applicazione ad una precisa disciplina, e diamo appuntamento per la seconda parte di questo numero XVII.

Buona lettura,
il Comitato di Redazione

Non sempre 2 è il doppio di 1

Una bugia detta con una statistica è fuorviante, ma non si può attribuirne la colpa

Darrell Huff [1]

Paolo Ciafaloni INFN & Dipartimento "Ennio De Giorgi" - Lecce

La realtà ci parla anche attraverso i numeri, ma non sempre chi utilizza i numeri per dare maggiore autorevolezza alle tesi che propugna, lo fa in maniera corretta. In effetti l'utilizzo dei numeri si presta ad ogni forma di falsa interpretazione e mistificazione. In questo articolo cerco di dare qualche nozione di base sulla interpretazione dei numeri per aiutare a sviluppare il senso critico di chi legge.

Numeri e realtà

Interagiamo in molti modi con la realtà, utilizzando i nostri sensi. Certamente non tutte le interazioni necessitano di numeri: un pittore che dipinge un quadro, un musicista che suona uno strumento, un appassionato di mare che surfa le onde, sono esempi di interazioni che non coinvolgono i numeri. Ma a volte i numeri servono, quando ad esempio vogliamo sapere le dimensioni di un tavolo per sapere se entrerà in cucina, quando ci pesiamo, quando utilizziamo il denaro eccetera.

I numeri, diversamente dall'emozione che posso provare di fronte a un tramonto, hanno sempre bisogno di interpretazione. Il numero "500"

acquista un significato solo se specifico, ad esempio, se si riferisce al prezzo in euro di un oggetto. Inoltre dovrei specificare di quale oggetto si tratta, e per sapere se posso permettermi di comprarlo lo devo confrontare con il mio conto in banca e il mio stipendio, e così via. In questo caso l'interpretazione che faccio del numero "500", ovvero l'elaborazione che faccio attraverso il cervello per dare un significato a questo numero, è piuttosto semplice.

Non sempre è così. Ad esempio, posso avere a disposizione una mole enorme di numeri, o dati, che riguardano pressione atmosferica, temperatura, venti di una gran quantità di località della terra in varie epoche, e tuttavia non capire nulla di cosa stia capitando al clima terrestre. In questo caso presumibilmente solo un climatologo è in grado di mettere ordine e dare una corretta interpretazione. Interpretazione che riguarda sia il passato, cioè come il clima si è evoluto nel tempo, sia il futuro, se il climatologo vuole predire l'evoluzione nei prossimi anni.

I numeri di cui parlo in questo articolo non sono di immediata interpretazione, ma non sono così complicati da richiedere un esperto per poter essere interpretati. Con qualche strumento matematico che qui introduco, tali numeri possono essere correttamente interpreta-

ti. Senza questi strumenti, l'interpretazione diventa completamente arbitraria e probabilmente sbagliata.

Parafrasando il divertente e istruttivo libretto di Darrell Huff [1], potremmo dire: "Datemi una qualsiasi tesi e troverò i dati che la confermano". I numeri in effetti sembrano dare un'aura di obiettività a quanto si asserisce, ma il loro significato può essere mistificato in mille modi: scegliendo le medie in maniera opportuna, ignorando alcuni insiemi di dati a favore di altri, ignorando il fatto che ogni numero andrebbe accompagnato dalla relativa incertezza, e così via.

Palline nere e palline bianche

Piero ha davanti a sé una enorme scatola piena di palline bianche e nere, e si chiede quale sia la percentuale di palline bianche. Le palline sono troppo numerose per poterle contare tutte, diciamo un milione per fissare le idee; quindi si chiede in quale modo possa stimare questa percentuale.

Piero inizia estraendo una pallina; è bianca, può concludere che tutte le palline nella scatola sono bianche? Certamente no, può solo concludere che non sono tutte nere. In effetti, da una sola estrazione, si rende conto Piero, non può concludere nulla: le probabilità che sia bianca oppure nera sono sostanzialmente uguali. Intuitivamente, Piero capisce che può avere maggiori informazioni rimettendo dentro la pallina e procedendo a fare molte estrazioni successive, rimettendo dentro ogni volta la pallina estratta per non alterare le percentuali.

Nella figura 1 possiamo vedere il risultato di tale serie di estrazioni: adottiamo la convenzione che il numero 1 sta ad indicare l'estrazione di una pallina bianca estratta e il numero -1 l'estrazione di una pallina nera¹. Piero decide di fare 100 estrazioni, e trova che è stata estratta 55 volte una pallina bianca e 45 volte una pallina nera.

A questo punto però a Piero sembra probabile che la percentuale di palline bianche sia del 55%, ma si rende conto che non può esserne sicuro: per esserlo, dovrebbe contare tutte le palline, che sono molte di più di 100. Allo stesso tempo capisce che è improbabile che la percentuale sia,

¹L'estrazione casuale è stata in realtà simulata da me al computer

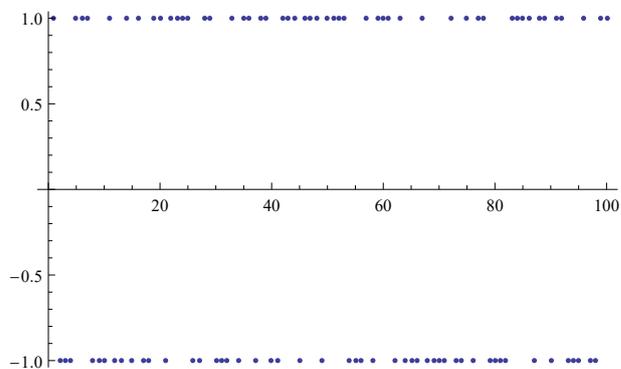


Figura 1: Estrazione casuale di palline ripetuta 100 volte. '1' significa che è stata estratta una pallina bianca, mentre '-1' significa che è stata estratta una pallina nera.

mettiamo, 10%. Come può quantificare il livello di incertezza con cui conosce quello che gli interessa, cioè la percentuale di palline bianche?

Per rispondere a questa domanda occorre utilizzare un po' di matematica. Supponiamo per il momento di imbrogliare, e di sapere che nella scatola ci sono metà palline bianche e metà nere. Quando estraiamo una pallina, la probabilità che sia bianca è $1/2$, così come quella che sia nera. Possiamo definire una variabile x tale che $x = 1$ se la pallina estratta è bianca, e $x = -1$ se la pallina estratta è nera. Avremo quindi $P(x = 1) = P(x = -1) = 1/2$. Possiamo definire la media di questa variabile (E sta per *Expectation value*):

$$E[x] \equiv \bar{x} = P_1 \times 1 + P_{-1} \times (-1) = 0 \quad (1)$$

Più in generale definiamo la media di una variabile casuale, cioè che non ha un valore determinato ma può assumere un certo numero di valori $v_1, v_2 \dots v_n$, come

$$\bar{x} = \sum_{i=1}^n P_i v_i \quad (2)$$

dove v_i sono i valori possibili che può assumere la variabile e P_i le relative probabilità. Che significato ha la media? Supponiamo di ripetere n estrazioni, rimettendo ogni volta la pallina estratta nella scatola; ogni estrazione assume un valore definito x_i .

Definiamo la media misurata come

$$\bar{x}^{(n)} \equiv \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Ci aspettiamo che se facciamo un numero grande di estrazioni (n grande) la media misurata sia molto vicina al valore di aspettazione di x , che possiamo anche chiamare *media vera*. Ovvero, in termini un po' più formali:

$$\lim_{n \rightarrow \infty} \bar{x}^{(n)} = E[x] . \quad (4)$$

Questa aspettativa, ovvero l'avvicinarsi della media misurata, anche detta *media empirica* o *media campionaria*, alla *media vera* con l'aumentare del numero di ripetizioni, è nota come "legge dei grandi numeri". Nella sua versione più formale, ovvero la (4), si chiama "Teorema di Bernoulli". Per quanto l'essenza della legge possa essere intuitivamente quasi ovvia, occorsero 20 anni al matematico Jacob Bernoulli per svilupparne, nel 1713, una dimostrazione sufficientemente rigorosa[2].

Possiamo visualizzare la legge dei grandi numeri facendo una piccola elaborazione sui dati rappresentati in Fig. 1. Fissato il numero di ripetizioni n , effettuiamo il calcolo della media misurata fino a quel momento, in base all'equazione (3). Ripetiamo il procedimento per $n + 1$ ripetizioni e così via. Le medie misurate per n che va da 1 a 100 sono riportate in Fig. (2).

Visualizziamo proprio quello che ci aspettiamo in base alla legge dei grandi numeri. Per piccoli valori di n , diciamo da 1 a 20, la media misurata si discosta molto dal valore vero $E[x] = 0$; per grandi valori di n (diciamo fra 70 e 100) la media misurata è invece vicina al valore vero. Nel linguaggio statistico si dice che le fluttuazioni, cioè le variazioni casuali della media misurata, sono grandi (di ordine 1) per n piccoli ma sono piccole (di ordine 0.1) per n grandi. Ci aspettiamo che al crescere di n oltre 100 diventino ancora più piccole.

Medie e Varianze

Quello che abbiamo detto finora, anche se interessante, è qualitativo; vogliamo sapere adesso qualcosa in più dal punto di vista quantitativo. Piero non si accontenta dell'equazione (4). Egli vuole sapere: fissato n , *quanto* la media misurata è vicina a quella vera? Ovvero, quanto è buona la stima che possiamo fare di $E[x]$?

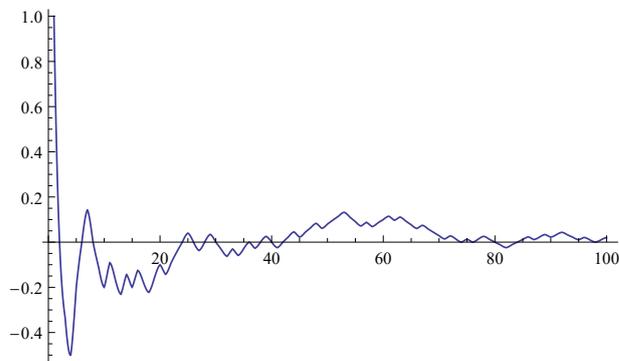


Figura 2: Nel grafico è riportato, per ogni valore di n e in base ai dati di Fig. 1, il valore della media misurata, che è definita in Eq. (3). Le fluttuazioni attorno alla media vera (che è 0) diminuiscono in maniera evidente all'aumentare di n .

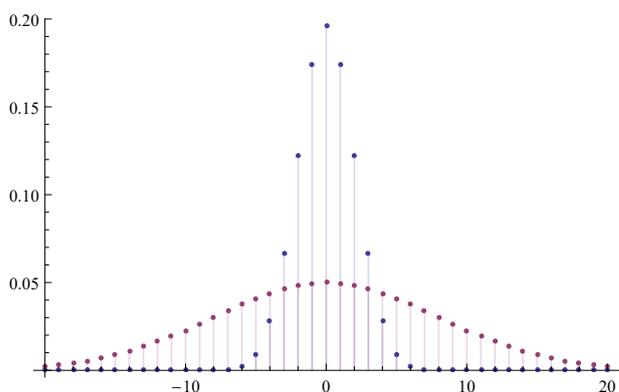


Figura 3: Due distribuzioni di probabilità con la stessa media ($\mu = 0$) ma diversa deviazione standard. La distribuzione coi puntini blu ha deviazione standard $\sigma = 4$, mentre quella coi puntini rossi ha $\sigma = 16$.

Per rispondere a questa domanda introduciamo la *deviazione standard*, detta anche *scarto quadratico medio* σ :

$$\sigma \equiv \sqrt{E[(x - \mu)^2]} = \sqrt{\sum_i P_i (x_i - \mu)^2} . \quad (5)$$

La deviazione standard è una misura della probabilità che una certa quantità devii dal valor medio μ : nella (5) vengono calcolate le deviazioni sia a destra che a sinistra (conta il modulo della differenza $x_i - \mu$) pesate dalle relative probabilità. La deviazione standard è quindi una misura di quanto sia probabile che una data variabile si discosti dal proprio valor medio. In Fig. 3 vediamo due distribuzioni con stessa media ($\mu = 0$ per entrambe) e diversa deviazione standard. È evidente che le probabilità sono significa-

tivamente diverse da 0 all'incirca in un intervallo $\mu - \sigma < x < \mu + \sigma$. La variabile relativa alla distribuzione con $\sigma = 4$ è quindi determinata con maggior accuratezza rispetto alla variabile relativa alla distribuzione con $\sigma = 16$. Idealmente vogliamo diminuire la deviazione standard per avere una stima più accurata. Procediamo quindi a calcolare la deviazione standard nel caso della prima estrazione di pallina (si ha $\mu = 0$ in questo caso):

$$\sigma^2 = E[x^2] = \sum P_i x_i^2 = \frac{1}{2}(1)^2 + \frac{1}{2}(-1)^2 = 1 \quad (6)$$

La deviazione standard in questo caso è quindi 1, un numero grande in quanto dello stesso ordine dei due possibili valori assunti dalla variabile. Ma andiamo avanti e calcoliamo la deviazione standard della media di due estrazioni, $\bar{x}^{(2)}$:

$$E\left[\left(\frac{x_1 + x_2}{2}\right)^2\right] = \frac{1}{4}(E[x_1^2] + E[x_2^2] + 2E[x_1 x_2]) \quad (7)$$

Abbiamo già calcolato $E[x_1^2] = E[x_2^2] = 1$. Per quanto riguarda la variabile $x_1 x_2$, essa può assumere valore 1 con probabilità 1/2, e -1 con probabilità 1/2; si ha quindi $E[x_1 x_2] = 0$.

Questo è un risultato del tutto generale quando due variabili sono incorrelate, cioè quando il risultato dell'una non influenza il risultato dell'altra: il valore di aspettazione del loro prodotto è 0 (se la media non è nulla, ad essere 0 è $E[(x_1 - \mu)(x_2 - \mu)]$).

Siamo in grado ora di calcolare la deviazione standard, che è la radice quadrata dell'equazione (7): $\sigma(\bar{x}^{(2)}) = 1/\sqrt{2}$. Quindi la deviazione standard è diminuita rispetto al valore ($\sigma = 1$) della singola variabile. Cosa succede quando si calcolano le fluttuazioni della variabile media dopo n estrazioni

$$\bar{x}^{(n)} = \sum_i^n \frac{x_i}{n} ?$$

Otteniamo

$$\begin{aligned} \sigma_n &= \sqrt{\frac{1}{n^2} \left(\sum_i^n E[x_i] \right)^2} \\ &= \sqrt{\frac{1}{n^2} \sum_i^n (E[x_i^2])} = \frac{1}{\sqrt{n}} \quad (8) \end{aligned}$$

avendo tenuto conto del fatto che, essendo le variabili incorrelate, $E[x_i x_j] = 0 \forall i, j$.

La (8) dà un significato quantitativo alla legge dei grandi numeri: non solo le fluttuazioni (deviazione standard) diminuiscono con il numero di ripetizioni, ma adesso sappiamo che diminuiscono come l'inverso della radice di n .

Possiamo fare un discorso analogo per il numero di palline bianche estratte, $b^{(n)}$ (nel nostro esempio, $b^{(100)} = 55$). La variabile $\bar{x}^{(n)}$ conta le palline bianche con peso 1, e quelle nere con peso -1; inoltre il numero delle palline nere è $n - b^{(n)}$, per cui:

$$\bar{x}^{(n)} = 2 \frac{b^{(n)}}{n} - 1; \quad b^{(n)} = (\bar{x}^{(n)} + 1) \frac{n}{2} \quad (9)$$

A una variazione $\Delta \bar{x}^{(n)}$ della media corrisponde quindi una variazione delle palline bianche $\Delta b^{(n)} = \Delta \bar{x}^{(n)} \frac{n}{2}$, per cui abbiamo

$$\sigma_b = \frac{n}{2} \sigma_n = \frac{1}{\sqrt{n}} \frac{n}{2} = \frac{\sqrt{n}}{2} \quad (10)$$

In definitiva, le fluttuazioni della nostra variabile $\bar{x}^{(n)}$ scalano con n come $1/\sqrt{n}$ mentre le fluttuazioni del numero di palline bianche scalano come \sqrt{n} .

Piero può quindi dare un significato più preciso alla sua stima del numero di palline bianche. In effetti Piero può prendere come misura dell'incertezza nel numero di palline bianche il valore di σ_b , quindi ritenere ragionevolmente che la media vera sia compresa fra il valore $\mu - \sigma_b$ e il valore $\mu + \sigma_b$. Avendo effettuato 100 ripetizioni ed essendo quindi $\sigma_b = \sqrt{n}/2 = 5$, Piero conclude che la sua stima della frazione di palline bianche è:

$$\frac{55 \pm 5}{100} = (55 \pm 5)\% \Leftrightarrow E[x] = \mu = 0.1 \pm 0.1 \quad (11)$$

Limite Centrale

Che significato ha l'equazione (11)? Più precisamente, con quale probabilità la stima $50\% \leq \mu \leq 60\%$ contiene la media vera, sconosciuta a Piero? Per rispondere a questa domanda quantitativa è utile introdurre un'approssimazione che utilizza una funzione dovuta al matematico e fisico Carl Friederich Gauss [3], scritta in termini dei

parametri μ, σ :

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} . \quad (12)$$

Questa distribuzione suppone l'uso di una variabile x continua, ed è quindi diversa dalle distribuzioni discrete finora discusse, in cui si assegna la probabilità P_i che una data variabile assuma un certo valore x_i (vedi l'equazione (6)). Per le distribuzioni continue di probabilità, il significato è che $G(a)da$ è la probabilità che la variabile x assuma un valore compreso fra a e $a + da$. Nel caso della Gaussiana otteniamo i valori di media e deviazione standard:

$$E[x] = \int xG(x)dx = \mu; \quad \sqrt{E[(x - \mu)^2]} = \sigma \quad (13)$$

L'importanza della gaussiana come distribuzione di probabilità è data dal Teorema del limite centrale [4], che possiamo enunciare nel modo seguente:

Una variabile casuale che sia somma di n variabili casuali, purché n sia sufficientemente grande, ha una distribuzione di tipo gaussiano indipendentemente dalle distribuzioni delle n variabili.

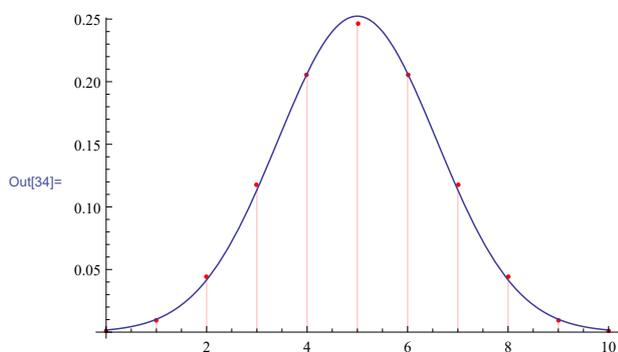


Figura 4: Il teorema del limite centrale: distribuzione della variabile $\bar{x}^{(n)}$ con $n=10$ (puntini rossi) e distribuzione gaussiana (linea blu).

Questo teorema è piuttosto sorprendente, in quanto vale qualsiasi siano le distribuzioni delle n variabili che concorrono a formare la variabile finale; questo significa che la distribuzione gaussiana è presente in moltissimi casi in cui l'effetto finale sia la somma di tanti piccoli effetti.

Ad esempio, la nostra variabile $\bar{x}^{(n)}$ definita dall'equazione (3) è la somma di tante variabili x_i , e quindi possiamo ritenere che la sua distribuzione sia gaussiana con $\mu = 0$ e $\sigma = 1/\sqrt{n}$. Ma

quando possiamo dire che n è **sufficientemente grande**?

In figura 4 appare il confronto fra la distribuzione di palline bianche con $n = 10$ (puntini rossi) e la distribuzione di Gauss; come si vede l'approssimazione data dalla gaussiana è buona, e ci si aspetta che sia ancora migliore per il nostro $n = 100$.

Come usiamo la distribuzione gaussiana per dare un significato quantitativo alla stima di Piero, Eq. (11)? Supponiamo di ripetere il nostro esperimento di estrazione di 100 palline per 1,2,...,N volte. Quante volte il risultato che otteniamo per la media cade in intervallo di $\pm 5\%$ intorno alla media vera? Ovvero, quanto volte la nostra stima $50\% < \mu < 60\%$ contiene la media vera? Rispondere a questa domanda è piuttosto semplice: tale probabilità P è data dall'integrale

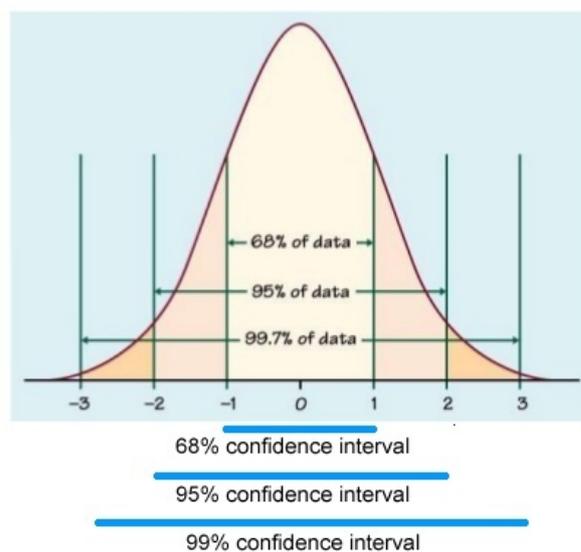


Figura 5: Gli intervalli di confidenza sono connessi all'area sotto la curva gaussiana: da $-\sigma$ a σ tale area vale circa 0.68, da -2σ a 2σ vale circa 0.95, da -3σ a 3σ vale circa 0.99

fra 50 e 60 della Gaussiana con media 50 e deviazione standard $\frac{\sqrt{n}}{2} = 5$, e vale circa 0.68, e se N è molto grande il numero di volte in cui la media misurata μ sta nell'intervallo prescelto è $PN = 0.68N$.

Piero sintetizza questa informazione dicendo che:

$$\frac{n_b}{n} = 55\% \pm 5\% \quad C.L. = 68\% , \quad (14)$$

dove $C.L.$ sta per confidence level, ovvero inter-

vallo di confidenza, ed ha il significato discusso sopra.

In genere si scelgono intervalli più ampi di intervallo di confidenza, come 2σ (=10 nel nostro caso) ad esempio:

$$\frac{n_b}{n} = 55\% \pm 10\% \quad C.L. = 95\% \quad (15)$$

in quanto l'integrale della distribuzione gaussiana fra $\mu - 2\sigma$ e $\mu + 2\sigma$ vale circa 0.95.

Le equazioni (14) e (15) sono infine ciò che Piero cercava. Esse quantificano, una volta scelto, un certo intervallo di confidenza, l'intervallo in cui riteniamo stia la media vera.

Queste sono informazioni di tipo probabilistico, e non potrebbe essere altrimenti dato che stiamo effettuando una stima. Ad esempio, se scelgo 68% C.L., esiste una probabilità del 32 % che la media vera **non** stia nell'intervallo definito dall'equazione (14). Posso diminuire tale probabilità fino al 5 %, ma l'intervallo della mia stima diventa più ampio (vedi l'equazione (15)). Per finire, abbiamo fin qui assunto di conoscere a priori la deviazione standard, ma in molti casi non è così. Si pensi ad esempio al caso in cui vogliamo stimare l'altezza delle donne adulte italiane ad esempio: non possiamo sapere quale sia la deviazione standard della vera distribuzione. Possiamo però stimare dal nostro campione la deviazione standard utilizzando l'espressione usuale:

$$\sigma^2 = \sum_i P_i (x_i - \mu)^2, \quad (16)$$

in cui μ è stimato dal campione nel modo usuale e P_i , invece che essere ricavato da una distribuzione nota a priori, è stimato in base alle frequenze dei valori: $P_i = n_i/n$ dove n_i è il numero di volte che ricorre un dato valore di altezza e n il numero totale. Ad esempio, nel caso di Piero $n_{bianche} = 55, n_{nere} = 45, n = 100$ per cui per la singola variabile x_i abbiamo $P_1 = .55, P_{-1} = .45$, da cui stimiamo media e varianza:

$$\begin{aligned} \mu &= 0.1 \\ \sigma^2 &= (1 - \mu)^2 0.55 + (-1 - \mu)^2 0.45 = 0.99 \end{aligned}$$

In base all'equazione (10), questo si ripercuote in una fluttuazione del numero di palline bianche pari a

$$\sigma_b = \sqrt{0.99} \frac{\sqrt{n}}{2} \approx 4.97,$$

un po' più piccola (ma non molto) della varianza vera pari a 5.

La stima di Piero per la media delle palline bianche a $2\sigma_b \approx 9.9\%$ diventa quindi:

$$\frac{n_b}{n} = 55\% \pm 9.9\% \quad 95\% C.L. \quad (17)$$

È bene rimarcare che il valore vero del numero di palline bianche può essere al di fuori dell'intervallo $45.1\% < \frac{n_b}{n} < 64.9\%$; la probabilità che questo accada è però piccola, circa del 5%.

Andamento col tempo

Spesso quello che ci interessa non è una fotografia della situazione, come nel caso del numero di palline bianche in una scatola, bensì capire come evolve nel tempo una data quantità. Ad esempio, potremmo essere interessati a come cambiano da 100 anni fa ad oggi l'altezza media e l'aspettativa di vita in Italia, o l'andamento negli ultimi sei mesi del numero di contagiati di Covid-19.

Va subito detto che nel fare questo dovremo necessariamente tener conto delle fluttuazioni statistiche (deviazione standard). Ad esempio, supponiamo che ci siano circa 10.000 tamponi positivi in un dato giorno durante la pandemia di Covid-19. Come abbiamo visto, le fluttuazioni sono dell'ordine di $\sqrt{10.000} = 100$. Queste fluttuazioni, che possiamo chiamare statistiche, non sono però le uniche. Se confrontiamo il numero di contagiati in un giorno con quelli del giorno successivo, ci saranno molte altre cause di fluttuazione: in un dato giorno ad esempio si sono effettuati molti più tamponi del giorno precedente. Oppure in un dato giorno si sono fatti più tamponi in una regione maggiormente colpita, e ancora si sono fatti più tamponi a una data fascia di età (gli anziani ad esempio, maggiormente a rischio). Per non parlare del fatto che si sono fatti tamponi solo a chi lo chiede oppure a chi mostra sintomi, e non sappiamo nulla di tutti coloro che sono positivi ma non mostrano sintomi; e così via. In definitiva, le fluttuazioni, cioè le variazioni nel conteggio dei positivi che nulla hanno a che vedere con il vero numero di positivi, saranno numericamente molto più grandi di 100.

Questo significa che se un giorno si misurano 10.000 nuovi positivi e il giorno dopo 11.000, non è corretto dire che i positivi sono aumenta-

ti. Quello che stiamo osservando sono piuttosto fluttuazioni non significative intorno alla media 10.500. In generale, possiamo dire che non è sensato dare i numeri, cioè analizzare le variazioni del numero di positivi da un giorno all'altro, cosa che peraltro i media fanno di continuo e non solo per il Covid-19. Più sensato sarebbe diminuire il peso delle fluttuazioni considerando un numero più elevato di tamponi. Si potrebbe valutare la differenza del numero di tamponi positivi da una settimana all'altra ad esempio, anziché da un giorno all'altro.

Una prima informazione che vogliamo estrarre (tenendo conto delle fluttuazioni) è se la quantità in oggetto sia sostanzialmente invariata con il tempo, oppure se aumenti (o diminuisca). Dopodiché può interessare con con quale velocità la quantità stia aumentando, cioè quale tipo di funzione descriva meglio l'andamento temporale. Spesso, almeno nel caso delle epidemie, si usano approssimazioni di tipo lineare oppure esponenziale. Una data quantità segue una curva lineare se la crescita (o decrescita) da un giorno all'altro è costante. Se invece la crescita (o decrescita) giornaliera è una percentuale costante del valore del giorno precedente, allora l'andamento è detto esponenziale.

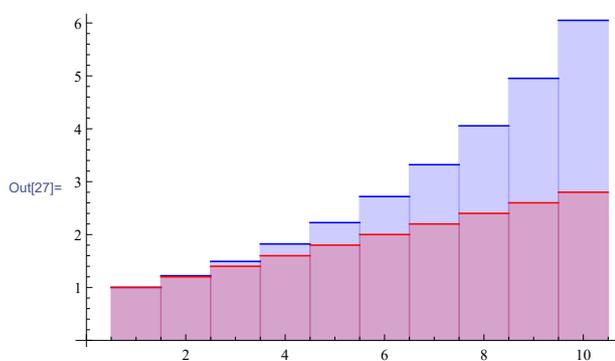


Figura 6: Confronto fra andamento lineare (in rosso) ed esponenziale (in blu). Nel caso lineare, la quantità varia di 0.2 ogni giorno, per un totale di 10 giorni. Per l'esponenziale l'aumento è del 20% ogni giorno.

In figura 6 sono riportati i grafici di un andamento lineare (in rosso) e uno esponenziale (in blu). Il valore iniziale della variabile è 1; nel caso lineare tale variabile cresce di 0.2 ogni giorno mentre nel caso esponenziale cresce del 20% ogni giorno. Inizialmente le due curve sono vicine ed è difficile distinguerle, ma poi l'esponenziale cre-

sce molto di più in quanto la quantità di cui aumenta ogni giorno non è costante bensì aumenta a sua volta col tempo. Da queste curve dovrebbe apparire chiara l'assurdità di certe conclusioni che si sentono trarre: se una quantità triplica da un giorno all'altro non ha nessun senso dire che "è aumentata esponenzialmente". Per capire se una crescita è di tipo esponenziale oppure no occorre seguire l'andamento per diversi giorni.

Supponiamo di fare, ad esempio, l'ipotesi che una data quantità x sia costante nel tempo. Naturalmente, a causa delle fluttuazioni, questa quantità non assumerà esattamente lo stesso valore ogni giorno. Come possiamo fare per verificare la nostra ipotesi quindi? Per prima cosa determiniamo il valore della costante, chiamiamola k , in modo da minimizzare la somma degli scarti quadratici relativi ad ogni valore x_i corrispondente al giorno i , con n giorni totali:

$$\frac{\partial}{\partial k} \sum_{i=1}^n (x_i - k)^2 = -2 \sum_{i=1}^n x_i + 2kn = 0$$

$$\Rightarrow k = \frac{1}{n} \sum_{i=1}^n x_i . \quad (18)$$

Determinare la costante k (che poi coincide, e non è una sorpresa, con la usuale media μ) ammonta a fare un *fit* dei dati con una funzione costante; il metodo qui utilizzato si chiama metodo dei minimi quadrati. Ma come possiamo giudicare la bontà del *fit*?

Definiamo per prima cosa una variabile che quantifichi le fluttuazioni, sommando gli scostamenti dalla media:

$$\chi^2 \equiv \sum_{i=1}^n \frac{(x_i - \mu)^2}{\mu} . \quad (19)$$

Gli scostamenti quadratici vengono qui divisi per la media in quanto ci si attende che $\sigma_i^2 \sim \mu$; ogni punto x_i contribuisce circa 1 al χ^2 . Se le distribuzioni degli x_i sono di tipo gaussiano allora la distribuzione del χ^2 è nota [6].

In maniera analoga a quanto visto in precedenza, possiamo stabilire un *C.L.*. Se il χ^2 viene costruito a partire da n variabili indipendenti $x_1 \dots x_n$, ci aspettiamo che il suo valore sia vicino a n . Se invece il valore del χ^2 è molto più grande di n , allora diciamo che la qualità del *fit* è cattiva e la nostra ipotesi che, ad esempio, i dati vengano

descritti da una funzione costante, è sbagliata.

Supponiamo di avere ad esempio 4 variabili indipendenti. Allora, dall'integrale della distribuzione del χ^2 [5] otteniamo che la probabilità di avere un valore maggiore o uguale a 7.78 è del 90 %. Quindi, con il C.L. del 90%, se il valore che otteniamo dai nostri dati è maggiore di 7.78 rigettiamo l'ipotesi che i dati siano costanti nel tempo.

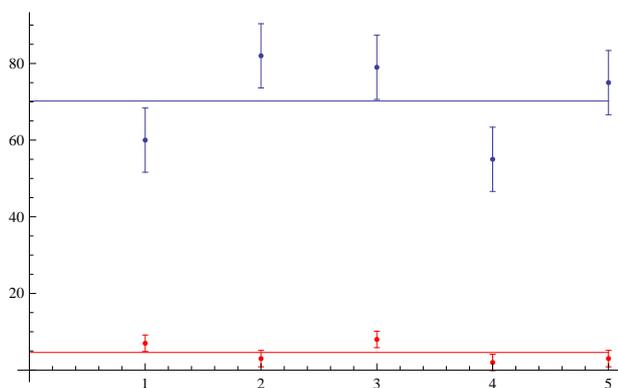


Figura 7: Due insiemi di dati al variare del giorno (da 1 a 5)

In figura 7 sono rappresentati, in rosso e in blu, due insiemi di dati insieme alle rispettive medie, in un arco temporale di 5 giorni. Nella figura sono rappresentate anche le incertezze, o fluttuazioni ($\pm 1\sigma$), come segmenti verticali per ogni dato. In entrambi i casi il numero di variabili indipendenti (o gradi di libertà) è 4: infatti, una volta fissata la media si ha un vincolo che permette di ricavare uno dei dati (ad esempio quello relativo al quinto giorno) dagli altri.

Il valore di χ^2 che si ottiene per i dati in rosso è 6.35, per quelli in blu è 8.19. Siccome il valore critico al livello di confidenza del 90% per 4 gradi di libertà è 7.78, rigettiamo l'ipotesi che i dati siano descritti da una costante per il caso blu mentre la accettiamo per il caso rosso. Notare che nel caso dei dati in rosso, il valore numerico è 3 nel giorno 2 e 8 nel giorno 3: malgrado questo i dati sono ben descritti da una costante. Questo fa capire l'importanza di tener conto delle fluttuazioni nell'analizzare i dati e capire gli andamenti.

In maniera analoga a quanto fatto finora per la funzione costante, si possono fare *fit* dei dati con qualsiasi tipo di funzione, come ad esempio funzioni lineari (retta), esponenziale, polinomi

ecc. Cruciale è l'importanza del numero di gradi di libertà (o variabili indipendenti).

È evidente che dato un insieme qualsiasi di punti posso sempre tracciare una curva che passi per ognuno di essi, ma in questo caso il *fit* è privo di significato. Ad esempio, se ho solo due dati posso sempre fare un *fit* con una retta, ma questo non mi dice nulla né sull'andamento dei dati né sulle possibili estrapolazioni (previsioni).

Supponiamo di fare un *fit* con un polinomio di grado n ; la funzione in questo caso è definita da $n + 1$ parametri. Il numero di gradi di libertà rilevante in questo caso è

$$\text{gradi di libertà} = \text{numero di variabili} - (n+1).$$

Se faccio un fit di n punti con un polinomio di grado $n - 1$, il numero di gradi di libertà è 0: questo corrisponde al fatto che posso sempre determinare i parametri della mia funzione in modo che passi per gli n punti. In questo caso il test del χ^2 non è significativo. Per aver un buon *fit*, oltre ad avere un valore basso di χ^2 , la mia curva deve essere definita da un numero limitato di parametri.

Conclusioni

Se si dovesse estrapolare un unico messaggio da questo articolo sarebbe questo: **qualsiasi numero, per essere correttamente interpretato, va accompagnato dalla relativa incertezza**. Quindi, ha poco significato dire ad esempio "il numero attuale di contagiati da Covid-19 in Italia è 100.000", mentre ha senso dire "il numero di contagiati è compreso fra 93.000 e 105.000 con una probabilità (livello di confidenza) del 90%".

Oltre a questa consapevolezza, dovremmo sempre tener presente che se si fa un conteggio, le fluttuazioni statistiche sono più grandi nel caso di numeri piccoli. Per cui, non è sempre vero che 2 è il doppio di 1: i due numeri sono di fatto equivalenti una volta tenuto conto delle fluttuazioni. Quello di cui non ho parlato qui invece è di come la scelta di quello che si chiama un campione non significativo possa portare a false interpretazioni e mistificazioni: invito alla lettura di [1] per questo tema.



- [1] Darrell Huff: *How to Lie with Statistics*, Norton, New York (1954).
- [2] https://en.wikipedia.org/wiki/Law_of_large_numbers
- [3] https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss
- [4] https://en.wikipedia.org/wiki/Central_limit_theorem
- [5] <https://math-wiki.com/images/5/5a/Chi-table.pdf>
- [6] https://en.wikipedia.org/wiki/Chi-square_distribution#Probability_density_function



Paolo Ciafaloni: è dipendente dell'Istituto di Fisica Nucleare (INFN). Svolge attività di ricerca nel campo della fisica teorica delle particelle e della Cosmologia, e insegna teoria delle interazioni forti.

Dinamica e controllo del Covid-19: commenti da due matematici

Bernhelm Booß-Bavnbek

Università di Roskilde, Danimarca

Klaus Krickeberg

Manglieu, Francia e Bielefeld, Germania

Poiché la pandemia di Covid-19 non è ancora terminata, potrebbe sembrare prematuro trarre delle conclusioni. Tuttavia, potremmo anche essere già in tempo per ricapitolare alcune lezioni che noi, come matematici, dovremmo aver imparato e che sono da applicare adesso. Quindi noi poniamo la domanda: perché la dinamica ed il controllo del Covid-19 sono di interesse per i matematici, e perché i matematici sono urgentemente necessari per controllare la pandemia?

Presenteremo dapprima i nostri commenti in un'esposizione dal basso verso l'alto, cioè seguendo gli eventi dal loro inizio come si sono evoluti nel tempo. Sono accaduti in modo diverso in diversi paesi, e l'obiettivo principale di questa prima parte è di confrontare tra loro queste evoluzioni nei pochi paesi selezionati.

Tuttavia, vi sono alcune caratteristiche generali, che presentiamo separatamente, come siamo abituati a fare in matematica. Queste includono la storia di alcune epidemie che hanno influenzato le reazioni delle persone in molti paesi, e alcuni strumenti matematici di base. Inoltre, vi

Since the Covid-19 pandemic is not yet over, it may appear to be premature to draw some conclusions. However, we may also be just in time to recapitulate some lessons we as mathematicians should have learned and are urged to apply now. Thus, we pose the question: Why are the dynamics and control of Covid-19 most interesting for mathematicians and why are mathematicians urgently needed for controlling the pandemic?

We shall first present our comments in a bottom-up approach, i.e., following the events from their beginning as they evolved through time. They happened differently in different countries, and the main objective of this first part is to compare these evolutions in a few selected countries with each other.

Still, there are some general features, which we present separately as we are used to doing in mathematics. They include the history of certain epidemics which have influenced the reactions of people in many countries, and some basic mathematical tools. In addition, there is a common

è un fattore comune, che uno degli autori (KK) ha fatto presente in un messaggio inviato il 12 marzo 2020 a un ufficio sanitario tedesco:

"L'estensione e l'evoluzione del Covid-19 nei vari paesi e regioni riflette lo stato dei loro sistemi sanitari. Questo era per esempio già evidente nel caso dell'epidemia di Ebola."

Infatti è la componente di sanità pubblica del sistema sanitario a giocare un ruolo cruciale.

La seconda parte dell'articolo non è orientata verso il paese, ma è focalizzata sul problema. Partendo da un determinato problema procediamo dall'alto verso il basso arrivando alle sue soluzioni e alle loro applicazioni in situazioni concrete. Abbiamo organizzato questa parte in base ai metodi matematici che hanno un ruolo nella loro soluzione. Ecco un esempio specifico in cui molta matematica è necessaria: sviluppare un vaccino e la strategia per applicarlo senza perdere di vista i principi etici di base.

Dal basso verso l'alto

1. La preistoria

Nel seguito, il gentile lettore può consultare, quando necessario, il libro [1] per i concetti di base dell'epidemiologia.

La demografia come materia matematica era già sviluppata secoli fa ben oltre le sue origini elementari. Per molto tempo è rimasta l'unico strumento matematico nello studio dell'evoluzione delle malattie infettive. Ecco un primo esempio famoso. In Cina, India e Europa si è cercato di conferire l'immunità contro il vaiolo infettando leggermente gli individui in modo che contraessero una forma lieve della malattia in modo da essere in seguito immuni. Alcuni di questi morirono dopo una tale procedura, ma nel 1766 il matematico svizzero Daniel Bernoulli mostrò mediante un approccio demografico che la procedura aumenterebbe l'aspettativa di vita se applicata a ogni individuo [2]. Al giorno d'oggi, si pratica ampiamente la valutazione del rapporto costo-efficacia di una misura di sanità pubblica; ciò è fatto basandosi su metodi di economia matematica.

factor, which one of the authors (KK) defined on 12 March 2020 in an e-mail to a German health office:

"The extension and evolution of Covid-19 in various countries and regions reflects the state of their health systems. This was for instance already very obvious in the case of Ebola."

It is in fact the public health component of the health system that plays a crucial role.

The second part of the article is not country-oriented but problem-oriented. From a given problem we go top-down to its solutions and their applications in concrete situations. We have organised this part by the mathematical methods that play a role in their solution. Here is an example where specially much mathematics is needed: to develop a vaccine and the strategy for applying it without losing sight of basic ethical principles.

Bottom-up

1. Prehistory

In the following, the gentle reader may consult when necessary the book [1] for the basic concepts of epidemiology.

Demography as a mathematical subject area was already developed centuries ago well beyond its elementary beginnings. For a long time, it remained the only mathematical tool in the study of the evolution of infectious diseases. Here is a famous early example. In China, India and Europe one tried to confer immunity against smallpox by infecting individuals slightly so they would contract a mild form of the disease and be immune afterwards. Some of them died by this procedure, but in 1766 the Swiss mathematician Daniel Bernoulli showed by a demographical approach that the procedure would increase life expectancy if applied to everybody [2]. Nowadays, evaluating the cost-effectiveness of a public health measure is done widely; it is based on methods of mathematical economy.

Il XIX secolo ha visto la scoperta di microrganismi come agenti patogeni di molte malattie e il loro studio principalmente mediante metodi microbiologici. Gli strumenti matematici per seguire un'epidemia sono rimasti essenzialmente demografici fin ben dentro il XX secolo. Alcuni medici hanno suggerito che ogni epidemia finisce perché alla fine non sono rimaste abbastanza persone che possano essere infettate: è questo un predecessore ingenuo del concetto di epidemiologia matematica di immunità del gregge (cfr. Sez. 8). Tuttavia, anche l'abbondante letteratura sulla pandemia influenzale del 1918-19, erroneamente chiamata influenza Spagnola, discute solo due possibili vie per la sua fine: un migliore trattamento clinico e mutazioni dell'agente patogeno.

Da un punto di vista virologico, l'influenza spagnola è stata una forma estrema della cosiddetta influenza stagionale. Il virus che la causa può essere uno di una grande varietà, il suo genere è indicato da A, B, C o D, dove alcuni di essi comprendono diverse specie. A è la più grave e ha sottotipi $A(H_xN_y)$, $x = 1, \dots, 18$ e $y = 1, \dots, 11$, dove x e y rappresentano proteine sulla superficie del virus. La strategia di controllo della normale influenza stagionale è ampiamente conosciuta anche tra i profani: identificare il ceppo del virus in autunno, sviluppare un vaccino il più velocemente possibile, e vaccinare le persone che si pensa siano a rischio. Tuttavia, il numero di infezioni e di decessi causati da un'influenza stagionale può essere alto quanto quello di alcune delle pandemie che saranno descritte ora.

L'influenza spagnola era dovuta all' $A(H_1N_1)$. Immagini di quel tempo mostrano persone che indossano maschere che assomigliano a quelle usate adesso. Negli anni 1957-58 si è verificata un'altra variante chiamata influenza Asiatica e causata da $A(H_2N_2)$. È iniziata in Cina e poi è diventata una pandemia, passando dagli stati confinanti attraverso il Regno Unito e gli Stati Uniti. Le stime del numero di casi variano intorno a 500 milioni e quelle del numero di morti intorno a 3 milioni di persone. I suoi inizi apparivano molto simili a quelli dell'influenza spagnola, ma verso la fine si è reso disponibile un vaccino, un predecessore di quelli che sono usati abitualmente contro l'influenza stagionale.

L'influenza di Hong Kong del 1968-69, genera-

The 19th century saw the discovery of microorganisms as pathogens of many diseases and their study by mainly microbiological methods. The mathematical tools for following up an epidemic remained essentially demographical well into the 20th century. A few physicians suggested that every epidemic ends because there are finally not enough people left to be infected, which is a naïve predecessor to the mathematical-epidemiologic concept of herd immunity (see Sect. 8). Nevertheless, even the abundant literature on the influenza pandemic of 1918-19, wrongly called Spanish flu, only discusses two possible ways for it to end: better clinical treatment and mutations of the pathogen.

Seen from a virological viewpoint, the Spanish flu was an extreme form of the so-called seasonal influenza. The virus which causes them can be one of a large variety, its genus being denoted by A, B, C or D, where some of them include several species. A is the most serious one; it has subtypes $A(H_xN_y)$, $x = 1, \dots, 18$ and $y = 1, \dots, 11$, where x and y represent proteins on the surface of the virus. The strategy for controlling the normal seasonal influenza epidemic is widely known even among laymen: identify the strain of the virus in the autumn, develop a vaccine as fast as possible, and vaccinate people thought to be at risk. Nevertheless, the number of infections and deaths by a seasonal influenza can be as high as those caused by some of the pandemics to be described now.

The Spanish flu was due to $A(H_1N_1)$. Pictures from that time show people wearing masks that resembled those used now. In the years 1957-58 another digression from seasonal influenza occurred, called the Asian flu and caused by $A(H_2N_2)$. It started in China and then became a pandemic, passing from neighbouring states through the UK and the USA. Estimations of the number of cases vary around 500 million and of the number of deaths around 3 million. Its beginnings looked much like those of the Spanish flu, but towards the end a vaccine became available, a predecessor to the ones being routinely used now against the seasonal flu.

The Hong Kong influenza of 1968-69, genera-

ta dal virus A(H₃N₂), aveva caratteristiche simili e non sarà descritta oltre.

Parallelamente a queste e ad altre epidemie che entrano in scena, e in parte motivati da queste, sono emersi nella prima parte del XX secolo strumenti matematici radicalmente nuovi per lo studio della sanità pubblica, preceduti da alcuni studi sul finire del XIX secolo. Avevano un duplice aspetto. Il primo strumento si chiamava modello matematico-statistico. Il suo scopo è lo studio dell'influenza di fattori, chiamati anche determinanti, sulla salute della popolazione. Tali fattori possono, ad esempio, essere una mancanza di igiene o un ambiente inquinato. Un fattore può anche essere un trattamento preventivo o curativo mediante un'immunizzazione o un farmaco, rispettivamente; in questo caso l'obiettivo principale di uno studio è di stimare l'efficacia del trattamento. I piani di campionamento sono modelli statistici di un tipo diverso ma correlato. Essi formano la base di indagini a campione, che sono effettuate con profusione anche nel Covid-19, e non sono sempre molto illuminanti.

Il secondo strumento si chiama modellizzazione matematica dell'evoluzione di un'epidemia, o brevemente modellizzazione matematica. Ve ne sono di due tipi. In primo luogo, si può mirare alla curva epidemica, che è il numero cumulativo di casi fino ad un momento t in funzione di t . In questo caso, la modellizzazione matematica serve a stimare o prevedere questa curva sotto varie ipotesi sull'infettività di soggetti infetti. I primi predecessori sono presentati in [3], vedi Figura 1. La questione se l'infettività rimanga costante o diminuisca ha già avuto un ruolo importante. Versioni raffinate sono ancora in uso, in particolare per il Covid-19 (Sez. 7).

In secondo luogo, si possono costruire i cosiddetti modelli compartimentali (Sez. 8). Il primo, per il morbillo, è stato pubblicato nel 1889 da P.D. En'ko; cfr. [4]. Intorno al 1900, sono apparsi modelli compartimentali per la malaria. Quindi negli anni 20 del Novecento, furono messi a punto, e studiati intensamente, nuovi modelli per l'evoluzione del morbillo in popolazioni chiuse. Questi sono diventati molto influenti perché hanno già mostrato molte caratteristiche di base che sono riapparse più tardi nei modelli matematici di epidemie in altri e più complessi contesti.

Tali strumenti hanno trovato molte applicazio-

ated by the virus A(H₃N₂), had similar characteristics and will not be described further.

Parallel to these and other epidemics entering the scene, and partly motivated by them, basically new mathematical tools of public health emerged in the first part of the 20th century, preceded by a few studies in the late 19th. They were twofold. The first tool was called a statistical-mathematical model. Its aim is the study of the influence of factors, also called determinants, on the health of the population. Such factors may, for instance, be a lack of hygiene or a polluted environment. A factor can also be a preventive or curative treatment by an immunisation or a drug, respectively; in this case the main objective of a study is to estimate the efficacy of the treatment. Sampling plans are statistical-mathematical models of a different but related kind. They form the basis of sample surveys, which are likewise being done in profusion about Covid-19, and are not always very illuminating.

The second tool is called mathematical modelling of the evolution of an epidemic or briefly mathematical modelling. There are two kinds of it. First, one may aim for the epidemic curve, which is the cumulated number of cases up to a moment t as a function of t . In this case, mathematical modelling serves to estimate or predict this curve under various assumptions on the infectivity of infected subjects. Early predecessors are presented in [3], see Figure 1. The question whether the infectivity remains constant or decreases already played a role. Refined versions are still being used, in particular for Covid-19 (Sect. 7).

Second, one may build so-called compartmental models (Sect. 8). The first one, for measles, was published in 1889 by P.D. En'ko; see [4]. Around the year 1900, compartmental models for malaria appeared. Then in the 1920s, new models for the evolution of measles in closed populations were defined and intensively studied. They became very influential because they already displayed many basic features that reappeared later in mathematical models of epidemics in other and more complex settings.

Such tools found many applications. Dealing

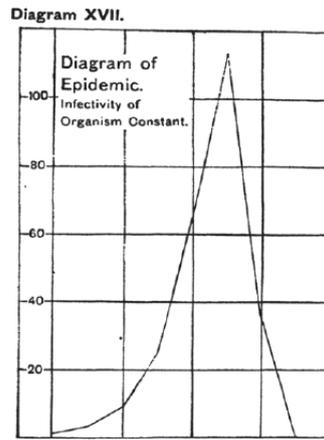


Figura 1: Prime simulazioni numeriche di una curva epidemica fatte da J. Brownlee, 1907, discusse in [3].
Early numerical simulation of an epidemic curve by J. Brownlee, 1907, discussed in [3].

ni. Trattare matematicamente grandi epidemie non era più un problema di sola demografia, anche se questo ha continuato ad essere lo strumento principale per la stima del numero di casi e di decessi. Sono stati utilizzati modelli statistici per stimare l'efficacia dei farmaci antivirali, ad esempio contro le infezioni da HIV e l'efficacia di varie immunizzazioni, comprese quelle contro le forme di influenza. La modellizzazione matematica delle epidemie è stata utilizzata nel pianificare strategie per debellare il vaiolo, la poliomielite, il morbillo e forse altre malattie. Nella letteratura scientifica, sono apparsi i primi articoli sulla modellizzazione delle epidemie influenzali. La pianificazione di una strategia di vaccinazione comporta sia una strategia statistico-matematica sia modelli matematici [5].

Questi progressi possono aver prodotto una sensazione generale di successo nell'affrontare le epidemie. Poi, nel periodo 2002-2019 si sono verificati alcuni eventi che hanno evocato ricordi di pandemie precedenti e minato tale fiducia.

2. Eventi inattesi 2002-2018

Nel novembre 2002 è scoppiata la prima epidemia di SARS (Severe Acute Respiratory Syndrome). Si trattava di una zoonosi generata dal virus SARS-CoV-1, un ceppo della specie SARS-CoV. È stata identificata per la prima volta in Cina e non si è mai diffusa molto al di fuori dei paesi circostanti e il Canada. Nel luglio 2003 è stato dichiarato sradicato dopo aver causato 8.096 casi e 774 morti.

with large epidemics mathematically was no longer a matter of demography alone, although that continued to be the main tool for estimating the number of cases and deaths. Statistical-mathematical models were employed to estimate the efficacy of antiviral drugs, for instance against HIV-infections, and the efficacy of various immunisations, including those against forms of influenza. Mathematical modelling of epidemics was used in planning strategies to eradicate smallpox, poliomyelitis, measles and maybe others. The first articles on modelling influenza epidemics appeared in the scientific literature. Planning a vaccination strategy involves both statistical-mathematical and mathematical models [5].

These roads to progress may have produced a general feeling of success in dealing with epidemics. Then in the period from 2002 to 2019, a few events occurred that evoked memories of previous pandemics and undermined such beliefs.

2. Unexpected Events 2002 - 2018

In November 2002 the first SARS (Severe Acute Respiratory Syndrome) epidemic broke out. It was a zoonosis generated by the virus SARS-CoV-1, a strain of the species SARS-CoV. It was first identified in China and never spread much beyond the surrounding countries and Canada. In July 2003 it was declared eradicated after having caused 8,096 cases and 774 deaths.

Allo stesso modo, la Sindrome Respiratoria del Medio Oriente (MERS), dovuta al corona-virus MERS-CoV, portò a circa 2.500 casi e 870 decessi tra il 2012 e il 2020. Fu concentrata essenzialmente in Arabia Saudita e, in misura minore, in Corea del Sud, con la maggior parte delle infezioni che si verificarono tra il 2014 e il 2015. Essendo una Zoonosi trasportata in gran parte dai cammelli, è anche chiamata l'influenza dei cammelli

Inoltre, si è diffusa nel mondo una pandemia di influenza che assomigliava all'influenza spagnola sotto diversi aspetti. Era un nuovo ceppo chiamato A(H₁N₁)09 del virus H₁N₁ dell'influenza. La sua origine è oggetto di dibattito; una probabile ipotesi dice che, essendo una zoonosi portata dai maiali, essa ha infettato un essere umano in un allevamento di maiali messicano intorno al gennaio 2009. È stata quindi chiamata influenza suina o influenza messicana. Si è diffusa dal Nord America a tutto il mondo e è stata dichiarata estinta nell'agosto 2010. Le stime del numero di infezioni e di decessi variano enormemente, ma apparentemente ci sono stati più casi e meno morti che per l'influenza spagnola. Sono state sollevate accuse contro l'OMS in merito alla sua gestione dei vaccini contro l'influenza suina.

Infine, è apparsa un'altra influenza zoonotica, popolarmente chiamata influenza aviaria e nel linguaggio scientifico HPAI (Highly Pathogenic Avian Influenza). Il principale agente patogeno era un virus influenzale A(H₅N₁). Era noto da molto tempo ma ha raggiunto un picco negli anni 2013-2017. Se vi sia stata una trasmissione per via aerea dal pollame all'uomo è stata una questione molto dibattuta, con ovvie implicazioni economiche. L'influenza aviaria si è diffusa ampiamente in tutto il mondo, ma il numero di casi umani conosciuti è rimasto basso, poco più di 70.

Oltre alle varie forme di influenza e alla epidemie generate dai corona-virus SARS-CoV-1, SARS-CoV-2 o MERS-CoV, si sono verificate altre epidemie.

È istruttivo confrontarle con quelle appena citate, applicando in aggiunta anche criteri matematici. Ci limiteremo alle epidemie di Ebola. La maggiore ondata epidemica è stata l'epidemia del virus Ebola nell'Africa occidentale dal 2013 al 2016, che ha causato 28.646 casi e 11.323 morti. Vi è una differenza fondamentale nell'evoluzione

Similarly, the Middle East Respiratory Syndrome (MERS), due to the corona virus MERS-CoV, lead to around 2,500 cases and 870 deaths between 2012 and 2020. It was essentially concentrated in Saudi Arabia and, to a minor degree, in South Korea, with most infections happening around the years 2014 to 2015. Being a zoonosis carried largely by camels, it is also called the camel flu.

Moreover, a pandemic influenza invaded the world that resembled the Spanish flu in several respects. Its pathogen was a new strain called A(H1N1)09 of the H1N1 influenza virus. Its origin is being debated; a likely hypothesis says that, being a zoonosis carried by pigs, it infected a human on a Mexican pig farm around January 2009. It was therefore called swine flu or Mexican flu. It spread from North America to the whole world and was declared extinguished in August 2010. Estimations of the number of infections and deaths fluctuate enormously, but there were apparently more cases and fewer deaths than by the Spanish flu. Accusations against the WHO were raised about its handling of vaccines against the swine flu.

Finally, another zoonotic influenza appeared, popularly called bird flu and in scientific language Highly Pathogenic Avian Influenza (HPAI). The main pathogen was an A(H₅N₁) influenza virus. It had been known long ago but reached a peak in the years 2013 - 2017. Whether there was airborne transmission from poultry to humans was a hotly debated question with obvious economic consequences. The bird flu spread widely over the whole world, but the number of known human cases remained small, at just over 70.

In addition to various forms of influenza and the epidemics generated by the corona virus SARS-CoV-1, SARS-CoV-2 or MERS-CoV, other epidemics occurred.

It is instructive to compare them with those just mentioned, also applying in addition mathematical yardsticks. We shall restrict ourselves to Ebola epidemics. Their most widespread outbreak was the Western African Ebola virus epidemic from 2013 to 2016, which caused 28,646 cases and 11,323 deaths. There is a fundamental

di un caso di influenza o di SARSCoV-1 o SARS-CoV-2 da un lato e di un caso di Ebola dall'altro, il che porta ad una differenza fondamentale nella loro modellizzazione matematica (sezioni 4 e 8). Un vettore del virus dell'influenza o di un corona-virus può trasmetterlo ad altri molto prima della comparsa dei primi sintomi, cioè molto prima della fine del periodo di incubazione. I soggetti infettati dal virus Ebola diventeranno contagiosi solo verso la fine del periodo di incubazione. Potrebbero quindi essere immediatamente isolati insieme ai loro ultimi contatti per evitare un'ulteriore trasmissione dell'infezione, a condizione che nella zona vi sia un servizio sanitario che lo faccia. Pertanto, il virus Ebola non si è diffuso in paesi che hanno una rete sanitaria primaria sufficientemente fitta, ma ha causato grande sofferenza in paesi che non l'hanno. La strategia dell'OMS per controllare l'epidemia era sbagliata. Ha insistito sui farmaci e la ricerca di un vaccino (che è diventato disponibile solo nel dicembre 2019), ma ha trascurato l'assistenza sanitaria di base. Allo scopo sarebbe stato ancora più utile, nei villaggi, formare rapidamente operatori sanitari e medici scalzi come era stato fatto decenni fa.

3. Uno sguardo ad alcuni paesi

Solo pochissimi paesi hanno tratto profitto dalle esperienze di questi 18 anni premonitori per prepararsi con molto anticipo ad una possibile, e probabile, nuova epidemia. Altri hanno adottato misure appropriate solo ai primi segni del Covid-19, e molti hanno iniziato a pianificare quando l'epidemia aveva quasi raggiunto il suo apice. Faremo alcuni esempi. Per semplicità descriveremo sempre il risultato della strategia di un paese indicando il numero cumulativo di casi confermati e morti intorno al 1^o giugno 2020. Per quanto riguarda l'affidabilità di questi dati, si vedano le Sezioni 5 e 6. Cominciamo con quelli che avevano pianificato per tempo.

Taiwan: Già nel 2004, l'anno dopo l'epidemia di SARS, il governo ha istituito il National Health Command Center (NHCC), che doveva preparare il Paese ad una possibile nuova epidemia. A partire dal 2017 era guidato dal popolare Ministro della Salute, Chen Shih-chung, che aveva studiato odontoiatria al Taipei Medical College. Il

difference in the evolution of a case of influenza or SARSCoV-1 or SARS-CoV-2 on the one hand and of an Ebola case on the other, which leads to a basic difference in their mathematical modelling (Sects. 4 and 8). A carrier of an influenza or corona virus can transmit it to others well before the first symptoms appear, that is well before the end of the incubation period. A subject infected by Ebola will only become infectious around the end of the incubation period. They could then be immediately isolated together with their latest contacts in order to avoid further transmission of the infection, provided that there is a health service nearby to do it. Therefore, Ebola did not spread to countries that have a sufficiently dense primary healthcare network, but it caused much suffering in countries that do not have it. The strategy of the WHO to control the epidemic was wrong. It insisted on drugs and the search for a vaccine (which only became available in December 2019), but neglected primary healthcare. For the present purpose it would even have been most useful to rapidly train village health workers and barefoot doctors as it had been done decades ago.

3. Looking at some countries

Only very few countries profited from the experiences of these premonitory 18 years to prepare much in advance for a possible, and probable, new outbreak of an epidemic. Some others only took appropriate measures at the first signs of Covid-19, and many started planning when the epidemic had almost reached its zenith. We shall sketch some examples. For simplicity we will always describe the result of the strategy of a country by indicating its cumulated numbers of confirmed cases and deaths around the 1 June 2020. Regarding the reliability of these data, see Sects. 5 and 6. We begin with those that had planned early.

Taiwan: Already in 2004, the year after the SARS-epidemic outbreak, the government established the National Health Command Center (NHCC), which was to prepare the country for a possible new epidemic. From 2017 on it was headed by the popular Minister of Health, Chen Shih-chung, who had studied dentistry at the

vicepresidente di Taiwan dal 2016 al 2020, Chen Chien-jen, era stato Ministro della Salute dal 2003 al 2005 dopo aver studiato genetica umana, sanità pubblica e epidemiologia all'Università Nazionale di Taiwan e all'Università Johns Hopkins negli Stati Uniti, ed essersi per un periodo dedicato alla ricerca. Così le decisioni sul controllo del Covid-19 sono state prese da politici competenti in materia di sanità, compresa quella pubblica.

Taiwan conta 23 milioni di abitanti, e molti di loro viaggiano da e per la Cina. Dal 31 dicembre 2019 in poi, quando l'OMS fu informata dell'epidemia a Wuhan, tutti i voli in arrivo da lì furono controllati e furono seguiti da controlli sui passeggeri in arrivo da qualsiasi altro posto. Un Tavolo d'azione fu istituito nel periodo dal 20 gennaio al 24 febbraio 2020: elencava 124 misure da adottare. Ogni giorno il pubblico riceveva informazioni chiare attraverso tutti i mezzi esistenti. Sulla base della tessera elettronica dell'assicurazione sanitaria che tutti i cittadini hanno fu rapidamente istituito il Tracciamento di Contatto, che significa il *follow-up* ripetuto delle persone sintomatiche, dei casi confermati e di tutti i loro contatti. I test virologici PCR utilizzati (Sez. 4) erano già disponibili e si dette inizio a quarantene ben organizzate. A fine gennaio, si pubblicarono le regole sull'uso delle mascherine; ne esisteva già una scorta sufficiente.

Come risultato sono stati trovati 442 casi confermati e registrati 7 decessi sino al 1^o giugno.

Vietnam: La strategia vietnamita assomiglia a quella taiwanese in quasi tutti gli aspetti, con l'eccezione della ricerca di contatti. Un comitato direttivo con cui trattare nuove epidemie esisteva al Ministero della Sanità. Mise in atto il suo piano subito dopo il 23 gennaio, quando i primi individui infetti arrivarono negli aeroporti vietnamiti, tra i quali un vietnamita di ritorno dal Regno Unito. Tutte le scuole furono chiuse il 25 gennaio, e dal 1^o febbraio tutti quelli che entravano in Vietnam dovevano passare due settimane in quarantena.

Altre misure sono state imposte o alleggerite in base all'evoluzione dell'epidemia, ad esempio un confinamento limitato o l'uso di mascherine. Il Ministero della Sanità ha rilasciato regolarmente informazioni precise e chiare per l'intera popolazione con tutti i mezzi disponibili, compresi

Taipei Medical College. The Vice-President of Taiwan from 2016 to 2020, Chen Chien-jen, had been Minister of Health from 2003 to 2005 after having studied human genetics, public health, and epidemiology at the National Taiwan University University and the Johns Hopkins University in the USA, followed by research. Thus, decisions about the control of Covid-19 were taken by politicians competent in matters of health, including public health.

Taiwan counts 23 million inhabitants, and many of them travel to and from China. From the 31 December 2019 onwards, when the WHO was notified of the epidemic in Wuhan, all incoming flights from there were checked, followed by controls of passengers arriving from anywhere else. An Action Table was produced in the period of 20 January to 24 February 2020, which listed 124 measures to be taken. The public obtained daily revised clear information by all existing means. "Contact tracing", which means repeated follow-up of symptomatic persons, of confirmed cases and of all of their contacts, was rapidly established on the basis of the electronic health insurance card that everybody has. The virological PCR-tests used (Sect. 4) were already available and quarantines well organised. In late January, rules about the wearing of masks were edited; a sufficient supply existed already.

As a result, 442 confirmed cases had been found and 7 deaths recorded up to the 1 June.

Vietnam: The Vietnamese strategy resembles the Taiwanese one in almost all aspects, with the exception of contact tracing. A Steering Committee to deal with new epidemics existed in the Ministry of Health. It put into effect its plan right after the 23 January, when the first infected individuals arrived at Vietnamese airports, among them a Vietnamese returning from the UK. All schools were closed on the 25 January, and since the 1 February everybody entering Vietnam had to spend two weeks in quarantine.

Other measures were imposed or relieved in accordance with the evolution of the epidemic, for instance a limited confinement or the wearing of masks. The Ministry of Health issued regular precise and clear information for the entire population by all available means, including smart-

gli smartphone. Inoltre, è stato fornito a tutti un sistema informativo personalizzato sulle cosiddette guide alla sopravvivenza. Ogni guida alla sopravvivenza definisce tre categorie di persone: F0: un caso confermato; F1: sospettato di essere infetto o di aver avuto contatti con una persona infetta; F2: aver avuto contatti con una persona in F1. Ci si aspetta che ogni persona determini la categoria alla quale appartiene. La guida alla sopravvivenza fornisce poi informazioni stampate su cosa si debba fare a seconda della categoria di appartenenza, ad esempio sottoporsi a un test. Sono utilizzati solo test PCR.

A differenza di Taiwan, la ricerca di contatti non utilizza strumenti elettronici. Lo fa la popolazione stessa, aiutata dalle guide alla sopravvivenza, insieme a un gran numero di membri ben addestrati dei servizi sanitari, ad esempio docenti universitari.

Alla fine del 2019, il Vietnam contava 98.257.747 abitanti. Al 1^o giugno c'erano stati 328 casi confermati e 0 morti. Questi dati si basano su una forte sezione demografica dell'Ufficio statistico generale e su diversi sistemi informativi sanitari [6] e può difficilmente essere contestato.

La breve descrizione delle misure di controllo a Taiwan e in Vietnam ha mostrato le tre componenti principali del loro lato epidemiologico: tracciamento dei contatti; *lockdown*, che è la distanza fisica o sociale in senso lato compresi i controlli di quarantena e di frontiera; e indossare le mascherine. Possiamo chiamarlo strategia di sorveglianza e contenimento. A ciò si aggiunge il lato medico-clinico, dall'assistenza sanitaria di base, come i medici di base, fino ai grandi ospedali. Il suo stato è cruciale per il numero di morti causate dal virus SARS-CoV-2.

A differenza di Taiwan e del Vietnam, sembra che tutti gli altri paesi del mondo fossero impreparati alla fine del dicembre 2019. Alcuni di loro hanno adottato misure abbastanza sistematiche e rigorose che coprivano l'intera popolazione non appena si sono manifestati i primi casi. Questo è stato per esempio vero per la **Cina** alla fine di gennaio 2020, per la **Slovacchia** e la **Grecia** il 27 e 28 febbraio, per l'**Austria** il 10 marzo e per la **Danimarca** il 12 marzo. Una strategia danese alternativa, basata su una rigorosa rintracciabilità dei contatti e sulla quarantena, ma sinora non messa in atto fu propugnata in [7].

phones. In addition, there is a personalised information system of so-called survival guides given to everybody. Every survival guide defines three categories of persons: F0: a confirmed case; F1: suspected to be infected or having had contact with an infected person; F2: having had contact with a person in F1. Each person is expected to find the category to which it belongs. The survival guide then provides printed information about what it must do depending on its category, for example submit to a test. Only PCR-tests are used.

In contrast to Taiwan, contact tracing does not use electronic tools. It is done by the population itself, aided by the survival guides, together with a large number of well-trained members of the health services, for example university lecturers.

At the end of 2019, Vietnam had 98,257,747 inhabitants. On the 1 June there had been 328 confirmed cases and 0 deaths. These data are based on a strong demographic section of the General Statistical Office and on several Health Information Systems [6] and can hardly be contested.

The preceding sketch of control measures in Taiwan and Vietnam has shown us the three main components of their epidemiologic side: contact tracing; lockdown, that is physical or social distancing in the wide sense including quarantine and border controls; and wearing of masks. We may call this the surveillance-containment strategy. In addition, there is the medical-clinical side, from primary healthcare such as general practitioners up to large hospitals. Its state is crucial to the number of deaths caused by the virus SARS-CoV-2.

In contrast to Taiwan and Vietnam, it seems that all other countries in the world were unprepared at the end of December 2019. A few of them took fairly systematic and strict measures that covered the entire population as soon as the first cases had declared themselves. This was for example true for China at the end of January 2020, for Slovakia and Greece on the 27 and 28 February, for Austria on the 10 March and for Denmark on the 12 March. An alternative Danish strategy, based on rigorous contact tracing and quarantine, but not implemented until now was argued for in [7].

Per quanto riguarda i risultati, l'evoluzione turbolenta in Cina è ben nota. In Danimarca, con una popolazione di 5.806 milioni, sono stati confermati circa 12000 casi e 593 decessi registrati; le cifre corrispondenti per l'Austria sono state di 8.86 milioni di persone, 16979 casi e 672 decessi.

Il confronto della Slovacchia, un paese di circa 5.5 milioni di abitanti, con la Grecia, che conta 10.72 milioni di persone, colpisce particolarmente perché rende visibile il ruolo di medici e ospedali. In Slovacchia ci sono stati 1528 casi confermati e 28 decessi. I dati corrispondenti per la Grecia sono stati 3058 e 183. Il numero relativamente molto più alto di vittime in Grecia, nonostante la reazione altrettanto precoce e quasi lo stesso numero di casi per numero di abitanti, è senza dubbio dovuto allo stato catastrofico del suo sistema sanitario, causato principalmente dalla crisi del debito a partire dal 2010.

Ci spostiamo quindi a un gruppo di paesi che hanno reagito tardi e non sistematicamente, applicando le varie misure in modo disordinato e solo a una parte della popolazione. Ecco alcuni di loro con il loro numero di abitanti in milioni, il numero cumulativo di casi confermati e il numero di vittime:

Belgio: 11.46; 59348; 9606.

Spagna: 46.94; 289046; 27136.

Italia: 60.36; 235561; 34043.

Francia: 66.99; 154591; 29296.

Germania: 83.02; 187000; 8831.

Il numero relativamente basso di morti in Germania riflette principalmente un sistema sanitario efficiente che ha saputo adattarsi facilmente all'epidemia. L'opposto è stato vero in Francia. Lì, circa 100000 letti d'ospedale erano stati eliminati nel periodo tra il 1993 e il 2018. Un arbitrario e rigoroso confinamento, non determinato da un ragionamento epidemiologico, è stato imposto il 17 marzo.

Infine, vi sono paesi che hanno deciso di non fare nulla, almeno per un lungo periodo. La loro motivazione, o pretesto, era soprattutto la fiducia nell'immunità di gregge (Sezioni 1 e 8), secondo cui l'epidemia si sarebbe fermata da sola. Questa era la strategia della **Svezia**, un paese che contava 10.23 milioni di persone, che ha portato a 37814 casi e 4403 morti. Nel **Regno Unito** ci sono stati,

Regarding the results, the turbulent evolution in China is well known. In Denmark, with a population of 5.806 million, about 12,000 cases had been confirmed and 593 deaths recorded, and the corresponding figures for Austria were 8.86 million people, 16,979 cases and 672 deaths.

The comparison of Slovakia, a country of around 5.5 million inhabitants, with Greece, which counts 10.72 million people, is particularly striking because it makes visible the role of their physicians and hospitals. In Slovakia there were 1,528 confirmed cases and 28 deaths. The corresponding data for Greece were 3,058 and 183. The relatively much higher number of fatalities in Greece, in spite of equally early reaction and almost the same number of cases per number of inhabitants, is no doubt due to the catastrophic state of its medical-clinical system, caused mainly by the debt crisis from 2010 onwards.

Next, we move to a group of countries that reacted late and not systematically, applying the various measures in a haphazard way and only to part of the population. Here are some of them with their numbers of inhabitants in millions, cumulated numbers of confirmed cases and numbers of fatalities:

Belgium: 11.46; 59,348; 9,606.

Spain: 46.94; 289,046; 27,136.

Italy: 60.36; 235,561; 34,043.

France: 66.99; 154,591; 29,296.

Germany: 83.02; 187,000; 8,831.

The relatively low number of deaths in Germany mainly reflects a sufficient medical-clinical system that could readily adapt itself to the epidemic. The opposite was true in France. There, about 100,000 hospital beds had been eliminated in the period between 1993 and 2018.

An arbitrary strict confinement, not determined by epidemiologic reasoning, was imposed on the 17 March. Finally, there are countries that decided to do nothing, at least for a long while. Their motivation, or pretext was above all a belief in herd immunity (Sects. 1 and 8), according to which the epidemic would stop by itself. This was the strategy of Sweden, a country counting 10.23 million people, which resulted in 37,814 cases and 4,403 deaths. In the United Kingdom there were, among 66.65 million inhabitants, around 290,000 cases and 41,128 fatalities, and in the

tra i 66.65 milioni di abitanti, circa 290000 casi e 41128 morti e negli **Stati Uniti** questi dati sono stati 328.2 milioni, 2.04 milioni, 115000 morti.

Questa panoramica delle strategie conferma che, come detto nell'introduzione, i risultati in effetti dipendono fortemente dallo stato della sanità pubblica. Si noti che al giorno d'oggi in ogni lingua del mondo il concetto di sanità pubblica è designato con una traduzione letterale o una leggera modifica di questa espressione. Per esempio, in danese è *folkesundhed*, cioè salute del popolo.

Dall'alto verso il basso

In questa seconda parte, descriveremo la parte scientifica e in particolare i principi matematici coinvolti nello studio delle fasi successive delle pandemie. In breve: Sez. 4: Scoperta del nuovo virus, proprietà di base, test per la sua presenza in una persona. 5 e 6: Dati sull'evoluzione del Covid-19 in una popolazione. 7: Tentativi di analizzare matematicamente e prevedere tale evoluzione rappresentandola con una curva epidemica. 8: L'analogo per una rappresentazione da parte di un modello compartimentale. 9: Provare a fermare l'epidemia con un vaccino. 10: Cosa imparare e cosa fare?

4. Il nuovo virus SARS-CoV-2

Dopo il primo focolaio, alla fine di dicembre 2019, spesso descritto, di casi di polmonite di eziologia sconosciuta intorno a Wuhan, nel corso del gennaio 2020 gli scienziati cinesi hanno identificato un nuovo virus come agente patogeno. Essi hanno seguito le procedure abituali, cioè hanno determinato il carico di 26 agenti patogeni respiratori comuni nei pazienti. Non ne hanno trovato nessuno in abbondanza. Sospettavano che si trattasse di SARS-CoV, ma anche qui non sono riusciti a trovarlo. Hanno poi indagato su tutti i tipi di carica virale che avevano una leggera somiglianza (coincidenza in alcuni genomi) con la SARS-CoV e hanno rilevato un nuovo virus che ha mostrato un'abbondanza di virioni nei campioni respiratori dei pazienti. Microscopia elettronica e analisi di modelli matematici [8], [9] hanno mostrato che appartiene alla stessa specie

USA these data were 328.2 million, 2.04 million, 115,000 deaths.

This overview of strategies confirms that, as said in the introduction, the results indeed depend heavily on the state of public health. Note that nowadays in every language of the world the concept public health is designated by a literal translation or a slight modification of this expression. For instance, in Danish it is *folkesundhed* that is, health of the people.

Top-down

In this second part, we will sketch the scientific and in particular mathematical principles involved in the study of successive stages of the pandemics. In short: Sect. 4: Discovery of the new virus, basic properties, testing for its presence in a person. 5 and 6: Data on the evolution of Covid-19 in a population. 7: Attempts at analysing mathematically and predicting such an evolution by representing it by an epidemic curve. 8: The analogous for a representation by a compartmental model. 9: Trying to stop the epidemic by a vaccine. 10: What to learn and what to do?

4. The New Virus SARS-CoV-2

After the often-depicted outbreak in late December 2019 of cases of pneumonia of unknown aetiology around Wuhan, in the course of January 2020 Chinese scientists identified a new virus as the pathogen. They followed the usual procedures, i.e., they determined the load of 26 common respiratory pathogens in the patients. They found none of them in abundance. They suspected SARS-CoV, but could not find it either. Then they investigated all kinds of viral load that had a slight similarity (coincidence in a number of genomes) with SARS-CoV and detected a novel virus which displayed abundant virions in respiratory specimens from patients. Electron microscopy and mathematical pattern analysis [8], [9] showed that it belongs to the same species as SARS-CoV-1 and MERS-CoV (Sect. 2); hence the name SARS-CoV-2.

di SARS-CoV-1 e MERS-CoV (Sez. 2); da qui il nome SARS-CoV-2.

A partire da questo lavoro in Cina, è apparso un gran numero di pubblicazioni sulle proprietà peculiari dell'agente patogeno e sui modi in cui agisce. Dal punto di vista virologico, è stata determinata la sua sequenza genetica. Si ritiene che il nuovo virus abbia origini zoonotiche, ma si è rapidamente stabilito che vi era contagi da uomo a uomo. La combinazione di caratteristiche della SARS e dell'influenza, che è intensa inibizione respiratoria dei pazienti e una trasmissione rapida, rendono il Covid-19, la malattia causata dalla SARS-CoV-2, particolarmente pericolosa. Per ulteriori lavori si veda [10].

Nel contesto clinico, sono stati determinati diversi periodi nell'evoluzione di un caso (vedi la loro definizione in [1], Sez. 5.2): Il periodo d'incubazione mediano è di 5,2 giorni; il periodo di latenza medio è di 4,6 giorni, cioè, in generale il periodo infettivo inizia infatti prima della fase prodromica. Abbiamo discusso le implicazioni nella Sez. 2 nel confronto con Ebola. La durata media del periodo infettivo è di 6 giorni per i casi lievi e asintomatici; per i casi gravi e critici questo periodo dura in media 22 giorni e termina solo con la guarigione o la morte.

Le molteplici applicazioni per il controllo della pandemia sia delle loro caratteristiche virologiche sia di quelle cliniche appariranno nelle Sezioni 7, 8, 9 e 10. Il loro studio è ancora attivo e può persino ribaltare i risultati precedenti; questo è accaduto di recente, per esempio, nelle cosiddette immunità incrociate. Tuttavia, in questo articolo tratteremo solo applicazioni all'elemento fondamentale delle strategie di controllo ben progettate, ovvero i test per le infezioni.

Il primo passo di un programma di test consiste nel definire la popolazione di interesse. Chi sarà sottoposto al test? I soggetti entrati in contatto con persone infette? O chi si lamenta dei propri sintomi? O tutti quelli che provengono da una regione dove esistono casi? Vedi l'esempio del Vietnam nella Sez. 3.

E poi, quale sarà lo scopo? Scoprire il presenza del virus o di qualche tipo di anticorpo? A seconda dello scopo ci sono test virologici e test sierologici. Il consueto test virologico si chiama test PCR-(*Polymerase Chain Reaction*). Decine di test sierologici di qualità variabile sono stati, e

Starting with this work in China, a large number of publications about the peculiar properties of the pathogen and the ways it is acting have appeared. On the virological side, its genetic sequence was determined. The new virus is believed to have zoonotic origins, but human to human infection was rapidly established. The combination of SARS and influenza features, that is intensive respiratory inhibition of patients and rapid transmission, make Covid-19, the disease caused by SARS-CoV-2, particularly dangerous. For further work see [10].

In the clinical context, several periods in the evolution of a case were determined (see their definition in [1], Sect. 5.2): The median incubation period is 5.2 days; the mean latency period is 4.6 days, i.e., in general the infectious period indeed starts before the prodromal phase. We have discussed the implications in Sect. 2 in comparison with Ebola. The mean length of the infectious period is 6 days for mild and asymptomatic cases; for severe and critical cases this period lasts on average 22 days and ends only by recovery or death.

The manifold applications for the control of the pandemic of both their virological and their clinical characteristics will appear in Sects. 7, 8, 9 and 10. Their study is still active and may even reverse former results; this happened recently for example about so-called crossimmunities. However, in this article we will only treat applications to the basic element of well-designed control strategies, namely testing for infections.

The first step of a test programme is to define the target population. Who will be tested? Subjects who had contact with infected people? Or those who complain about symptoms? Or everybody coming from a region where cases exist? See the example of Vietnam in Sect. 3.

Next, what will be the objective? To discover the presence of the virus or that of some kind of antibodies? Depending on the objective there are virological and serological tests. The usual virological test is called the PCR-(*Polymerase Chain Reaction*) Test. Dozens of serological tests of vary-

sono tuttora, sviluppati e in alcuni paesi persino offerti al grande pubblico. Ricordiamo che la caratterizzazione di un test con una popolazione di interesse e un dato scopo è un argomento classico di epidemiologia clinica (cfr. [1], Sect. 19.2).

Tornando al ruolo fondamentale del *test* nelle strategie di controllo, osserviamo solo che nei paesi poveri o nei paesi ricchi con funzionari della sanità pubblica disattenti, la popolazione di interesse era spesso determinata dalla carenza di kit di prova e dall'influenza delle istituzioni che li richiedeva per sè.

5. Demografia: Epidemiologia descrittiva

È questa la statistica medica classica, che fornisce per una specifica malattia il numero di casi e di morti insieme a il quando e il dove e ad alcuni dati aggiuntivi come sesso, età e, talvolta, professione dei soggetti. Nel contesto attuale abbiamo a che fare con il Covid-19 come la malattia dovuta a un'infezione da virus SARS-CoV-2.

In linea di principio, i metodi per trovare il numero di i casi confermati e i decessi causati dal Covid-19 sono gli stessi di qualsiasi altra malattia. Essi oscillano ampiamente tra paesi. Sia la diagnosi di un caso di malattia sia la descrizione della causa del decesso può essere relativamente corretta o molto inaffidabile. In particolare, trovare una diagnosi corretta per qualcuno che si lamenta di seri problemi di salute dipende molto dai metodi locali di tracciamento dei contatti e dallo stato del sistema clinico-medico. Un'altra difficoltà deriva dall'esistenza di forme asintomatiche della malattia, cioè soggetti infettati dal SARS-CoV-2 che non mostrano sintomi.

Nella Sez. 3 abbiamo citato il Vietnam, che utilizza i suoi normali sistemi informativi demografici e sanitari [6]. Include nelle sue statistiche casi asintomatici trovati tramite il tracciamento dei contatti. Altri paesi ottengono il loro dati di morbosità e mortalità da un sistema che riporta lo stato della sanità. Un tale sistema è in parte basato su metodi di campionamento da varie fonti, per esempio ospedali e uffici sanitari locali. In Germania, l'Istituto Robert Koch, un istituto centrale dedicato principalmente alle malattie infettive, riferisce sui risultati sul Covid-19. Negli

ing quality have been and still are being developed and even offered to the general public in some countries. Recall that the characterisation of a test with a given target population and a given objective is a classical subject of clinical epidemiology ([1], Sect. 19.2.)

Coming back to the fundamental role of testing in control strategies, we only remark that in poor countries or in rich countries with inattentive public health officials, the target population was often determined by the shortage of test kits and by the influence of institutions that required them for themselves.

5. Demography: Descriptive Epidemiology

This is classical medical statistics, which gives for a specific disease the number of cases and deaths together with the when and where and a few additional data such as sex, age and sometimes profession of the subjects. In the present context we are dealing with Covid-19 as the disease due to an infection by the virus SARS-CoV-2.

In principle, the methods for finding the number of confirmed cases and fatalities by Covid-19 are the same as for any other disease. They fluctuate widely between countries. Both the diagnosis of a case of a disease and the description of the cause of a death may be relatively correct or most unreliable. In particular, finding a correct diagnosis for somebody who complains about acute health problems depends very much on the local contact tracing methods and on the state of the clinical-medical system. An additional difficulty arises from the existence of asymptomatic forms of the disease, that is, subjects infected by SARS-CoV-2 who display no symptoms.

In Sect. 3 we have mentioned Vietnam, which uses its normal demographic and health information systems [6]. It includes in its statistics asymptomatic cases found by contact tracing. Other countries obtain their morbidity and mortality data from a health reporting system. Such a system is partly based on sampling methods from various sources, for example hospitals and local health offices. In Germany, the Robert Koch Institute, a central institute mainly devoted to infectious diseases, reports on the results for Covid-19. In the USA, the Johns Hopkins University plays

Stati Uniti, la Johns Hopkins University svolge un ruolo simile. Anche altri paesi utilizzano i dati degli uffici dell'assicurazione sanitaria.

Tuttavia, molti paesi non dispongono né di un sistema di informazione sanitaria né di un sistema di segnalazione sanitaria, o non lo usano per il Covid-19. Si usa un insieme di metodi alternativi. Ad esempio, la Francia conta solo i casi confermati di ricovero in ospedale e solo i decessi che si verificano in ospedale o in una residenza per anziani collegata a una struttura medica.

Riassumendo, possiamo dire che i dati sulla morbosità, e in misura minore i dati sulla mortalità, per il Covid-19 che si trovano in varie pubblicazioni periodiche sono abbastanza inaffidabili, con pochissime eccezioni. Le fonti non sono sempre chiaramente indicate.

Un'importante idea alternativa è di confrontare la situazione attuale con quella degli anni passati. Parlando in maniera ingenua anche in questo caso, si presume che il più alto numero di casi attuali e il numero di morti, e solo questi, sono il risultato del Covid-19. Date le difficoltà diagnostiche sopra menzionate, questa idea si applica principalmente ai decessi e difficilmente ai casi non letali. Così, nel metodo dell'eccesso di mortalità, misuriamo solo quanti decessi in più per qualsiasi causa sono avvenuti quest'anno rispetto al periodo corrispondente nel passato. Per il Regno Unito, ad esempio, abbiamo citato nella Sez. 3 la cifra di 41128 decessi fino al 1^o giugno come fornito dal Servizio Sanitario Nazionale. Per contro, l'Ufficio Nazionale di Statistica ha innalzato a circa 62000 decessi la mortalità in eccesso!

Infine, ecco un'idea interessante basata sulla più classica forma di modello statistico. Un grafico nell'articolo [11] mostra in doppia scala logaritmica per ognuno dei 16 paesi selezionati il punto del piano le cui coordinate sono, rispettivamente, il numero stimato di infezioni per milione di abitanti nel giorno del *lockdown* e l'eccesso di mortalità. Un veloce occhiata ci convince che sono correlati positivamente. Una semplice analisi di regressione basata su questo grafico permetterebbe anche di stimare uno di questi valori in funzione dell'altro per qualsiasi altro paese.

6. Demografia avanzata

a similar role. Still other countries use data from health insurance offices.

However, many countries have neither a health information system nor a health reporting system, or they do not use it for Covid-19. A host of alternative methods is being employed. For example, France only counts hospitalised confirmed cases and only deaths which happen in a hospital or in a retirement home that is connected with a medical structure.

Summing up, we may say that morbidity data, and to a lesser degree mortality data, for Covid-19 that one finds in various periodic publications are fairly unreliable, with very few exceptions. The sources are not always clearly indicated.

An important alternative idea is to compare the present situation with that in years past. Speaking naïvely again, we assume that the present higher case frequencies and death toll, and only these, are the result of Covid-19. Given the diagnostic difficulties mentioned above, this idea is mainly applied to fatalities and hardly to non-lethal cases. Thus, in the method of excess mortality we only measure how many more deaths by any cause happened this year than in the corresponding period in the past. For the UK we have for instance quoted in Sect. 3 the figure of 41,128 deaths up to the 1 June as supplied by the National Health Service. By contrast, the National Statistical Office advanced about 62,000 deaths as excess mortality!

Finally, here is an interesting idea based on the most classical form of a statistical-mathematical model. A graphic in the paper [11] shows in double logarithmic scale for every one of 16 selected countries the point in the plane whose coordinates are, respectively, the estimated number of infections per million inhabitants on lock-down day, and the excess mortality. A short glance convinces us that they are positively correlated. A simple regression analysis based on this graphic would also allow us to estimate one of these values by the other one for any other country.

6. Advanced Demography

Questa procede in diverse direzioni al di là delle classiche statistiche sanitarie, tutte ugualmente rilevanti per il Covid-19. In primo luogo, si conducono indagini a campione invece di utilizzare i dati da tutta la popolazione di interesse. Queste indagini sono state, ad esempio, utilizzate per studiare l'influenza dei fattori sociali sull'evoluzione di vari aspetti della malattia. In particolare in alcuni Paesi è stato indagato a fondo il fattore essere un immigrato o discendere da immigrati. In secondo luogo, si raccolgono più tipi di dati sui casi e sui decessi, per esempio sulla morbosità e la mortalità per gruppi di età. In terzo luogo, l'insieme dei dati non solo è registrato, e forse pubblicato, ma anche trasformato e interpretato in vari modi. Qui la standardizzazione è la procedura più conosciuta. Un esempio fittizio sarebbe il numero di vittime da Covid-19 in Danimarca, se la Danimarca avesse la stessa struttura di età del Vietnam e se in ciascun gruppo d'età avesse avuto la stessa mortalità da Covid-19 della stessa fascia d'età in Vietnam.

Nella Sez. 9 incontreremo i modelli statistici come strumento matematico di base per sviluppare un trattamento preventivo del Covid-19. Con il loro aiuto, si studia l'influenza di vari fattori su alcune variabili E di interesse in uno studio clinico. Svolge qui un ruolo l'idea di controllo per l'influenza di un altro fattore, che potrebbe essere un elemento di disturbo nello studio dell'azione di E . Sembra che la maggior parte dei demografi da un lato, e la maggior parte degli epidemiologi clinici, dall'altro, ignorino che la procedura matematica di standardizzazione è la stessa di quella di controllo per un elemento di disturbo (Ref. [1], Lezione 21). Però un matematico non si stupirà!

7. Modellando la curva epidemica

Abbiamo accennato a questo concetto classico nella Sez. 1; (cfr. [1], Sez. 4.6). Sia C un'epidemia, V una regione geografica, t_0 un istante di tempo che può essere quello del primo caso di C in V , e $f(t)$ per $t \geq t_0$ il numero di casi osservati e riportati di C che si erano manifestati in V prima o all'istante t . Quindi si chiama f curva epidemica di C in V . In particolare, occorre dire se siano inclusi o no i casi non confermati. Misurare $f(t)$ al variare del tempo t è il compito

It goes in several directions beyond classical health statistics, all of them likewise relevant to Covid-19. Firstly, sample surveys are conducted instead of using the data from the entire target population. They have for example been used to study the influence of social factors on the evolution of various aspects of the disease. In particular the factor to be an immigrant or to descend from them was thoroughly investigated in some countries. Secondly, more types of data about cases and deaths are collected, for example about morbidity and mortality by age groups. Thirdly, data sets are not only being registered and perhaps published, but also transformed and interpreted in various ways. Here, standardisation is the best-known procedure. A fictitious example would be the number of fatalities by Covid-19 in Denmark, if Denmark had the same age structure as Vietnam and in each age group it had the same Covid-19 mortality as in the same age group in Vietnam.

In Sect. 9 we shall meet statistical-mathematical models as a basic mathematical tool in developing a preventive treatment of Covid-19. With their help, one studies the influence of various factors on some outcome variable E of interest in a clinical trial. Here the idea of controlling for the influence of another factor, which might be a confounder in the study of the action of E , plays a role. It looks as if most demographers on the one hand, and most clinical epidemiologists on the other, ignore that the mathematical procedure of standardising is the same as that of controlling for a confounder (Ref. [1], Lesson 21). A mathematician will not be astonished, though!

7. Modelling the Epidemic Curve

We have mentioned this classical concept in Sect. 1; (see [1], Sect. 4.6). Let C be an epidemic, V a geographical region, t_0 a moment of time which may be that of the first case of C in V , and $f(t)$ for $t \geq t_0$ the number of observed and reported cases of C that had declared themselves in V before or at the instant t . Then f is called the epidemic curve of C in V . In particular, it needs to be said whether unconfirmed cases are included or not. Measuring $f(t)$ as the time t goes along is the

dei servizi demografici competenti (Sezioni 5 e 6). Questo processo è quindi soggetto a tutte le carenze elencate sopra.

Ottenere qualche informazione su f per le varie regioni V è naturalmente una delle principali preoccupazioni della popolazione di un paese colpito da C . Tale informazione è altrettanto vitale per le autorità sanitarie che cercano di controllare C . Tuttavia, è auspicabile una conoscenza molto maggiore. Cosa possiamo apprendere sui meccanismi di C osservando $f(t)$? Questo era già stato oggetto dei lavori descritti in [3]; si veda la Sez. 1. In particolare, vi è un modo per prevedere gli aspetti dell'evoluzione futura di f , avendo osservato i valori $f(t)$ per un certo periodo di tempo?

A queste domande si risponde generalmente con una modellizzazione di f , cioè facendo alcune ipotesi sulla sua forma e stimando alcuni parametri in essa contenuti. È stato pubblicato un gran numero di articoli su questo problema. Alcuni di questi usano metodi di estrapolazione noti dall'economia matematica. Una rassegna recente di varie idee e tecniche di base si possono trovare in [12], dove si descrive un modello mediante un'equazione integro-differenziale.

Ci limiteremo a discutere di un'applicazione, ovvero il cosiddetto numero di riproduzione di base R_0 . Questo compare costantemente nelle pubblicazioni divulgative. Per definirlo, consideriamo un soggetto s che risulta infetto in un momento $t^* \geq t_0$. Sia $\mu(s, t^*)$ il numero di tutti i soggetti infettati da s dopo il tempo t^* sotto forma di infezioni secondarie, terziarie, ecc. Quindi R_0 è la media di $\mu(s, t^*)$ su tutti i soggetti s . Così dipende da t^* . Proprio a questa dipendenza è interessata la gente: un valore inferiore a 1 è considerato come una previsione dell'estinzione di C dopo t^* . Nel caso $C = \text{Covid-19}$, erano stati stimati valori fino a 5.7, nella fase iniziale, vale a dire t^* vicino al momento dell'inizio dell'epidemia di C . L'articolo [7] presenta un'interessante fattorizzazione di R_0 per confrontare i diversi approcci al controllo della sua grandezza.

8. Modelli compartimentali

Abbiamo accennato alla loro origine storica nella Sez. 1. Abbiamo distinto due maniere di modellare matematicamente l'evoluzione di un'epide-

task of the relevant demographic services (Sects. 5 and 6). This process is therefore subject to all the deficiencies listed there.

To get some knowledge about f for various regions V is of course one of the main concerns of the population of a country invaded by C . Such knowledge is equally vital for health authorities who attempt to control C . However, much more knowledge is desirable. What can we learn about the mechanism of C by observing $f(t)$? This was already the subject of the papers described in [3]; see Sect. 1. In particular, is there a way to predict aspects of the future evolution of f , having observed the values $f(t)$ for a while?

Answers to these questions are generally given by modelling f , that is by making certain assumptions about its shape and by estimating certain parameters in it. A very large number of papers have been published about this issue. Some of them use extrapolation methods known from mathematical economy. A recent survey on various basic ideas and techniques can be found in [12], where a model is described in terms of an integro-differential equation.

We will restrict ourselves to a discussion of an application, namely a so-called basic reproduction number R_0 . It appears constantly in popular publications. To define it, let us look at a subject s that is infected at a time $t^* \geq t_0$. Let $\mu(s, t^*)$ be the number of all subjects infected by s after t^* in the form of secondary, tertiary etc. infections. Then R_0 is the average of $\mu(s, t^*)$ over all s . Thus, it depends on t^* . It is precisely this dependence in which people are interested: a value less than 1 is looked upon as predictor of the extinction of C after t^* . In the case $C = \text{Covid-19}$, values as high as 5.7 had been estimated in the beginning, that is, for t^* close to the time of the first outbreak of C . The article [7] presents an interesting factorisation of R_0 in order to compare different approaches to control the size of it.

8. Compartmental Models

We have sketched their historical origin in Sect. 1. We distinguished between two ways of mathematically modelling the evolution of an epidemic.

mia. I modelli del primo tipo (Sez. 7) rappresentano l'evoluzione temporale del numero di soggetti in un certo stato, per esempio lo stato essere infetti. Al contrario, i modelli compartimentali rappresentano anche i *cambiamenti* di questo stato in alcuni momenti sotto la forma di transizioni di un soggetto da un comparto ad un altro.

Il modello SIR, che abbiamo indicato nella Sez. 1 come "intensamente studiato negli anni '20 del secolo scorso", è particolarmente semplice ed è servito per in confronto con molti altri, in particolare per quelli applicati a Covid-19. Si tratta di tre comparti: S sono i soggetti suscettibili, non ancora infetti, I quelli infetti, e R consiste di soggetti rimossi o per la guarigione con immunità o per la morte. Le transizioni tra i comparti sono descritte da equazioni differenziali per le variabili $S(t)$, $I(t)$ e $R(t)$ che danno il numero dei soggetti nei comparti in funzione del tempo t . In queste equazioni sono presenti alcuni parametri come le probabilità di transizione da un comparto all'altro. Sotto varie ipotesi, il sistema risultante delle equazioni differenziali per S , I e R può essere risolto in modo esplicito o numericamente.

Una prima importante applicazione è quella di stimare un numero di riproduzione R_0 come definito nella Sez. 7. Può essere espresso per mezzo dei parametri di base.

In secondo luogo, si scopre che il limite S_∞ di $S(t)$ per $t \rightarrow \infty$ è strettamente positivo, il che significa che una certa parte della popolazione non sarà mai contagiata. Questo ha portato al concetto di immunità di gregge, che, tuttavia, ha dato origine a molta confusione tra le persone che pensavano di avere qualcosa da dire sulla questione.

Dopo l'inizio dell'epidemia di Covid-19, sono stati definiti e analizzati molti altri modelli compartimentali più complessi. I loro parametri hanno rappresentato, tra le altre caratteristiche, la strategia di controllo sottostante da utilizzare. Vi era ad esempio la strategia del non fare niente e anche la strategia della mitigazione, che consisteva nelle componenti meno stringenti della strategia di sorveglianza e contenimento definita nella Sez. 3. Nel molto discusso articolo [13], Neil Ferguson e i collaboratori hanno descritto la forma della funzione I , che è il numero di soggetti infetti nella strategia del "non fare niente". A partire dal valore 0 in poi aumenta,

Models of the first kind (Sect. 7) represent the temporal evolution of the number of subjects in a certain state, for instance the state to be infected. By contrast, compartmental models also represent changes of this state at some moments in the form of transitions of a subject from one compartment to another one.

The SIR-model, which we designated in Sect. 1 "as intensively studied in the 1920s", is particularly simple and has served as a paragon for many others, in particular for those applied to Covid-19. It involves three compartments: S are the susceptible, not yet infected subjects, I the infected ones, and R consists of subjects removed by recovery with immunity or death. The transitions between compartments are described by differential equations for the numbers $S(t)$, $I(t)$ and $R(t)$ of subjects in the compartments as a function of time t . They involve certain parameters such as transition probabilities from one compartment to another one. Under various assumptions, the resulting system of differential equations for S , I and R can be solved explicitly or numerically.

A first important application is to estimate a basic reproduction number R_0 as defined in Sect. 7. It can be expressed by the basic parameters.

Secondly, it turns out that the limit S_∞ of $S(t)$ for $t \rightarrow \infty$ is strictly positive, which means that a certain part of the population will never be infected. This led to the concept of herd immunity, which, however, gave rise to much confusion among people who thought they had something to say about the matter.

After the outbreak of Covid-19, many more involved compartmental models were defined and analysed. Their parameters represented among other features the underlying control strategy to be used. There was for instance the do nothing strategy and also the mitigation strategy, which consisted of the less stringent components of the surveillance-containment strategy defined in Sect. 3. In the much-discussed paper [13] Neil Ferguson and collaborators described the shape of the function I , that is the number of infected subjects, for the do nothing strategy. From the value 0 on it increases, reaches a maximum, decreases and finally reaches 0 at a certain

raggiunge un massimo, diminuisce e infine raggiunge lo 0 ad un certo momento t_{happy} . Questo aveva apparentemente motivato i paesi Regno Unito, Stati Uniti, Svezia e Brasile ad adottare questa strategia per troppo tempo, ignorando che Ferguson aveva previsto (vedi Figura 2) circa 500 mila morti causate dall'epidemia in il Regno Unito e 2.2 milioni negli Stati Uniti prima della sua estinzione al momento t_{happy} .

moment t_{happy} . This had apparently motivated the countries UK, USA, Sweden and Brazil to adopt this strategy for too long, ignoring that Ferguson predicted (see Figure 2) about 500,000 deaths caused by the epidemic in the UK and 2.2 million in the USA before extinction at the moment t_{happy} .

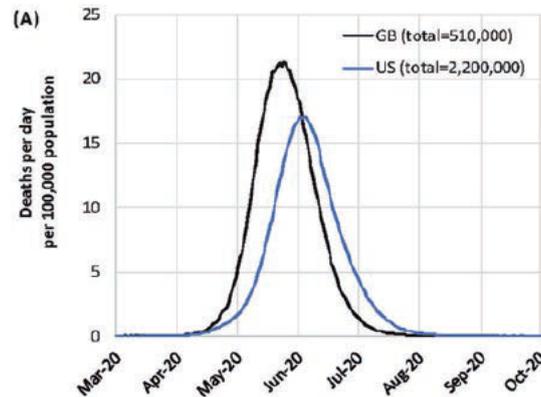


Figura 2: *Morti attese causate da una epidemia con una strategia "do-nothing". Figura riprodotta da [13] con il permesso della School of Public Health, Imperial College London.*
Expected deaths caused by the epidemic for the do-nothing strategy, reproduced from [13] with permission of School of Public Health, Imperial College London.

Al momento, i modelli compartimentali non svolgono quasi alcun ruolo pratico, soprattutto perchè contengono troppi parametri sconosciuti. Alcuni parametri come l'infettività vengono prima stimati con l'aiuto di un modello della curva epidemica, che non sembra essere un percorso di grande successo.

At present, compartmental models play hardly any practical role, mainly because they contain too many unknown parameters. Some parameters such as infectivity are first estimated with the help of a model of the epidemic curve, which does not seem to be a very successful detour.

9. Trattamenti preventivi e curativi

Non sarà certo una sorpresa che diverse compagnie farmaceutiche abbiano iniziato un percorso di sviluppo di prodotti curativi e trattamenti preventivi di vari disturbi che la SARS-CoV-2 può infliggere a una persona. Fino ad ora, non è stato trovato un trattamento curativo. Vi sono solo i metodi ben noti da utilizzare nel trattamento di aspetti non specifici di un caso, come la riduzione del dolore, facilitare la respirazione o accorciare i tempi di recupero con un antivirale. Ci limiteremo quindi ai trattamenti *preventivi*, cioè alle vaccinazioni.

L'obiettivo di un'immunizzazione con un vaccino contro un danno sanitario collegato al Covid-

9. Preventive and Curative Treatments

It will hardly be a surprise that several pharmaceutical companies have started a run for developing curative and preventive treatments of various ailments which SARS-CoV-2 may inflict on a person. Up until now, no curative treatment was found. There are only the well known methods to be used in the treatment of non-specific aspects of a case, such as reducing pain, facilitating breathing or shortening the recovery time by an antiviral drug. We shall therefore restrict ourselves to preventive treatments, that is, to immunisations.

The objective of an immunisation by a vaccine against a Covid-19 connected health deficiency

19 deve essere definito allo stesso modo di qualsiasi altra malattia infettiva. Innanzitutto occorre determinare la popolazione di interesse "chi intendiamo proteggere"? Poi, quali sono le carenze sanitarie che vogliamo prevenire? Per quanto tempo deve durare l'effetto preventivo? Questo è un aspetto particolarmente importante del vaccino, ma di solito lo si trascura quando se ne annuncia uno nuovo. Ad esempio, nella maggior parte dei soggetti la vaccinazione contro il morbillo rimane attiva per tutta la loro durata della vita. Per il Covid-19, l'azienda che sta cercando di sviluppare il vaccino potrebbe accontentarsi di qualche mese, sperando che la SARS-CoV-2 sia poi scomparsa. Infine, occorre determinare l'efficacia, il che rappresenta la parte della popolazione obiettivo effettivamente protetta. Può anche essere definita in termini epidemiologici considerando come esposti tutti i soggetti che non hanno ricevuto il trattamento. Allora l'efficacia è la "frazione eziologica tra i soggetti esposti", (cfr. anche [1], Sez. 15.1, 16.1, 17.1).

Vi è oggi un consenso generale sul fatto che il processo di sviluppo di un vaccino contro una malattia infettiva debba seguire una linea comune ben definita (cfr. [1], Lezione 18, e [5]). Questo dovrebbe valere per il Covid-19, e quindi lo ricorderemo qui.

Per prima cosa, si selezionano una o più sostanze, che, per alcuni motivi, di solito virologici, sembrano possibili candidati per un vaccino. Ognuna di queste deve essere sottoposta a un test clinico per esplorarne le proprietà più importanti. Un tale studio clinico consiste di tre fasi I, II e III. La fase I si occupa di vari aspetti principalmente farmacologici come effetti collaterali per vari possibili dosaggi.

I modelli statistici sono gli strumenti essenziali delle fasi II e III. La fase II mira a fornire una prima idea dell'efficacia del vaccino selezionato. Quindi, si costruisce artificialmente una popolazione obiettivo relativamente piccola. Qui sorgono due problemi di base. Il primo è la definizione di variabile di risposta di interesse. Spesso si studia solo l'immunogenicità, cioè la formazione di anticorpi, ma non di protezione contro la malattia. È un problema particolarmente complesso e sfaccettato nel caso del Covid-19. In secondo luogo, la popolazione di interesse deve includere tra i soggetti vaccinati un numero sufficiente

needs to be defined in the same way as for any other infectious disease. First the target population needs to be determined: whom do we intend to protect? Next, what are the health deficiencies we want to prevent? For how long is the preventive effect to last? This is a particularly important aspect of the vaccine, but is usually suppressed when a new one is announced. For instance, the measles vaccination remains active in most subjects for their entire lives. For Covid-19, the company which is trying to develop the vaccine may be satisfied with a few months, hoping that SARS-CoV-2 will have disappeared after that. Finally, the efficacy needs to be found, which represents the part of the target population actually protected. It may also be defined in epidemiologic terms by regarding as exposed all subjects that have not obtained the treatment. Then the efficacy is the etiological fraction among the exposed subjects (see also [1], Sects. 15.1, 16.1, 17.1).

Nowadays, there is general agreement that the process of developing a vaccine against an infectious disease needs to run along a well-defined common line ([1], Lesson 18, and [5]). This ought to hold for Covid-19, too, and we shall therefore recall it here.

First, one or several substances are selected, which, for some reasons, usually virological ones, look like possible candidates for a vaccine. Each of them needs to be submitted to a clinical trial in order to explore its most important properties. Such a clinical trial consists of three phases I, II and III. Phase I deals with various mainly pharmacologic aspects such as side effects for various possible dosages.

Statistical-mathematical models are the essential tools of the phases II and III. Phase II aims at providing a first idea of the efficacy of the selected vaccine. Thus, a relatively small target population is built artificially. Here two basic problems arise. The first is the definition of the outcome variable of interest. Often only the immunogenicity is being studied, which means the formation of antibodies, but not protection against the disease. It is a particularly complex and manifold problem in the case of Covid-19. Secondly, the target population needs to include among the vaccinated subjects a sufficient number of people who

di persone che sarebbero esposte alla malattia qualora non fossero vaccinate. Poiché la morbilità del Covid-19 nell'intera popolazione di un paese è piccola, tale gruppo deve essere costruito con sfida, cioè infettando artificialmente i suoi membri. Di solito sono volontari e il loro rischio di morire è ridotto, tranne che nelle fasce d'età in cui la mortalità da parte della malattia è alta, cioè, nel caso del Covid-19, per le persone anziane. Di fronte a questo problema etico, in passato gli Stati Uniti hanno usato, per varie malattie infettive, detenuti i cui termini di pena sono stati abbreviati come ricompensa. Vi è stato un tempo in cui il Vietnam, mentre sviluppava un certo vaccino, inviava i campioni per la sperimentazione della Fase II negli Stati Uniti per provarli in questo modo perché le norme dell'etica vietnamita vietano ogni tipo di rischio.

Vi sono di solito diverse prove nella fase II allo scopo di selezionare il potenziale vaccino da studiare alla fine in una prova della fase III. Si tratta di una prova sul campo nel senso che un campione di soggetti viene estratto dall'intera popolazione di interesse, ad esempio tra tutti gli abitanti di un paese all'interno di una certa fascia d'età. La variabile risultante non è l'immunogenicità, ma la protezione contro la malattia nel senso dell'efficacia desiderata. La numerosità del campione è determinata in anticipo dalla precisione della stima prevista dell'efficacia. Come già detto in precedenza, la decisione sulla durata del processo di prova è un elemento cruciale. Se si considera sufficiente un'elevata efficacia durante le prime due settimane dopo la vaccinazione, la prova può essere interrotta dopo due settimane; questa filosofia è alla base delle vaccinazioni contro l'influenza stagionale. Se siamo interessati alla sua efficacia durante i primi dieci anni dopo la vaccinazione, la prova deve durare dieci anni. Oltre ad altri problemi, questo ha causato il lungo ritardo nello sviluppo di un vaccino contro l'Ebola (fine della Sez. 2). Speriamo che non sorvoli su questo aspetto chi cerca di vendere molto presto un vaccino per il Covid-19.

10. Conclusioni e prospettive

La pandemia ha funzionato come una lente d'ingrandimento. In alcuni luoghi, ha mostrato una società fondamentalmente ben funzionante.

would attract the disease when not vaccinated. Since Covid-19 morbidity in the entire population of a country is small, such a group must be constructed by challenge, that is, by infecting its members artificially. They are usually volunteers and their risk of dying is small except in the age groups where the lethality by the disease is high, that is, in the case of Covid-19, for elderly people. Faced with this ethical problem, the USA used, for various previous infectious diseases, prison inmates whose terms were shortened as a reward. There was a time when Vietnam, while developing a certain vaccine, sent its samples for the Phase II trial to the USA to be tested in this way because Vietnamese ethical standards forbade all kinds of challenge.

There are usually several phase II trials in order to select the potential vaccine to finally be studied in a phase III trial. This is a field trial in the sense that a sample of subjects is drawn from the entire population of interest, for instance from among all inhabitants of a country within a certain age group. The outcome variable is not immunogenicity but protection against the disease in the sense of the desired efficacy. The size of the sample is determined beforehand by the precision of the intended estimate of the efficacy. As noted above, the decision about the duration of the trial is a crucial element. If high efficacy during the first two weeks after vaccination is considered sufficient, the trial may be stopped after two weeks; this philosophy underlies the vaccinations against the seasonal influenza. If we are interested in its efficacy during the first ten years after vaccination, it must last ten years. This has, in addition to other problems, caused the long delay in developing an Ebola vaccine (end of Sect. 2). We hope that it will not be glossed over by those who are trying to sell a Covid-19 vaccine very soon.

10. Outlook

The pandemic has functioned like a magnifying glass. In some places, it has shown a basically well-functioning society. In other places it has

In altri luoghi ha rivelato scandali e intollerabili disuguaglianze sociali. In particolare, ha rispecchiato lo stato del sistema sanitario pubblico di un paese.

Questo articolo si è proposto di descrivere il ruolo della matematica nella pandemia. Come detto sopra, vi sono due parti di queste conclusioni. Prendiamo la prima, cioè: cosa si può imparare dall'epidemia? Nella Sezione 1 abbiamo dato una panoramica dei principali rami della matematica che svolgono un ruolo importante. Poi le Sezioni 4-9 hanno dato un cenno delle applicazioni più frequenti; titoli e ordine delle sezioni corrispondono vagamente ai rami della matematica in questione. Così, vi sono state analisi di modelli matematici nel lavoro di laboratorio e nei modelli matematico-statistici nel giudicare la qualità dei test; metodi demografici nella raccolta dei dati, diversi modi per modellare matematicamente l'evoluzione della pandemia e l'epidemiologia clinica nei tentativi di sviluppare un vaccino.

In questo modo, l'articolo mirava a chiarire il potenziale ruolo della matematica nel prendere decisioni. Da un lato si è scoperto che in pratica il ruolo della curva epidemica o dei modelli compartimentali è molto più limitato rispetto a quanto divulgato in molte pubblicazioni. Le decisioni basate su di esse possono anche avere conseguenze disastrose, ad esempio quelle basate sul concetto matematico di immunità di gregge. Pertanto, la fiducia cieca nelle argomentazioni matematiche è ingiustificata.

D'altra parte, negare l'esistenza di un valido fondamento matematico per una strategia di controllo è altrettanto dannoso. È stato fatto in Danimarca con la strategia del tracciare e confinare secondo un rapporto di un gruppo di esperti di accademici e funzionari sanitari, che rifletteva gli interessi degli ambienti medici, industriali e governativi.

Questo commento ci porta alla seconda parte delle nostre conclusioni, ovvero: cosa fare in futuro? Gli autori di questo articolo avevano incominciato all'inizio di maggio con

"Dal momento che la pandemia di Covid-19 non è finita...".

Mentre stiamo per finire il nostro lavoro alla metà luglio, non è ancora finita! È ancora molto atti-

revealed scandals and intolerable social inequalities. In particular, it has reflected the state of a country's public health system.

The present article aimed at describing the role of mathematics in the pandemic. As said above, there are two parts to this outlook. Let us take up the first one, namely: What can be learnt from the epidemic? In Sect. 1 we gave an overview of the main branches of mathematics that play a role. Then the Sects. 4–9 sketched the most frequent applications; their titles and their order correspond vaguely to the branches of mathematics concerned. Thus, there were mathematical pattern analysis in laboratory work and statistical-mathematical models in judging the quality of tests; demographic methods in the collection of data; different ways to model the evolution of the pandemic mathematically; and clinical epidemiology in attempts to develop a vaccine.

In this way, the article aimed to clarify the potential role of mathematics in making decisions. On the one hand, it turned out that in practise the role of the epidemic curve or compartmental models is much more restricted than advertised in many publications. Decisions based on them may even have disastrous consequences, for instance those based on the mathematical concept of herd immunity. Thus, blind trust in mathematical arguments is unjustified.

On the other hand, denying the existence of a valid mathematical-scientific foundation for a control strategy is just as detrimental. It was done in Denmark with the tracing and lock-down strategy by a report of an expert group of health academics and officials, which reflected the interests of medical, industrial and governmental circles.

This comment leads us to the second part of our outlook namely: what to do in the future? The authors of this article started it in early May with the line

"Since the Covid-19 pandemic is not over, ...".

While we are finally finishing our work in the middle of July, it is still not over! It is still even

va, ma ha preso una forma largamente diversa. Per questo sembra naturale analizzarne le attuali caratteristiche alla luce dei fatti che abbiamo descritto nelle precedenti Sezioni 4-9 e domandarci: Quali lezioni possiamo trarre sulle strategie di controllo da applicare ora?

Il Covid-19 non ha più una sola fonte. Riappare in regioni grandi o piccole di molte parti del mondo, che possono essere di varie forme ed estensioni: un'unica casa di riposo per anziani in Francia, due distretti in Germania, una grande città come Pechino, un'intera provincia in Spagna, o un intero paese come la Nuova Zelanda. Li chiameremo focolai per distinguerli dai cluster, che indicano insiemi discreti di persone. Un seguito preciso dell'evoluzione dei casi in questi focolai si scontra con le molteplici difficoltà spiegate nelle Sezioni 5 e 6 e non saranno ripetute qui.

Una prima domanda naturale da porsi è, quindi: perché i focolai attivi persistono e riappaiono? La Sez. 3 ha presentato tre componenti di strategie di controllo di successo: la ricerca dei contatti, il *lockdown* e le mascherine. Mentre il tracciamento dei contatti continua a malincuore, il *lockdown* e l'uso di mascherine sono stati ampiamente abbandonati, spesso a causa delle politiche di governi in cerca di popolarità.

E poi, cosa si dovrebbe fare? Nelle Sezioni 7, 8 e 9 abbiamo spiegato, utilizzando in particolare argomentazioni matematiche, in che misura le strategie di controllo lì trattate soffrono di gravi inconvenienti. Questo ci lascia con la combinazione di due misure: all'interno di un nido un rigoroso *lock-down* come il distanziamento sociale e il divieto di grandi assembramenti; ai suoi confini: chiuderli o permettere il passaggio solo se combinati con la quarantena. Ad esempio, la Nuova Zelanda, considerata come singolo nido, ha adottato queste misure rigorose. Di conseguenza, non ci sono ora nuovi casi, tranne due casi intorno al 14 luglio in strutture di isolamento gestite. Altri focolai si comporteranno in modo simile, speriamo.

very active but has taken a largely different form. Hence it seems natural to analyse its present characteristics in the light of the facts we have described in the Sections 4–9 above and to ask ourselves: Which lessons can we draw regarding the control strategies to be applied now?

Covid-19 no longer surges from a single source. It reappears in large or small regions of many parts of the world, which may be of various forms and extensions: a single home for the elderly in France, two districts in Germany, a large city like Beijing, an entire province in Spain, or a whole country like New Zealand. We shall call them “nests” to distinguish them from “clusters”, which denote certain discrete sets of people. A precise followup of the evolution of cases in these nests meets with the manifold difficulties explained in Sects. 5 and 6 and will not be repeated here.

A first natural question to ask is, then: Why do active nests persist and reappear? Sect. 3 presented three components of successful control strategies: contact tracing; lock-down and masks. While contact tracing continues reluctantly, lock-down and wearing masks were widely abandoned, often as a result of governmental policies seeking popularity.

Next, what should be done? In the Sections 7, 8 and 9 we have explained, using in particular mathematical arguments, to what extent the strategies of control treated there suffer from serious drawbacks. This leaves us with the combination of two measures: within a nest a rigorous lock-down such as social distancing and preventing larger assemblies of people; at its borders: closing them or only allowing passage when combined with quarantine. For example, New Zealand, regarded as a single nest, has taken such rigorous measures. As a result, there are now no new cases, except two cases around the 14 July in managed isolation facilities. Other nests will act similarly, we hope.

Questo articolo è apparso in "Newsletter of the the European Mathematical Society 117 (Sept. 2020) 29". Ringraziamo gli autori e la European Mathematical Society per il permesso di tradurlo e riprodurlo.



- [1] K. Krickeberg, V. T. Pham, T.M.H. Pham: *Epidemiology. Key to Public Health*, Springer, Berlin (2019).
- [2] K. Dietz: *The first epidemic model: a historical note on P.D. En'ko*, Austral. J. Statist., 30A (1988) 56.
- [3] P. E. M. Fine: *John Brownlee and the measurement of infectiousness: an historical study in epidemic theory*, J. R. Statist. Soc. A, 142 (1979) 347.
- [4] K. Dietz, J. A. P. Heesterbeek: *Daniel Bernoulli's epidemiological model revisited*, Mathematical Biosciences, 180 (2002) 1.
- [5] M. E. Halloran, Jr. Longini, M. Ira, C. Struchiner: *Design and Analysis of Vaccine Studies*, Springer, Berlin (2010).
- [6] K. Krickeberg: *Principles of health information systems in developing countries*, Health Information Management J., 36 (2007) 8.
- [7] L. Simonsen, V. Andreasen: *Alternative corona strategy: we can beat the infection down for the count with quarantine and tracing of infectors (in Danish)*. Videnskab.dk. (2020)
<https://videnskab.dk/kultur-samfund/alternativ-corona-strategi-vi-kan-banke-smitten-helt-i-bund-med-karantaene-og>
- [8] D. Mumford, A. Desolneux: *Pattern Theory. The Stochastic Analysis of Real-world Signals*, AK Peters, Ltd., Natick, MA (2010).
- [9] N.C. Jones, P.A. Pevzner: *An Introduction to Bioinformatics Algorithms*, MIT Press, Cambridge, MA (2004).
- [10] K.G. Andersen: *The proximal origin of SARS-CoV-2*, Nature Medicine, 26 (2020) 450.
- [11] *FT analysis of mortality data from M. Wolf, The risks of lifting lockdowns prematurely are very large*, Financial Times. 4 June 2020.
<https://www.ft.com/content/2a6cd7c2-a5b5-11ea-92e2-cbd-9b7e28ee6>
- [12] F. Koehler-Rieper, C.H.F. Roehl, E. De Micheli: *A novel deterministic forecast model for COVID-19 epidemic based on a single ordinary integro-differential equation*, European Journal of Physics PLUS, 135 (2020) 599.
- [13] N. Ferguson, et al. : *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*, MRC Centre for Global Infectious Disease Analysis, Imperial College London, UK, (2020) doi: 10.25561/77482.
<https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/>



Bernhelm Booß-Bavnbek: nato nel 1941, ha studiato matematica dal 1960 al 1965 all'Università di Bonn. Ha svolto ricerca, insegnamento e prassi prima in econometria e ricerca operativa, poi in analisi geometrica e processi di membrane in fisiologia cellulare. Affiliato all'Università di Roskilde dal 1977.

Klaus Krickeberg: nato nel 1929, ha studiato matematica dal 1946 al 1951 all'Università Humboldt di Berlino. Professore in diverse Università europee ed estere. Ha svolto ricerca, insegnamento e prassi prima in matematica e poi in epidemiologia e salute pubblica. Molto di questo lavoro è stato svolto in paesi in via di sviluppo. Pensionato dal 1998.

Bernhelm Booß-Bavnbek: born in 1941, studied mathematics from 1960 to 1965 at Bonn University. Research, teaching and practical work first in econometrics and operations research and then in geometric analysis and membrane processes of cell physiology. Affiliated to Roskilde University since 1977.

Klaus Krickeberg: born in 1929, studied mathematics from 1946 to 1951 at the Humboldt-University Berlin. Professor at several universities in Europe and outside; research, teaching and practical work first in mathematics and then in epidemiology and public health. Much of this was done in developing countries. Retired since 1998.

Il trattamento per gruppi

Christian Genest

Università McGill, Montréal, Canada

Christiane Rousseau

Università di Montréal, Montréal, Canada

Il tracciamento massiccio è un elemento essenziale della lotta contro la propagazione del coronavirus. Ma come far fronte a una possibile penuria di reagenti e di materiale? Operando i test su miscele di prelievi e facendo appello alla matematica.

Con il timore anticipato di una seconda ondata di COVID-19, numerosi esperti sono concordi nel dire che l'instaurazione di un piano di tracciamento su grande scala è necessario per arrestare la propagazione del coronavirus. Realizzati su campioni rappresentativi della popolazione, i test immunologici, sierologici o antigenici permetterebbero anche di stimare la diffusione della malattia, di giudicare il grado di immunità collettiva e di adattare i mezzi di gestione della pandemia.

Perché sia coronata da successo, il varo di una strategia di tracciamento presuppone l'accesso a risorse adeguate di personale e di materiale. Con la crescita della richiesta a livello planetario si profila all'orizzonte una penuria dei reagenti necessari alle analisi di laboratorio; questa resta una preoccupazione delle autorità della sanità pubblica in Canada e nel resto del mondo.

Sapendo che la maggior parte dei test si rivelano (molto fortunatamente) negativi, si può mettere a profitto la matematica per fare meglio? Sì, si può, segnatamente realizzando test di gruppo su miscele di prelievi costruiti in maniera giudiziosa.

Il tracciamento per gruppi

Immaginiamo che un laboratorio abbia ricevuto 100 prelievi per il tracciamento. Li divide a caso in 5 gruppi di 20 ciascuno. Poi, gruppo per gruppo, utilizza la metà di ciascuno dei 20 prelievi per costituire una miscela alla quale si applica il test.

Se il test effettuato su una miscela è negativo si può subito concludere che nessun membro del gruppo interessato è infetto. Se il test è invece positivo, allora si procede a test individuali sulla seconda metà di ognuno dei 20 prelievi.

Se i 100 prelievi originali provengono tutti da persone sane, questa procedura permette di accertarsene facendo 5 test invece di 100. Se un solo individuo è infetto, bastano $5 + 20 = 25$ test per individuarlo. Se sono infettate due persone, le si possono individuare ancora con 25 test se sono nello stesso gruppo, ma ne occorrono $5 + 20 + 20 = 45$ se appartengono a gruppi differenti. E così via se vi sono tre o più persone infette.

Come si può constatare, il tracciamento per gruppi permette dunque di realizzare economie importanti, purché la sensibilità e la specificità del test non siano modificate dalla miscela, come si è supposto qui e come è spesso il caso in pratica.

Il laboratorio avrebbe anche potuto applicare la stessa strategia di tracciamento a 10 gruppi di 10 prelievi. Se un solo individuo è infettato, non vi sarebbe stato allora bisogno che di 20 test per identificarlo. In cambio, sarebbero stati necessari 10 test per concludere che nessuno è infetto.

Quale di queste due strategie è migliore? E ne esistono altre che siano preferibili a queste? La risposta dipende dalla prevalenza della malattia, cioè dalla proporzione della popolazione che è infetta.

In una nota apparsa nel 1943 in *The Annals of Mathematical Statistics*, l'americano Robert Dorfman [1] riporta che, nella sua forma più elementare, il rintracciamento per gruppi era già stato usato nella Seconda Guerra mondiale per individuare i casi di sifilide tra i coscritti. Questo approccio si è affermato e ne esistono oggi molte varianti che sono utilizzate dovunque in Nord America per controllare la presenza dell'HIV, dell'influenza, o del Virus del Nilo occidentale.

Ottimizzare l'algoritmo

Dorfman ha mostrato come determinare la numerosità ottimale di un gruppo in funzione della prevalenza $p \in [0, 1]$ della malattia. Indichiamo con $n \geq 2$ la numerosità del gruppo e supponiamo che i suoi membri costituiscano un campione aleatorio rappresentativo della popolazione.

Se X denota il numero sconosciuto di persone infette nel gruppo, questa variabile obbedisce a una legge binomiale¹ di parametri n e p , donde

$$Pr(X = 0) = (1 - p)^n .$$

Poiché ogni individuo ha probabilità $1 - p$ di essere sano

$$Pr(X > 0) = 1 - Pr(X = 0) = 1 - (1 - p)^n .$$

Se $X = 0$, non si faranno allora che $N = 1$ test. Tuttavia, se $X > 0$, si faranno $N = n + 1$ test. In media, il numero di test che saranno effettuati, chiamato speranza di N e denotato con $E(N)$, è eguale a

$$\begin{aligned} E(N) &= 1 \times Pr(X = 0) + (n + 1) \times Pr(X > 0) \\ &= n + 1 - n(1 - p)^n . \end{aligned}$$

Questa è una funzione crescente di p . Se $p = 0$, si ha $E(N) = 1$, questo che è evidente perché

¹Si tratta qui di un'approssimazione che è giustificata nella misura in cui la popolazione è molto grande rispetto alla numerosità dei gruppi.

nessuno ha la malattia e dunque basta un solo test per confermarlo. Se $p = 1$, si ha $E(N) = n + 1$ perché il primo test sarà necessariamente positivo.

Per ogni valore di $p \in [0, 1]$, è possibile determinare il costo relativo legato all'uso del tracciamento per gruppi studiando il comportamento del rapporto

$$\frac{E(N)}{n} = 1 + \frac{1}{n} - (1 - p)^n ,$$

in funzione di n . Più $E(N)/n$ è piccolo, più conviene ricorrere ai test per gruppo, naturalmente a condizione che il rapporto sia inferiore a 1. Quando $p = 0$, si trova $E(N)/n = 1/n$, di modo che si ha interesse a prendere n più grande possibile. Quando $p = 1$, si ha sempre

$$\frac{E(N)}{n} = 1 + \frac{1}{n} > 1 ,$$

perché il test sul gruppo è sempre positivo dunque e non fa quindi che aggiungere altri test.

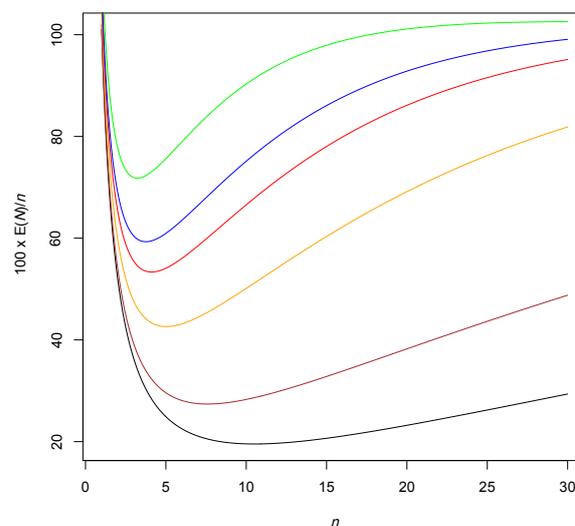


Figura 1: Tracciato della curva $100 \times E(N)/n$ in funzione di n per diversi valori di p : 1% (nero), 2% (marrone), 5% (arancio), 8% (rosso), 10% (blu) e 15% (verde).

Per un valore fissato di p , la funzione $100 \times E(N)/n$ rappresenta la percentuale media di test effettuati in funzione della numerosità, n , del gruppo. La Figura 1 mostra il grafico di questa funzione per diversi valori di p , corrispondenti a una prevalenza dell'1% (nero), 2% (marrone), 5%

(arancio), 8% (rosso), 10% (blu) e 15% (verde). Come si può vedere, la numerosità ottimale dipende dalla miscela, che corrisponde al minimo della curva, e varia in funzione della proporzione, p , degli individui infetti nella popolazione. La Tabella 1, riportata da Dorfman [1], dà la scelta ottimale di n per qualche valore di p .

p (%)	n	Costo relativo (%)
1	11	20
2	8	27
5	5	43
8	4	53
10	4	59
15	3	72

Tabella 1: Scelta ottimale di n per alcuni valori della percentuale p degli infetti.

Generalizzazioni

Il protocollo di *test* descritto sopra è l'esempio di algoritmo adattativo a due passi. Si definisce adattativo perché la scelta (e dunque il numero) dei *test* da effettuare nel secondo passo dipende dal risultato del *test* realizzato al primo. Esistono diversi modi di migliorare il successo di questo tipo di algoritmo. In particolare la procedura può essere estesa aumentando il numero di passi.

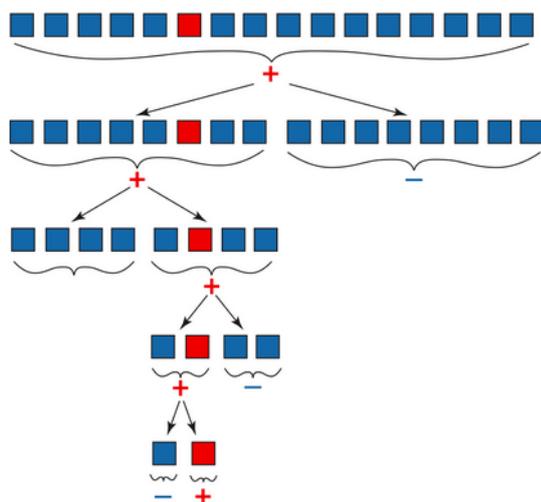


Figura 2: Rappresentazione grafica di un algoritmo di divisione binaria adattativo a 5 passi applicato ad un insieme iniziale di $n = 2^4 = 16$ prelievi, di cui solo uno è infetto.

Ecco un algoritmo classico, che chiameremo algoritmo di divisione binaria, e che possiede certe proprietà di ottimalità (si veda la Figura 2).

- Si prende un intero n della forma $n = 2^s$ e si effettuano k passi del *test*, ove $k \leq s + 1$.
- Al primo passo si effettua un *test* della miscela dei prelievi di tutto il gruppo.
- Se il *test* risulta positivo, si divide allora il gruppo in due sottogruppi di 2^{s-1} prelievi e si effettua un *test* su ognuno di essi.
- Si procede nello stesso modo sino al k -esimo passo, al quale si effettua il *test* sui membri di un sottogruppo dichiarato positivo al passo precedente. Nel caso particolare $k = s + 1$, questo sottogruppo ha solo due elementi.

Se il gruppo contiene un solo individuo infetto, questo algoritmo premetterà di identificarlo in esattamente $s + 1 = \log_2(n) + 1$ passi. Come regola generale, più alto è il numero di passi, migliori sono le economie realizzate tramite questo approccio. Tuttavia, se bisogna attendere da 24 a 48 ore perché il *test* dia il risultato, i ritardi nella consegna dei risultati rischiano di essere controproducenti. Si noti anche che questa miglioria richiede prelievi biologici più importanti. Questo non è considerato veramente un problema e lo si ritroverà in tutti gli algoritmi presentati di seguito.

Un algoritmo non adattativo

Per controllare meglio il tempo di risposta, si può anche pensare di utilizzare metodi non adattativi di tracciamento per gruppi. Questi protocolli comportano un solo passo, ciò che permette di effettuare tutti i *test* simultaneamente. Essi si rivelano inoltre molto efficaci per il tracciamento dei casi se si dispone di una stima affidabile della prevalenza della malattia.

Spieghiamo questo concetto mediante il seguente esempio, sviluppato da un gruppo di ricercatori ruandesi nel quadro della lotta attuale al COVID-19. Si forma dapprima un campione aleatorio di numerosità $n = 3^m$. Si stabilisce quindi una corrispondenza tra i 3^m individui e i

punti di un ipercubo discreto $\{0, 1, 2\}^m$. Si veda la figura 3 per un esempio nel caso $m = 3$.

L'approccio proposto consiste allora nell'effettuare contemporaneamente $3m$ test su miscele di campioni che comprendono ognuna 3^{m-1} individui. Le miscele sono tuttavia formate secondo modalità molto precise, ossia dei tagli nell'ipercubo. In effetti, se x_1, \dots, x_m denotano gli assi coordinati dell'ipercubo, allora ogni miscela corrisponde agli individui situati nell'iperpiano $x_i = t$, ove $i \in \{1, \dots, m\}$ e $t \in \{0, 1, 2\}$ è una tranche di 3^{m-1} individui.

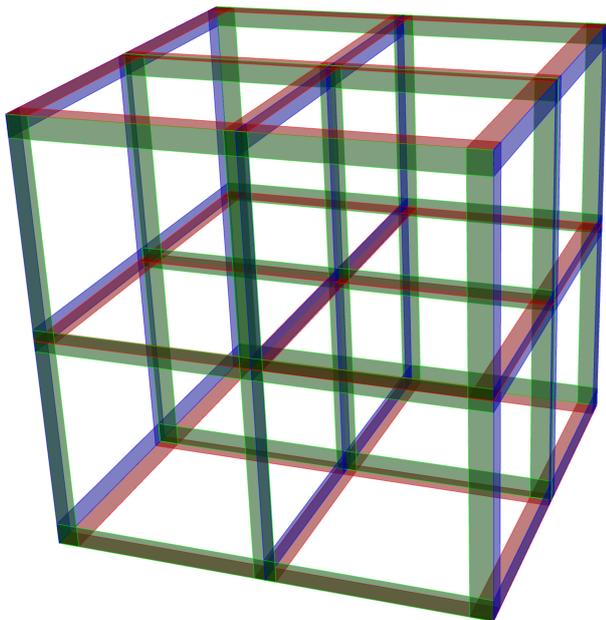


Figura 3: Ipercubo discreto $\{0, 1, 2\}^3$. Ogni punto del reticolo corrisponde ad un individuo di un campione casuale di dimensioni $n = 3^3 = 27$. Gli individui $3^2 = 9$ di ciascuna delle 9 miscele si trovano lungo una fetta rossa, blu o verde.

Quando $m = 3$, come nella Figura 3, si effettuano allora $3 \times 3 = 9$ test su gruppi di $3^2 = 9$ individui. Quando $m = 4$, che è il valore riportato nel caso del Ruanda, si effettuano piuttosto 12 test a partire da un campione di $n = 81$ individui. Ciò significa che ogni prelievo è diviso in quattro porzioni uguali e contribuisce a quattro test differenti. Inoltre ogni test è fatto su una miscela di 27 campioni.

Questo approccio si basa su una tecnica di costruzione dei codici correttori d'errori descritta nel riquadro². Uno dei grandi vantaggi è che

²Si veda anche <http://accromath.uqam.ca/accro/wp-content/uploads/2020/02/Codes.pdf> per una introduzione ai codici correttori d'errori.

la composizione delle miscele è tale che se nel campione vi è un solo individuo infetto questo può essere individuato con certezza. Di contro, se è presente più di una persona infetta, occorre allora procedere a un secondo giro di test.

Esaminiamo ora l'esempio ruandese nel caso $n = 81 = 3^4$. Sapendo che il numero X di individui infetti nel campione obbedisce a una legge binomiale di parametri $n = 81$ e p , si ha

$$Pr(X \leq 1) = (1 - p)^{81} + 81p(1 - p)^{80}.$$

Questa strategia è interessante se è grande la probabilità che sia $X \leq 1$. Ma perché sia $Pr(X \leq 1) \geq 0.95$, per esempio, occorre avere $p \leq 0.44\%$; in altre parole, la prevalenza della malattia deve essere bassa. Già a 1% di prevalenza si ha $Pr(X \leq 1) = 0.806$ e in quasi il 20% dei casi bisognerà ricorrere a un secondo giro di test. Tuttavia, ancora per l'1% di prevalenza si ha

$$Pr(X = 2) = \binom{81}{2} p^2(1 - p)^{79} = 0.146,$$

da cui $Pr(X \leq 2) = 0.952$.

È possibile controllare facilmente i costi se ci si mostra astuti nel modo di condurre il secondo passo, quando vi è più di un individuo infetto. Per esempio, tutti i casi che corrispondono a $X = 2$ soddisfano alla seguente proprietà (si faccia riferimento alla Figura 4 dove si mostrano, nel caso $m = 3$, le tranche per le quali il test è positivo):

(P) Per ogni valore di $i \in \{1, \dots, m\}$, vi sono al più due tranche della forma $x_i = s$ e $x_i = t$ che portano a un test positivo, ed esiste almeno un valore di i per il quale si hanno esattamente due test positivi di questa forma.

Se k è il numero di valori di $i \in \{1, \dots, m\}$ per i quali si hanno due test positivi della forma specificata in (P), il numero dei test aggiuntivi richiesti per individuare tutti gli individui infetti è dato nella Tabella 2.

Se non è verificata la proprietà (P), si debbono allora controllare tutti gli individui delle tranche che hanno portato ad un test positivo, ciò che si traduce in al più 81 test supplementari. Alla fine, il costo totale è dunque inferiore o eguale a $100 \times 13.06/81 \approx 16.1\%$.

Costruire un algoritmo non adattativo

Si vuole costruire un algoritmo non adattativo per *testare* un gruppo di persone. Si rappresenterà questo algoritmo mediante una tabella di T righe e n colonne, o, in maniera equivalente, mediante una matrice di dimensione $T \times n$ con tutti gli elementi eguali a 0 o a 1. Gli elementi della j -esima colonna rappresentano i *test* ai quali partecipa il j -esimo individuo. Così, $m_{ij} = 1$ se il j -esimo individuo contribuisce all' i -esimo *test* e 0 altrimenti.

Costruiamo un vettore X di lunghezza n che rappresenti il gruppo che ci si accinge a *testare*: la sua j -esima coordinata, x_j , vale 1 se il j -esimo individuo è infetto, 0 altrimenti. Si decide di trattare X come una parola di lunghezza n e le sue coordinate eguali a 1 come errori in una parola iniziale X_0 in cui tutte le coordinate erano nulle. Degli algoritmi di codici correttori d'errori permettono di correggere gli errori che si sarebbero prodotti nella trasmissione di X_0 , ciò che equivale a identificare quali siano le coordinate x_j di X che valgono 1, ossia esattamente lo scopo cercato. Come illustrato nell'esempio, un algoritmo di correzione degli errori non può correggere che un numero massimo prefissato di errori k scelto al momento della sua costruzione.

La matrice M è la matrice del codice. Una proprietà sufficiente perché il codice possa correggere k errori è che la matrice sia k -disgiunta. Definiamo questa nozione. Non si vuole che se degli individui j_1, \dots, j_k (non necessariamente distinti) sono infetti, allora un j_{k+1} -esimo individuo infetto rimane inosservato. Dare la colonna j (ossia l'individuo j) è lo stesso che dare il sottoinsieme A_j di $\{1, \dots, T\}$ degli elementi uguali a 1 (vale a dire l'insieme dei *test* ai quali contribuisce quell'individuo). Se gli individui j_1, \dots, j_k sono infetti, allora tutti i *test* corrispondenti all'unione $A_{j_1} \cup \dots \cup A_{j_k}$ saranno positivi. Perché non passi inosservato il j_{k+1} -esimo individuo infetto bisogna che $A_{j_{k+1}}$ non sia incluso in $A_{j_1} \cup \dots \cup A_{j_k}$. La matrice M è k -disgiunta se ciò si verifica per ogni j_1, \dots, j_k e per ogni j_{k+1} distinto da j_1, \dots, j_k . La matrice 12×81 che corrisponde all'esempio precedente utilizzato in Ruanda sarebbe una matrice 1-disgiunta.

Esistono due principali tipi di metodi per costruire matrici k -disgiunte. Il primo è probabilistico: si generano a caso matrici i cui elementi sono 1 con probabilità q e 0 altrimenti. Per n, T e k ben scelti la probabilità che la matrice sia k -disgiunta è non nulla; per tentativi si finisce dunque per generare una matrice k -disgiunta.

Il secondo metodo, algebrico, è preso a prestito dalla teoria dei codici correttori d'errori di Reed-Solomon. Questo permette di costruire matrici M nelle quali tutte le righe hanno lo stesso numero m di elementi non nulli e tutte le colonne hanno lo stesso numero c di elementi non nulli. Così si effettuano tutti i *test* su sottogruppi di numerosità m e ogni prelievo individuale è diviso in c parti e può essere incluso in c *test* distinti.

k	Numero di <i>test</i> supplementari
1	0
2	4
3	8
4	16

Tabella 2: Scelta ottimale di n per alcuni valori della percentuale p degli infetti.

Prospettive

La ricerca di algoritmi adatti alla lotta contro il coronavirus raggiunge il culmine. Per esempio, recentemente è stato elaborato e implementato in laboratorio un algoritmo da un gruppo di ricerca israeliano: consiste nella realizzazione di 48 *test* sullo stesso campione di numerosità $8 \times 48 = 384$. Ogni prelievo individuale è diviso in sei parti uguali. Ogni *test* si basa sulla miscela di 48 di queste parti, una per ogni individuo. Ogni individuo è dunque presente in sei *test* differenti.

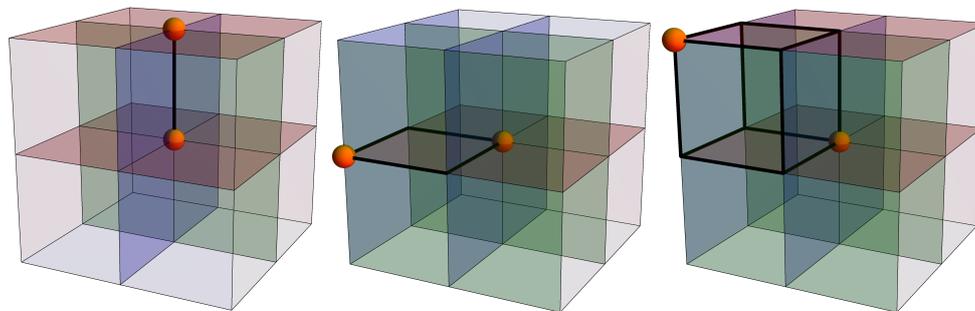


Figura 4: L'ipercubo discreto $\{0, 1, 2\}^3$ e le tre possibilità per due individui infetti: sulla stessa linea retta parallela ad un asse (sinistra), ai vertici opposti di un quadrato in un piano (centro) o ai vertici opposti di un cubo (destra). In ogni caso, le fette che portano ad un test positivo sono identificate in rosso, verde o blu con le stesse convenzioni di colore della Figura 3.

Nel laboratorio è stato programmato un *robot* per preparare le 48 miscele. L'algoritmo consente di identificare sino a quattro persone infette. Si ha così bisogno di un numero di *test* otto volte inferiore a quello degli individui. Di nuovo, minore è la percentuale di infetti, migliore è la *performance* dell'algoritmo.

Naturalmente, nell'elaborazione di una procedura di un *test* statistico entrano in gioco altre considerazioni. Per semplificare, noi abbiamo qui supposto implicitamente che il *test* usato sia infallibile. Nella pratica anche le migliori procedure possono dare falsi positivi o falsi negativi. La sensibilità e la specificità dei *test* sono elementi importanti da prendere in considerazione al momento di raccomandarne l'implementazione, come pure la loro fattibilità in termini di tempo, di costi e di complessità delle manipolazioni.

Questo articolo è apparso in lingua originale francese in:

<http://accromath.uqam.ca/volume/volume-15-2-ete-automne-2020/>

e nella sua traduzione italiana in:

<http://accromath.uqam.ca/2020/09/il-tracciamento-per-gruppi/>

Ringraziamo gli autori e Accromath per il permesso di riprodurlo.



[1] R. Dorfman: *The Detection of Defective Members of Large Populations*, Ann. Math. Statist., 4 (1943) 436.



Applicazioni del Formalismo Quantistico alle Scienze Cognitive

La vita è complessa: ha sia una parte reale che una parte immaginaria.

— Anonimo

Sandro Sozzo

*Department of Economics, Finance and Accounting (EFA) and Centre IQSCS
School of Business, University of Leicester
223 London Road, Leicester LE2 1ZE, United Kingdom*

In questo articolo verranno presentati alcuni risultati recenti sull'applicazione del formalismo matematico della meccanica quantistica (QM) ai domini cognitivi e socio-economici, ottenuti nell'ambito di una decennale collaborazione tra i gruppi di ricerca di Bruxelles e Leicester. In particolare, si mostrerà come il formalismo quantistico in spazi di Hilbert permetta di modellizzare e fornire una spiegazione unitaria per una serie di fenomeni cognitivi complessi, compresi i giudizi e le decisioni in presenza di incertezza, per i quali l'applicazione di formalismi di tipo classico porta invece a risultati paradossali. I risultati presentati in questo articolo supportano un programma di ricerca emergente, noto con il nome di *quantum cognition programme*, che presuppone l'esistenza di strutture quantistiche in domini non-fisici.

I modelli delle scienze cognitive

I modelli sviluppati in psicologia cognitiva tradizionalmente impiegano strutture matematiche di tipo insiemistico, per esempio, algebre Booleane, probabilità Kolmogoroviane, ecc., per rappresentare i fenomeni cognitivi complessi, per esempio, i fenomeni che coinvolgono il linguaggio, la memoria, la percezione, i giudizi e le decisioni [1]. Tali strutture, originariamente adottate in fisica classica, sono state successivamente estese a numerose discipline, dalla statistica alla psicologia, dall'economia alla finanza e alle scienze informatiche. Esse sono perciò comunemente dette strutture classiche. In particolare, le seguenti nozioni, relative a strutture classiche, saranno utilizzate nel seguito di questo articolo.

Sia Ω un insieme non vuoto, sia $\mathcal{P}(\Omega)$ l'insieme delle parti di Ω , e sia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$. Si dice che \mathcal{A} è una σ -algebra (Booleana) se soddisfa le seguenti condizioni:

- (i) $\Omega \in \mathcal{A}$,
- (ii) per ogni $A \in \mathcal{A}$, $A' = \Omega \setminus A \in \mathcal{A}$,

(iii) se $\{A_i\}_{i \in \mathbb{N}}$ è una famiglia numerabile di elementi di \mathcal{A} , $\cup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

Una σ -algebra \mathcal{A} è **distributiva**, ovvero, per ogni $A, B, C \in \mathcal{A}$, valgono le seguenti proprietà:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad (1)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \quad (2)$$

Sia ora Ω un insieme non vuoto, sia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ una σ -algebra, e sia $p : A \in \mathcal{A} \mapsto p(A) \in [0, 1]$. Si dice che p è una **misura (normalizzata) di probabilità** se soddisfa le seguenti condizioni:

$$(1) p(\Omega) = 1,$$

(2) se $\{A_i\}_{i \in \mathbb{N}}$ è una famiglia numerabile di elementi di \mathcal{A} , tale che, per ogni $A_i, A_j \in \mathcal{A}$, $A_i \cap A_j = \emptyset$, allora $p(\cup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} p(A_i)$.

Le condizioni (1) e (2) sono state introdotte da Kolmogorov [2]. La misura p è perciò detta **probabilità Kolmogoroviana**. Gli elementi dell'algebra \mathcal{A} sono in particolare utilizzati in matematica applicata per rappresentare accadimenti del mondo naturale. Essi perciò abitualmente designano eventi.

Se p è una probabilità Kolmogoroviana e $A, B \in \mathcal{A}$, la **probabilità condizionale** $p(A|B)$ di A dato B è definita dall'equazione

$$p(A|B) = \frac{p(A \cap B)}{p(B)}. \quad (3)$$

La probabilità condizionale $p(B|A)$ di B dato A è definita in modo analogo e vale la relazione

$$p(B)p(A|B) = p(A \cap B) = p(A)p(B|A) \quad (4)$$

nota con il nome di **formula di Bayes**. Se ne derivano le seguenti leggi:

$$\begin{aligned} p(A \cap B) &\leq \min [p(A), p(B)] \\ &\leq \max [p(A), p(B)] \leq p(A \cup B), \end{aligned} \quad (5)$$

detta **legge di monotonicità**, e

$$p(A) = p(B)p(A|B) + p(B')p(A|B'), \quad (6)$$

detta **legge della probabilità totale**.

Crescente evidenza empirica mostra tuttavia che i modelli basati su tali strutture classiche non sono in generale in grado di spiegare il modo concreto in cui gli individui esprimono giudizi o prendono decisioni in presenza di incertezza.

In particolare, le seguenti deviazioni empiriche dalle strutture classiche sono state identificate in psicologia cognitiva [3, 4].

(a) **Errori nei giudizi probabilistici**. Gli individui giudicano la probabilità che si verifichi la congiunzione "A e B" (rispettivamente, disgiunzione "A o B") di due eventi A e B maggiore (rispettivamente, minore) della probabilità che si verifichi uno dei due, il che implica una violazione della legge di monotonicità in (6).

(b) **Errori nelle decisioni**. Gli individui preferiscono un'azione A ad un'azione B quando sono sicuri del verificarsi di un evento Z, e anche quando sono sicuri del non verificarsi di Z, ma preferiscono B ad A quando non sanno se Z si verificherà o meno, il che implica una violazione della legge della probabilità totale in (5).

Deviazioni di tipo (a) sono state identificate, in particolare, negli effetti di sovra- e sotto-estensione in giudizi di appartenenza sulle combinazioni dei concetti [5, 6] e nelle fallacie della congiunzione [7] e della disgiunzione [8]. Deviazioni di tipo (b) sono state identificate, in particolare, nell'effetto disgiunzione [9] e nelle violazioni della teoria dell'utilità attesa [10, 11].

È legittimo allora chiedersi se la probabilità Kolmogoroviana sia l'unico modo a disposizione per formalizzare l'incertezza sul verificarsi degli eventi. Esiste in effetti almeno una possibile alternativa: quella di utilizzare la struttura matematica della probabilità quantistica, che formalizza la struttura probabilistica degli eventi che si verificano a livello microscopico e che sono governati dalle leggi della meccanica quantistica (MQ). La probabilità quantistica non soddisfa infatti le condizioni matematiche della probabilità Kolmogoroviana [12, 13]. In effetti, a partire dagli anni Novanta del secolo scorso, il formalismo matematico della MQ è stato applicato con successo alla modellizzazione dei fenomeni cognitivi e sociali, come quelli nei punti (a) e (b), in cui l'applicazione di strutture classiche si è rivelata altamente problematica [3, 4, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. In contrapposizione infatti al **programma di ricerca di Tversky-Kahneman**, secondo cui le deviazioni empiriche dalle strutture classiche sono dovute a fallacie del ragionamento umano e vengono spiegate in termini di euristiche e *bias* cognitivi [26], il programma di ricerca della *quantum co-*

gnition parte dal presupposto che tali deviazioni rivelino strutture genuinamente quantistiche, come una forma di ragionamento emergente, nel pensiero umano [27, 28].

Prima di procedere ulteriormente, è bene tuttavia anticipare qualche *caveat* riguardo al *quantum cognition programme*, allo scopo di evitare fraintendimenti. Innanzitutto, solo il formalismo matematico della MQ, non la sua interpretazione fisica, viene utilizzato in tali domini cognitivi. Ciò implica, in particolare, che non è richiesta alcuna assunzione riguardo all'esistenza di processi quantistici microscopici che determinerebbero le deviazioni dalle strutture classiche menzionate sopra. Inoltre, il *quantum cognition programme* si discosta nettamente da altri programmi di ricerca, solo apparentemente simili, quali i programmi del *quantum brain* e della *quantum mind*, in cui si ipotizza che fenomeni quantistici microscopici siano responsabili dei processi cerebrali e della coscienza.

Perché allora usare strutture quantistiche nei domini cognitivi? Per rispondere a questa domanda, osserviamo che un crescente numero di studiosi di scienze cognitive e sociali, compresi i premi Nobel per l'economia Paul Samuelson [29], Amartya Sen [30] e Daniel Kahneman [31], concordano sul fatto che i giudizi e le decisioni sono processi attivi che creano e non si limitano a registrare. Più precisamente, in un giudizio o decisione, si verifica una interazione, a livello concettuale-cognitivo, tra l'oggetto del giudizio o decisione e l'individuo che giudica o decide. Nel corso di tale interazione, i cui dettagli non possono in generale essere né controllati né previsti, se non a livello probabilistico, l'individuo agisce come contesto di misura e influenza l'oggetto alternandone lo stato. Da questo punto di vista, giudizi o decisione ricorda da vicino un processo di misura quantistica, dove si verifica una interazione, a livello fisico, tra l'apparato macroscopico che esegue la misura e l'oggetto fisico che viene misurato. Nel corso di tale interazione, l'apparato agisce come contesto di misura e influenza l'oggetto misurato in un modo non controllabile né predicibile, se non a livello probabilistico. Questa analogia tra l'indeterminismo dei processi fisici microscopici e quello dei processi concettuali ha costituito il punto di partenza per gli studi sull'applicazione di strutture

quantistiche ai sistemi concettuali, come si vedrà nelle prossime sezioni.

Struttura e dinamica dei concetti

Una delle grandi sfide della psicologia cognitiva, anche per le sue applicazioni alla linguistica computazionale e all'intelligenza artificiale, è rappresentata dalla comprensione dei meccanismi di combinazione dei concetti e del modo in cui il significato può essere espresso da tali combinazioni. Secondo la visione classica, risalente ad Aristotele, tutte le istanze di un concetto condividono un insieme comune di proprietà definitorie, che risultano quindi necessarie e sufficienti. Tale visione classica è tuttavia in contrasto con alcune caratteristiche empiriche rivelate dal modo in cui gli individui usano i concetti in pratica. In primo luogo, i concetti hanno una struttura graduata: per esempio, gli individui giudicano l'esemplare Pettirosso più tipico dell'esemplare Cicogna, quando la tipicità è definita rispetto al concetto Uccello. In secondo luogo, il significato di un concetto dipende fortemente dal contesto in cui il concetto è espresso: per esempio, l'esemplare Serpente non è un esemplare tipico del concetto Animale Domestico, quindi Serpente farà registrare una bassa tipicità rispetto a tale concetto. Tuttavia, se si considera il concetto Strano Animale Domestico, allora è ragionevole ritenere che Serpente farà registrare un'alta tipicità rispetto a tale concetto (si veda, per esempio, [32]).

Per superare le difficoltà della visione classica, sono state proposte numerose teorie psicologiche dei concetti, per esempio, la teoria dei prototipi, la teoria degli esemplari, la teoria della teoria, ecc. (si veda, per esempio [33]). Alcune delle teorie che sono state proposte hanno poi suggerito una rappresentazione matematica esplicita dei concetti e delle loro combinazioni. Per esempio, un tentativo di superare la concezione aristotelica dei concetti utilizzando strutture matematiche classiche è costituito dalla rappresentazione dei concetti mediante insiemi *fuzzy*, introdotta da Zadeh negli anni Ottanta del secolo scorso [34]. Nella rappresentazione *fuzzy*, dato un insieme di esemplari \mathcal{X} e un insieme di concetti \mathcal{C} su cui siano definite una congiunzione e una disgiun-

zione, per ogni esemplare $X \in \mathcal{X}$, si assume che esista una **funzione di appartenenza graduata** $\mu_X : A \in \mathcal{C} \mapsto \mu_X(A) \in [0, 1]$ tale che la congiunzione "A e B" e la disgiunzione "A o B" dei concetti A e B soddisfino le seguenti regole (in cui la dipendenza da X è sottintesa per semplicità):

$$\mu(A \text{ e } B) = \min [\mu(A), \mu(B)], \quad (7)$$

$$\mu(A \text{ o } B) = \max [\mu(A), \mu(B)]. \quad (8)$$

Le equazioni (7) e (8) esprimono la **regola del minimo per la congiunzione fuzzy** e la **regola del massimo per la disgiunzione fuzzy**, rispettivamente.

Tuttavia, l'applicazione di strutture classiche e, in particolare, la rappresentazione *fuzzy*, porta a previsioni che sono in disaccordo con una serie di risultati sperimentali su semplici combinazioni di concetti. Tale disaccordo è espresso dai seguenti effetti cognitivi.

Effetto Guppy.¹ Gli individui giudicano l'esemplare Guppy come esempio molto tipico della congiunzione Pet-Fish, senza giudicare Guppy né esempio tipico di Pet né esempio tipico di Fish. Questo risultato è in disaccordo con l'Equazione (7) [35].

Sovra-/sotto-estensione dei pesi di appartenenza. Per numerosi esemplari, gli individui giudicano la probabilità, o peso, di appartenenza dell'esemplare rispetto alla congiunzione (rispettivamente, disgiunzione) maggiore (rispettivamente, minore) della probabilità, o peso, di appartenenza dell'esemplare rispetto ad uno o entrambi i concetti componenti. Questo risultato è in disaccordo con l'Equazione (7) (rispettivamente, (8)) [5, 6].

Contraddizioni *borderline*. Una contraddizione *borderline* è un enunciato della forma "John è alto e John non è alto", e viene giudicato vero per alcuni casi limite di "John". Questo risultato è in disaccordo con l'Equazione (7) [36].

Ne consegue il **problema della combinazione**, vale a dire, l'impossibilità di rappresentare consistentemente la combinazione di due (o più) concetti in termini della rappresentazione dei concetti componenti. Tale problema venne rile-

vato in modo eclatante in una serie di test sulla combinazione dei concetti naturali, eseguiti da James Hampton, psicologo della *City University of London*, verso la fine degli anni Ottanta del secolo scorso. Nel seguito verranno riportati alcuni dei risultati ottenuti da Hampton.

Si considerino, per esempio, i concetti naturali Frutta e Verdura, la loro congiunzione Frutta e Verdura, e la loro disgiunzione Frutta o Verdura. Hampton chiese ad un gruppo di individui di indicare su un questionario l'appartenenza a questi concetti di una serie di esemplari in una lista contenente Mela, Pomodoro, Broccoli, Oliva, Cocco, Mandorla, Fungo, Uvetta, Ghianda, ecc. Più precisamente, agli individui che partecipavano al test veniva chiesto di giudicare su una scala di Likert a 7 valori, $\{+3, +2, +1, 0, -1, -2, -3\}$, l'appartenenza di tali esemplari ai concetti Frutta e Verdura e alle loro combinazioni Frutta e Verdura e Frutta o Verdura. Un valore +3 indicava un giudizio di forte appartenenza dell'esemplare al concetto, un valore -3 indicava un giudizio di forte non-appartenenza, un valore 0 indicava che l'individuo non aveva alcuna preferenza rispetto all'appartenenza dell'esemplare al concetto. Le frequenze relative delle risposte positive venivano considerate, nel limite dei grandi numeri, probabilità, o pesi, di appartenenza. Hampton raccolse i pesi di appartenenza di numerosi esemplari rispetto a diversi concetti naturali, alla loro congiunzione ed alla loro disgiunzione, trovando deviazioni significative da strutture classiche ed introducendo una specifica terminologia per classificare tali deviazioni, come segue [5, 6].

Sia X un esemplare e siano $\mu(A)$, $\mu(B)$, $\mu(A \text{ e } B)$ e $\mu(A \text{ o } B)$ i pesi di appartenenza di X rispetto ai concetti A, B, alla congiunzione "A e B" e alla disgiunzione "A o B", rispettivamente. Si dirà che X è (**doppiamente**) **sovra-esteso rispetto alla congiunzione** se $\mu(A \text{ e } B) > \mu(A)$ (e) o $\mu(A \text{ e } B) > \mu(B)$. Si dirà che X è (**doppiamente**) **sotto-esteso rispetto alla disgiunzione** se $\mu(A \text{ o } B) < \mu(A)$ (e) o $\mu(A \text{ o } B) < \mu(B)$.

Recentemente, si è trovato che le deviazioni da strutture classiche e, in particolare, da modelli di probabilità Kolmogoroviana, dei dati di Hampton sono ancora più profonde degli effetti di sovra- e sotto-estensione [17, 37]. Sia infatti X

¹Il *guppy*, dal nome del suo scopritore, il naturalista inglese R. J. L. Guppy, è un piccolo pesce d'acqua dolce, di origine tropicale, simile al pesce rosso.

un esemplare e siano $\mu(A)$, $\mu(B)$ e $\mu(A \text{ e } B)$ i pesi di appartenenza di X rispetto ai concetti A , B e alla congiunzione "A e B", rispettivamente. Con riferimento alle strutture classiche definite nella sezione introduttiva, si dirà che tali pesi corrispondono a **dati classici per la congiunzione** se possono essere rappresentati in un modello probabilistico Kolmogoroviano, cioè, se esiste una σ -algebra \mathcal{A} , una probabilità Kolmogoroviana $p : \mathcal{A} \rightarrow [0, 1]$, e $E_A, E_B \in \mathcal{A}$ tali che:

$$\mu(A) = p(E_A), \quad (9)$$

$$\mu(B) = p(E_B), \quad (10)$$

$$\mu(A \text{ e } B) = p(E_A \cap E_B). \quad (11)$$

Si dimostra allora che $\mu(A)$, $\mu(B)$ e $\mu(A \text{ e } B)$ sono dati classici per la congiunzione se e solo se le seguenti disuguaglianze sono simultaneamente soddisfatte:

$$\mu(A \text{ e } B) - \min [\mu(A), \mu(B)] \leq 0, \quad (12)$$

$$\mu(A) + \mu(B) - \mu(A \text{ e } B) \leq 1 \quad (13)$$

[17, 37]. L'Equazione (12) esprime la legge di monotonicità della probabilità Kolmogoroviana relativamente alla congiunzione di due eventi (si veda l'Equazione (5)). Si noti che l'Equazione (12) non è sufficiente a garantire l'esistenza di un modello di probabilità Kolmogoroviana per una data terna di pesi $\mu(A)$, $\mu(B)$ e $\mu(A \text{ e } B)$.

È facile trovare dati non-classici per la congiunzione nello studio in [5]. Infatti, i pesi di appartenenza dell'esemplare *Rasoio* rispetto ai concetti *Arma*, *Strumento* e alla congiunzione *Arma e Strumento* sono 0.63, 0.68 e 0.83, rispettivamente, il che viola l'Equazione (12) e anche la regola del minimo per la congiunzione *fuzzy* (7). Utilizzando la terminologia di Hampton, si ha un **effetto di doppia sovra-estensione**, in questo caso.

Analogamente, la natura non - Kolmogoroviana della probabilità nei dati di Hampton sulla disgiunzione di due concetti non si riduce a effetti di sotto - estensione. Sia infatti X un esemplare e siano $\mu(A)$, $\mu(B)$ e $\mu(A \text{ o } B)$ i pesi di appartenenza di X rispetto ai concetti A , B e alla disgiunzione "A o B", rispettivamente. Si dirà che tali pesi corrispondono a **dati classici per la disgiunzione** se possono essere rappresentati in un modello probabilistico Kolmogoroviano, cioè, se esiste una σ -algebra \mathcal{A} , una probabilità Kol-

mogoroviana $p : \mathcal{A} \rightarrow [0, 1]$, e $E_A, E_B \in \mathcal{A}$ tali che:

$$\mu(A) = p(E_A), \quad (14)$$

$$\mu(B) = p(E_B), \quad (15)$$

$$\mu(A \text{ o } B) = p(E_A \cup E_B). \quad (16)$$

Si dimostra allora che $\mu(A)$, $\mu(B)$ e $\mu(A \text{ o } B)$ sono dati classici per la disgiunzione se e solo se le seguenti disuguaglianze sono simultaneamente soddisfatte:

$$\mu(A \text{ o } B) - \max [\mu(A), \mu(B)] \geq 0, \quad (17)$$

$$\mu(A) + \mu(B) - \mu(A \text{ o } B) \geq 0 \quad (18)$$

[17, 37]. L'Equazione (17) esprime la legge di monotonicità della probabilità Kolmogoroviana relativamente alla disgiunzione di due eventi (si veda l'Equazione (5)). Anche in questo caso, l'Equazione (17) non è sufficiente a garantire l'esistenza di un modello di probabilità Kolmogoroviana per una data terna di pesi $\mu(A)$, $\mu(B)$ e $\mu(A \text{ o } B)$.

È facile trovare dati non-classici per la disgiunzione nello studio in [6]. Infatti, i pesi di appartenenza dell'esemplare *Occhiali da Sole* rispetto ai concetti *Abbigliamento Sportivo*, *Attrezzatura Sportiva* e alla disgiunzione *Abbigliamento Sportivo o Attrezzatura Sportiva* sono 0.40, 0.20 e 0.10, rispettivamente, il che viola l'Equazione (17) e anche la regola del massimo per la disgiunzione *fuzzy* (8). In questo caso, si ha un **effetto di doppia sotto - estensione**. Se invece si considerano i pesi di appartenenza dell'esemplare *Oliva* rispetto ai concetti *Frutta*, *Verdura* e alla disgiunzione *Frutta o Verdura*, si trovano i valori 0.50, 0.10 e 0.80, rispettivamente, il che soddisfa l'Equazione (17) ma non la (18). In quest'ultimo caso, si ha dunque una violazione dalla probabilità Kolmogoroviana che non si riduce ad un effetto di sovra- o sotto- estensione.

La maggior parte dei dati raccolti da Hampton e da altri studiosi che si sono occupati dei meccanismi di combinazione dei concetti non possono essere riprodotti da modelli di probabilità Kolmogoroviana. Ne consegue che strutture probabilistiche classiche di questo tipo non sono in grado di rappresentare nemmeno le più semplici combinazioni di due concetti.

Nelle prossime sezioni si presenterà la prospettiva teorica generale che è stata elaborata nell'ambito della collaborazione Bruxelles-Leicester per rappresentare i sistemi concettuali e le loro interazioni. Tale prospettiva adotta il formalismo quantistico in spazi di Hilbert ed è motivata da due decenni di studi sui fondamenti epistemologici e matematici della MQ, in particolare, sulle motivazioni che giustificano l'uso di strutture quantistiche in determinati domini empirici.

Rappresentazione dei sistemi concettuali in spazi di Hilbert

Negli anni Sessanta e Settanta del secolo scorso, furono sviluppati diversi approcci ai fondamenti della MQ, i quali si proponevano di giustificare l'uso degli spazi di Hilbert nel formalismo della teoria a partire da assiomi fisicamente giustificati, relativi a nozioni ben definite e collegate in modo diretto alle operazioni che sono tipicamente eseguite in un laboratorio di fisica. Alcuni di questi approcci vennero detti *realistico-operazionali* e uno di essi venne inizialmente sviluppato a Ginevra [38, 39] e successivamente completato a Bruxelles (si veda, per esempio, [40]). Nell'approccio di Ginevra-Bruxelles, ogni sistema fisico viene astrattamente descritto in termini delle nozioni elementari ed operativamente definite di *stato*, *contesto* e *proprietà* e delle mutue relazioni statistiche (*sistema SCoP*). Se un sistema SCoP soddisfa alcuni assiomi, opportunamente scelti, allora tale sistema può essere rappresentato univocamente (a meno di isomorfismi) da una struttura algebrica in uno spazio di Hilbert [39, 40].

Discuteremo ora le ragioni che suggeriscono di applicare il formalismo matematico della MQ ai sistemi concettuali. Come si vedrà, l'uso di tale formalismo è giustificato dalle analogie nelle descrizioni *realistico-operazionali* di sistemi fisici e sistemi concettuali.

Un esperimento su un sistema fisico viene tipicamente eseguito in un laboratorio di fisica, che viene comunemente identificato con un dominio spazio-temporale. Il sistema fisico viene preliminarmente sottoposto a una procedura di preparazione da parte dello sperimentatore, al termine della quale il sistema si trova in uno stato

definito. Questo stato esprime aspetti realistici del sistema misurato, nel senso che, come conseguenza dell'essere in quello stato, il sistema ha alcune proprietà attuali indipendentemente da qualsiasi misura che può essere eseguita su di esso (*parte realistica*). Quando una misura viene eseguita sul sistema fisico, l'apparato macroscopico agisce come contesto di misura e interagisce, a livello fisico, con il sistema misurato alterandone lo stato in modo in generale non-deterministico (*parte operativa*).² Quindi, il sistema fisico, i suoi stati, i contesti, le proprietà e le reciproche relazioni statistiche sono rappresentati nel formalismo della MQ in spazi di Hilbert.

Si è recentemente dimostrato che una siffatta descrizione *realistico-operazionale* può essere elaborata per ogni sistema concettuale, sia esso un concetto, una combinazione di concetti, oppure un più complesso sistema decisionale [41]. Più precisamente, in un sistema concettuale, si possono introdurre, in analogia con un sistema fisico, le seguenti nozioni basilari:³

- (i) i *test cognitivi* vengono eseguiti in un laboratorio di psicologia, che viene identificato con un dominio spazio-temporale;
- (ii) il complesso delle procedure sperimentali concepite dallo sperimentatore, il dispositivo sperimentale e l'effetto cognitivo che si intende studiare definiscono un **sistema concettuale** A ;
- (iii) le procedure sperimentali adottate (per esempio, il contenuto concettuale del testo introdotto in un questionario), inoltre, preparano A in un stato definito p_A ;

²La nozione di *determinismo* è piuttosto complessa ed è tuttora oggetto di accesi dibattiti in filosofia della scienza. Intuitivamente, si parla di *determinismo* quando ad uno stato presente completamente definito corrisponde un unico stato futuro ad esso compatibile e altrettanto definito. Con riferimento a tale interpretazione intuitiva, si può allora dire che le trasformazioni di stato che coinvolgono i sistemi fisici descritti dalla fisica classica sono fondamentalmente *deterministiche*. Per quanto visto brevemente in questa sezione, invece, le trasformazioni di stato che coinvolgono i sistemi fisici nel corso di una misura in MQ sono invece *non-deterministiche*, in quanto lo stato iniziale non determina in modo univoco lo stato finale.

³Si noti che questa descrizione di un sistema concettuale diverge dalla visione classica, e anche dalle successive estensioni *fuzzy*, dei concetti, in quanto un concetto non è più semplicemente considerato come un insieme di elementi (o esemplari), ma ha una struttura più complessa e dipende in modo cruciale dal contesto.

(iv) un contesto e ha natura concettuale e può influenzare A alterandone lo stato in modo non-deterministico, cioè tale che solo la probabilità di transizione $\mu(p, e, p_A)$ dallo stato iniziale p_A ad uno stato finale p è nota;

(v) il sistema concettuale A può possedere una data proprietà a , a seconda del contesto e e dello stato p_A .

Una volta introdotti, in analogia con la descrizione realistico-operazionale di un sistema fisico, gli elementi fondamentali di un sistema concettuale, si procederà alla rappresentazione matematica di sistemi concettuali e dei rispettivi stati, contesti, proprietà e mutue relazioni statistiche utilizzando il formalismo della MQ in spazi di Hilbert. Nella prossime due sezioni discuteremo tale rappresentazione, distinguendo tra sistemi concettuali individuali e combinazioni concettuali.

Rappresentazione dei concetti individuali

Si consideri un test cognitivo consistente in un questionario in cui viene chiesto ad un gruppo di partecipanti di giudicare se un dato esemplare, per esempio, *Oliva*, appartenga al concetto *Frutta*. Tale giudizio ha due possibili risultati, "si" e "no", e, pertanto, corrisponde alla misura di una proprietà a_{Oliva} sul concetto *Frutta*. Il concetto *Frutta* viene descritto come un sistema concettuale preparato, mediante il questionario del test, in uno stato iniziale p_{Frutta} . Ciascuno degli individui che partecipano al test interagisce con il sistema concettuale *Frutta* preparato nello stato iniziale p_{Frutta} e ne trasforma lo stato in un modo che è, in generale, non-deterministico (si veda la nota 2). Se l'interazione produce, per esempio, il risultato "si" relativamente all'esemplare *Oliva*, lo stato del sistema concettuale *Frutta* verrà trasformato nello stato finale p_{Oliva} .⁴ L'insieme degli individui che partecipano al test agisce quindi come un contesto (di misura) complessivo e per il sistema concettuale *Frutta*. Una volta raccolte tutte le risposte individuali, se ne ottiene una distribuzione statistica $\mu(p_{\text{Oliva}}, e, p_{\text{Frutta}})$, che viene considerata, nel limite dei grandi nume-

⁴Nel gergo quantistico, misure siffatte sono dette misure ideali di prima specie.

ri, come la probabilità, o peso, di appartenenza dell'esemplare *Oliva* rispetto al concetto *Frutta*.

Tale descrizione suggerisce che un sistema concettuale possa essere modellizzato utilizzando la rappresentazione quantistica di stati, contesti, proprietà e le mutue relazioni statistiche nello spazio di Hilbert. Più precisamente, utilizzando la notazione di Dirac tipicamente esposta nei manuali di MQ [42], il sistema concettuale *Frutta* viene associato ad uno spazio di Hilbert \mathcal{H} sul campo \mathbb{C} dei numeri complessi, lo stato iniziale p_{Frutta} viene rappresentato da un vettore unitario (normalizzato) $|p_{\text{Frutta}}\rangle$ di \mathcal{H} , la proprietà che *Oliva* sia giudicato un esemplare del concetto *Frutta* viene rappresentata da un operatore di proiezione ortogonale M_{Oliva} su \mathcal{H} , il contesto della misura e viene rappresentato da un operatore autoaggiunto su \mathcal{H} o, equivalentemente, dalla misura spettrale $\{M_{\text{Oliva}}, \mathbb{1} - M_{\text{Oliva}}\}$, dove $\mathbb{1}$ è l'operatore identità.⁵ Infine, la probabilità, o peso, di appartenenza $\mu(p_{\text{Oliva}}, e, p_{\text{Frutta}})$ è calcolata mediante la regola di Born della probabilità quantistica, cioè, $\mu(p_{\text{Oliva}}, e, p_{\text{Frutta}}) = \|M_{\text{Oliva}}|p_{\text{Frutta}}\rangle\|^2$.

Nella sezione successiva, si presenterà un'esposizione semplificata della rappresentazione matematica generale che è stata elaborata per la congiunzione e la disgiunzione di due concetti.

Rappresentazione della combinazione di due concetti

La rappresentazione matematica che tratteggiamo in questa sezione, sebbene semplificata e giustificata dalla prospettiva teorica generale presentata, per esempio, in [17, 19, 20, 37, 43, 44, 45], permette di riprodurre la maggior parte dei dati raccolti da Hampton sulla congiunzione e la disgiunzione di due concetti.

Iniziamo dal modello quantistico per la congiunzione di due concetti. Sia X un esemplare, e siano $\mu(A)$, $\mu(B)$ e $\mu(A \text{ e } B)$ i pesi di appartenenza di X rispetto ai concetti A , B e alla loro congiunzione " A e B ", rispettivamente. Ciascuno

⁵Ricordiamo che un operatore lineare $O : \mathcal{H} \rightarrow \mathcal{H}$ si dice autoaggiunto se, per ogni $|A\rangle, |B\rangle \in \mathcal{H}$, $\langle A|O|B\rangle = \langle B|O|A\rangle^*$, dove $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ denota il prodotto scalare complesso definito su \mathcal{H} e $\langle B|O|A\rangle^*$ è il complesso coniugato del numero complesso $\langle B|O|A\rangle$. Se l'operatore autoaggiunto O è anche idempotente, cioè, $O^2 = O$, O viene detto operatore di proiezione ortogonale.

di questi concetti è associato nel modello quantistico allo spazio di Hilbert \mathcal{H} sul campo \mathbb{C} dei numeri complessi. Lo stato iniziale del concetto A è rappresentato dal vettore unitario $|A\rangle \in \mathcal{H}$, cioè, $\langle A|A\rangle = 1$, e lo stato iniziale del concetto B è rappresentato dal vettore unitario $|B\rangle \in \mathcal{H}$, cioè, $\langle B|B\rangle = 1$. Si supponga, per semplicità, che $|A\rangle$ e $|B\rangle$ siano ortogonali, cioè, $\langle A|B\rangle = 0$. Inoltre, lo stato iniziale della congiunzione concettuale "A e B" è rappresentato dalla combinazione lineare⁶

$$|A \text{ e } B\rangle = \frac{1}{\sqrt{2}}(|A\rangle + |B\rangle). \quad (19)$$

Infine, la misura di un individuo che giudichi l'appartenenza dell'esemplare X rispetto ai concetti A , B e "A e B" è rappresentata dall'operatore di proiezione ortogonale M su \mathcal{H} .

Si usi ora la regola di Born della probabilità quantistica per calcolare i pesi, o probabilità, di appartenenza. Si ha:

$$\mu(A) = \langle A|M|A\rangle, \quad (20)$$

$$\mu(B) = \langle B|M|B\rangle, \quad (21)$$

$$\mu(A \text{ e } B) = \langle A \text{ e } B|M|A \text{ e } B\rangle. \quad (22)$$

Usando le equazioni (20) e (21), la linearità dello spazio di Hilbert \mathcal{H} e la proprietà $\langle B|M|A\rangle^* = \langle A|M|B\rangle$ (l'operatore di proiezione ortogonale M è in particolare autoaggiunto, si veda la nota 5), l'Equazione (22) diventa:

$$\begin{aligned} \mu(A \text{ e } B) &= \frac{1}{2}(\langle A| + \langle B|)M(|A\rangle + |B\rangle) \\ &= \frac{1}{2}(\langle A|M|A\rangle + \langle A|M|B\rangle \\ &\quad + \langle B|M|A\rangle + \langle B|M|B\rangle) \\ &= \frac{\mu(A) + \mu(B)}{2} + \text{Re} \langle A|M|B\rangle, \end{aligned} \quad (23)$$

dove $\text{Re} \langle A|M|B\rangle$ è la parte reale del numero complesso $\langle A|M|B\rangle$. Si riconosce in $\text{Re} \langle A|M|B\rangle$ il termine di interferenza, che è tipico dell'esperimento quantistico della doppia fenditura. La sua presenza, infatti, esprime interferenza quantistica, dal momento che questo termine produce una fluttuazione intorno al valore medio $\frac{1}{2}(\mu(A) + \mu(B))$, che è ciò che ci si aspette-

⁶Nel gergo quantistico, si dice che lo stato rappresentato dal vettore unitario in (19) è la "sovrapposizione" degli stati rappresentati dai vettori unitari $|A\rangle$ e $|B\rangle$.

rebbe classicamente nell'esperimento della doppia fenditura. Si noti, inoltre, il ruolo fondamentale svolto dai numeri complessi in questa rappresentazione.

È possibile elaborare ulteriormente l'Equazione (23) per ottenere la formula quantistica per la probabilità della congiunzione, distinguendo due casi, come segue.

(i) Se $\mu(A) + \mu(B) \leq 1$, si ha:

$$\mu(A \text{ e } B) = \frac{\mu(A) + \mu(B)}{2} + \sqrt{\mu(A)\mu(B)} \cos \theta_c. \quad (24)$$

(ii) Se $\mu(A) + \mu(B) > 1$, si ha:

$$\begin{aligned} \mu(A \text{ e } B) &= \frac{\mu(A) + \mu(B)}{2} \\ &\quad + \sqrt{1 - \mu(A)}\sqrt{1 - \mu(B)} \cos \theta_c. \end{aligned} \quad (25)$$

Nelle Equazioni (24) e (25), θ_c denota l'angolo di interferenza per la congiunzione.

Il modello quantistico presentato sopra può essere effettivamente realizzato nello spazio di Hilbert complesso \mathbb{C}^3 di tutte le triple ordinate di numeri complessi. Sia infatti $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ la base canonica ortogonale (ON) di \mathbb{C}^3 . Distinguiamo nuovamente due casi, come segue.

(I) Se $\mu(A) + \mu(B) \leq 1$, $\mu(A), \mu(B) \neq 0, 1$, le equazioni (20)–(24) sono soddisfatte dalla scelta:

$$|A\rangle = \left(\sqrt{1 - \mu(A)}, 0, \sqrt{\mu(A)} \right), \quad (26)$$

$$\begin{aligned} |B\rangle &= e^{i\theta_c} \left(-\sqrt{\frac{\mu(A)\mu(B)}{1 - \mu(A)}}, \right. \\ &\quad \left. -\sqrt{\frac{1 - \mu(A) - \mu(B)}{1 - \mu(A)}}, \sqrt{\mu(B)} \right), \end{aligned} \quad (27)$$

mentre l'operatore di proiezione ortogonale M proietta sul sottospazio 1-dimensionale generato dal vettore $(0, 0, 1)$.

(II) Se $\mu(A) + \mu(B) > 1$, $\mu(A), \mu(B) \neq 0, 1$, le equazioni (20)–(23) e (25) sono soddisfatte dalla scelta:

$$|A\rangle = \left(\sqrt{\mu(A)}, 0, \sqrt{1 - \mu(A)} \right), \quad (28)$$

$$\begin{aligned} |B\rangle &= e^{i\theta_c} \left(\sqrt{\frac{(1 - \mu(A))(1 - \mu(B))}{\mu(A)}}, \right. \\ &\quad \left. \sqrt{\frac{\mu(A) + \mu(B) - 1}{\mu(A)}}, -\sqrt{1 - \mu(B)} \right), \end{aligned} \quad (29)$$

mentre l'operatore di proiezione ortogonale M proietta sul sottospazio 2-dimensionale generato dai vettori $(1, 0, 0)$ e $(0, 1, 0)$.⁷

Si consideri nuovamente l'esemplare Rasoio rispetto ai concetti Arma, Strumento e alla congiunzione Arma e Strumento. Poiché $\mu(A) + \mu(B) = 0.63 + 0.78 = 1.41 > 1$, la rappresentazione quantistica richiede l'uso delle equazioni (25), (28) e (29). In particolare, dall'Equazione (25) si ottiene:

$$\theta_c = \arccos\left(\frac{2\mu(A \text{ e } B) - \mu(A) - \mu(B)}{\sqrt{1 - \mu(A)}\sqrt{1 - \mu(B)}}\right) = 64.02^\circ. \quad (30)$$

Pertanto, i pesi di appartenenza $\mu(A) = 0.63$, $\mu(B) = 0.78$ e $\mu(A \text{ e } B) = 0.83$ sono rappresentati da $\theta_c = 64.02^\circ$, $|A\rangle = (0.79, 0, 0.61)$ e $|B\rangle = e^{i64.02^\circ}(0.36, 0.81, -0.47)$ nel modello quantistico. In questo caso, la doppia sovra-estensione dell'esemplare Rasoio rispetto alla congiunzione Arma e Strumento può essere spiegata come un effetto di interferenza costruttiva ($\theta_c = 64.02^\circ < 90^\circ$) tra i concetti componenti Arma e Strumento nella sovrapposizione Arma e Strumento.

Veniamo ora al modello quantistico per la disgiunzione di due concetti. Nuovamente, sia X un esemplare e siano $\mu(A)$, $\mu(B)$ e $\mu(A \text{ o } B)$ i pesi di appartenenza di X rispetto ai concetti A , B e alla loro disgiunzione " $A \text{ o } B$ ", rispettivamente. I concetti in questione sono associati allo spazio di Hilbert complesso \mathcal{H} , mentre gli stati iniziali dei concetti A e B sono rappresentati dai vettori unitari ortogonali $|A\rangle$ e $|B\rangle$, rispettivamente, e lo stato iniziale della disgiunzione concettuale " $A \text{ o } B$ " è rappresentato dal vettore unitario $|A \text{ o } B\rangle = \frac{1}{\sqrt{2}}(|A\rangle + |B\rangle)$ ottenuto sovrapponendo gli stati A e B (si vedano l'equazione (19) e la nota 6). Infine, la misura di un individuo che giudichi l'appartenenza dell'esemplare X rispetto ai concetti A , B e " $A \text{ o } B$ " è rappresentata dall'operatore di proiezione ortogonale M su \mathcal{H} .

Usando nuovamente la regola di Born della probabilità quantistica per i pesi di appartenenza $\mu(A)$ e $\mu(B)$, come in (20) e (21), rispettivamente,

⁷La situazione in cui $\mu(A) = 0$ o $\mu(B) = 0$ richiede ulteriori dettagli tecnici e una struttura di spazio di Hilbert particolare, lo spazio di Fock, come si vedrà nel seguito. Non insistiamo su questo aspetto, per brevità.

si trova per $\mu(A \text{ o } B)$:

$$\begin{aligned} \mu(A \text{ o } B) &= \langle A \text{ o } B | M | A \text{ o } B \rangle \\ &= \frac{\mu(A) + \mu(B)}{2} + \text{Re} \langle A | M | B \rangle. \end{aligned} \quad (31)$$

Anche nel caso della disgiunzione, l'Equazione (31) può essere elaborata ulteriormente per ottenere la formula quantistica per la probabilità della disgiunzione, la quale prende matematicamente due forme, come segue.

(i') Se $\mu(A) + \mu(B) \leq 1$, si ha:

$$\begin{aligned} \mu(A \text{ o } B) &= \frac{\mu(A) + \mu(B)}{2} \\ &+ \sqrt{\mu(A)\mu(B)} \cos \theta_d. \end{aligned} \quad (32)$$

(ii') Se $\mu(A) + \mu(B) > 1$, si ha:

$$\begin{aligned} \mu(A \text{ o } B) &= \frac{\mu(A) + \mu(B)}{2}, \\ &+ \sqrt{1 - \mu(A)}\sqrt{1 - \mu(B)} \cos \theta_d. \end{aligned} \quad (33)$$

Nelle Equazioni (32) e (33), θ_d denota l'angolo di interferenza per la disgiunzione. Anche in questo caso, il modello quantistico può essere realizzato nello spazio di Hilbert complesso \mathbb{C}^3 , e formule analoghe alle equazioni (26)–(29) valgono con θ_d al posto di θ_c [17].

Si consideri nuovamente l'esemplare Occhiali da Sole rispetto ai concetti Abbigliamento Sportivo, Attrezzatura Sportiva e alla disgiunzione Abbigliamento Sportivo o Attrezzatura Sportiva. Poiché $\mu(A) + \mu(B) = 0.40 + 0.20 = 0.60 < 1$, la rappresentazione quantistica richiede l'uso delle equazioni (26) and (27), con θ_d al posto di θ_c , e (32). In particolare, dall'Equazione (32) si ottiene:

$$\begin{aligned} \theta_d &= \arccos\left(\frac{2\mu(A \text{ o } B) - \mu(A) - \mu(B)}{\sqrt{\mu(A)\mu(B)}}\right) \\ &= 135.00^\circ. \end{aligned} \quad (34)$$

Pertanto i pesi di appartenenza $\mu(A) = 0.40$, $\mu(B) = 0.20$ e $\mu(A \text{ o } B) = 0.10$ sono rappresentati da $\theta_d = 135.00^\circ$, $|A\rangle = (0.77, 0, 0.63)$ e $|B\rangle = e^{i135.00^\circ}(-0.37, -0.82, 0.45)$ nel modello quantistico. Perciò, la doppia sotto-estensione dell'esemplare Occhiali da Sole rispetto alla disgiunzione Abbigliamento Sportivo o

Attrezzatura Sportiva può essere spiegata come un effetto di interferenza distruttiva ($90^\circ < \theta_d = 135.00^\circ < 180^\circ$) tra i concetti componenti Abbigliamento Sportivo and attrezzatura Sportiva nella sovrapposizione Abbigliamento Sportivo o attrezzatura Sportiva.

Le formule quantistiche per la probabilità in (23) e (31) permettono di rappresentare la maggior parte dei dati di Hampton sulle congiunzioni e disgiunzioni di due concetti. Questo modello non è tuttavia in grado di rappresentare combinazioni di due concetti in tutta la loro generalità (si veda la nota 7). Infatti, in questo caso è necessaria una struttura più generale, a causa di ragioni che diventeranno chiare nella sezione finale. Questa struttura più generale è uno spazio di Fock a due settori \mathcal{F} , cioè, $\mathcal{F} = \mathcal{H} \oplus (\mathcal{H} \otimes \mathcal{H})$ (\oplus denota la somma diretta di spazi vettoriali), dove lo spazio di Hilbert individuale \mathcal{H} è chiamato 'settore 1 di \mathcal{F} ', mentre lo spazio di Hilbert prodotto tensoriale $\mathcal{H} \otimes \mathcal{H}$ è chiamato settore 2 di \mathcal{F} [17, 28, 37, 49].

Una prospettiva teorica generale

Partendo dalla modellizzazione della congiunzione e della disgiunzione di due concetti presentate nelle due sezioni precedenti, è stata elaborata una prospettiva teorica generale per rappresentare i concetti naturali e le loro combinazioni che utilizza gli spazi di Fock. Tale prospettiva generale ha permesso di riprodurre i dati sperimentali ottenuti in diversi test sulla combinazione dei concetti e ha mostrato che le deviazioni da strutture classiche osservate in questi test può essere spiegata in termini di effetti genuinamente quantistici, per esempio, contestualità, emergenza, interferenza e sovrapposizione, si veda la Figura 1. La stessa prospettiva è stata applicata con successo all'effetto *Guppy* [28] e alle contraddizioni *borderline* [37].

È stata inoltre identificata una violazione sperimentale delle disuguaglianze di Bell in una specifica combinazione concettuale e ne è stato elaborato un modello in spazi di Hilbert, mostrando che la violazione può essere spiegata in termini di *entanglement* tra i concetti componenti [46, 47]. Infine, è stato identificato un

fenomeno di indistinguibilità quantistica di tipo Bose-Einstein in specifiche combinazioni di concetti identici, per esempio, Undici Animali [48].

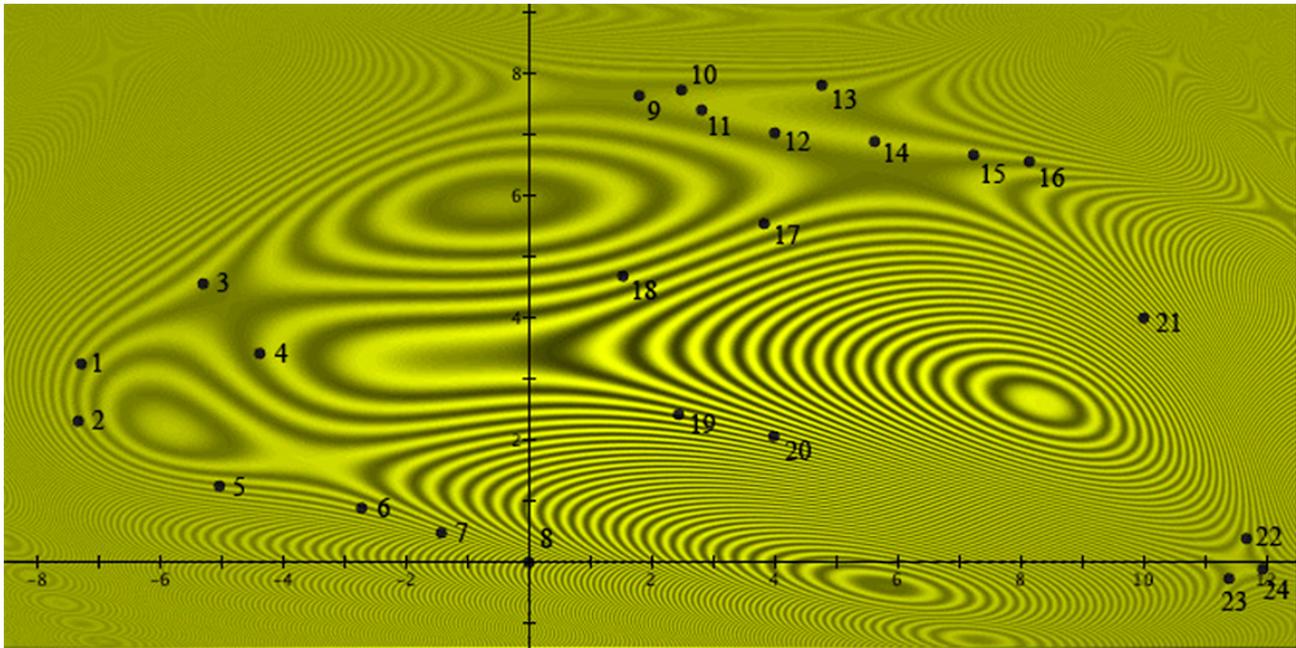
Più recentemente, i test di Hampton sono stati estesi a congiunzioni e negazioni di due concetti naturali [49, 50]. Più precisamente, i pesi di appartenenza di diversi esemplari sono stati testati rispetto ai concetti A , B , alle loro negazioni "non A " e "non B ", e alle congiunzioni " A e B ", " A e non B ", "non A e B " e "non A e non B ". In questi nuovi test, sono stati trovati sistematici effetti di sovra-estensione e doppia sovra-estensione in tutte le congiunzioni " A e B ", " A e non B ", "non A e B " e "non A e non B ", come previsto. Tuttavia, è stato anche scoperto un nuovo effetto non-classico. Infatti, è stata identificata una inaspettata, sistematica e significativa deviazione dalla legge marginale della probabilità Kolmogoroviana. Si consideri, ad esempio, la quantità

$$\mu(A \text{ e } B) + \mu(A \text{ e non } B) - \mu(A), \quad (35)$$

il cui valore dovrebbe essere identicamente nulla in un modello di probabilità Kolmogoroviana, per ogni esemplare e per ogni coppia di concetti. Dal momento che i pesi di appartenenza non possono in generale essere modellizzati da probabilità Kolmogoroviane, come si è visto nelle sezioni precedenti, ci si aspetterebbe che il valore dell'espressione nell'Equazione (35) sia differente da 0 e dipenda dall'esemplare e dalla coppia di concetti. Si è invece trovato che la seguente uguaglianza risulta soddisfatta:

$$\mu(A \text{ e } B) + \mu(A \text{ e non } B) - \mu(A) \approx \frac{1}{2}. \quad (36)$$

Inoltre, si è trovato che il valore numerico non dipende né dall'esemplare né dalla coppia di concetti che sono testati. Questo comportamento altamente non-classico è esattamente previsto nel modello generale in spazi di Fock per la congiunzione e la negazione di due concetti [50, 51]. Ne segue che la prospettiva teorica generale che abbiamo presentato in queste sezioni può essere ulteriormente testata e permette di prevedere nuovi effetti non-classici nella combinazione dei concetti che non erano stati osservati negli studi precedenti sull'argomento.



1 Almond	5 Coconut	9 Mustard	13 Garlic	17 Green Pepper	21 Broccoli
2 Acom	6 Raisin	10 Wheat	14 Mushroom	18 Yam	22 Rice
3 Peanut	7 Elderberry	11 Root Ginger	15 Watercress	19 Tomato	23 Parsley
4 Olive	8 Apple	12 Chili Pepper	16 Lentils	20 Pumpkin	24 Black Pepper

Figura 1: Effetti di interferenza ottenuti nella disgiunzione *Frutta o Verdura* dei concetti naturali *Frutta e Verdura*. Nella figura si riconoscono le frange di interferenza che sono tipicamente osservate nell'esperimento quantistico della doppia fenditura. Si veda [20], Figura 3, per i dettagli su come la figura è stata ottenuta.

Applicazioni a note fallacie cognitive

Altri fenomeni sono stati identificati nelle scienze cognitive dove l'applicazione di strutture matematiche classiche conduce a risultati in disaccordo con i dati sperimentali. In particolare:

(i) le fallacie della congiunzione e della disgiunzione costituiscono esempi di situazioni in cui i giudizi probabilistici non sono in accordo con la legge di monotonicità (5) della probabilità Kolmogoroviana.

(ii) l'effetto disgiunzione costituisce un esempio di situazione in cui le decisioni in presenza di incertezza non sono in accordo con la legge della probabilità totale (6) in probabilità Kolmogoroviana.

Mostreremo nelle prossime due sezioni che il formalismo quantistico in spazi di Hilbert usato per modellizzare la congiunzione e la disgiunzione di due concetti può essere applicato con successo alla modellizzazione sia delle fallacie della congiunzione e della disgiunzione al punto (i), che dell'effetto disgiunzione al punto (ii). Tali

effetti possono quindi essere spiegati in termini di strutture genuinamente quantistiche, come quelle identificate nelle sezioni precedenti.

Le fallacie della congiunzione e della disgiunzione

Tversky e Kahneman scoprirono nel 1983 la fallacia della congiunzione in un test che conteneva la seguente storia riguardante una ipotetica donna chiamata Linda.

“Linda ha 31 anni, è single, estroversa e molto brillante. Ha una laurea in filosofia. Da studentessa, era profondamente impegnata in questioni di discriminazione e giustizia sociale e ha anche partecipato a manifestazioni antinucleari.” [7]

Ai partecipanti veniva chiesto quale delle seguenti alternative essi giudicassero più probabile:

- (1) Linda è una cassiera di banca;
- (2) Linda è una cassiera di banca ed è attiva nel movimento femminista.

Complessivamente, l'85% dei partecipanti giudicava l'opzione (2) più probabile dell'opzione (1). Tale risultato sperimentale è in disaccordo con la legge di monotonicità della probabilità Kolmogoroviana relativamente alla congiunzione (Equazione (5)), donde la fallacia della congiunzione.

Esiste una vasta evidenza empirica che conferma i risultati di Tversky e Kahneman sull'esistenza della fallacia della congiunzione (si veda, per esempio, [3, 52, 53]), e numerose ipotesi sono state proposte per spiegarne l'origine. Nessuna di esse è però esente da critiche. Per esempio, Tversky e Kahneman ipotizzarono l'esistenza di *bias* cognitivi ed euristiche nei giudizi individuali. Tale proposta risulta, tra l'altro, compatibile con la teoria della *bounded rationality* proposta da Herbert Simon negli anni Cinquanta del secolo scorso [54]. Più precisamente, Tversky e Kahneman suggerirono l'euristica della rappresentatività come possibile spiegazione: gli individui giudicherebbero l'evento "essere femminista e cassiera di banca" più rappresentativo dell'evento "essere cassiera di banca" per le attività di Linda. Tale ipotesi funziona a livello intuitivo, tuttavia è necessario ipotizzare differenti euristiche per adattare diversi tipi di fallacie [52]. Una proposta alternativa suggerisce l'ipotesi dell'incomprensione: gli individui fraintenderebbero il problema di Linda nel test, nel senso che non interpreterebbero correttamente i termini "e" e/oppure "probabilità" in esso. Si osserva, tuttavia, che la fallacia si manifesta anche quando tali termini non vengono esplicitamente menzionati nel test [53]. Un primo tentativo di applicare strutture quantistiche viene dalla ipotesi degli effetti di ordine: gli individui giudicherebbero l'enunciato "Linda è una femminista ed è una cassiera di banca" come una sequenza ordinata, prima "Linda è una femminista", poi "Linda è una cassiera di banca". Secondo i proponenti, i due enunciati asseriscono proprietà incompatibili nel senso quantistico usuale, il che genera la fallacia della congiunzione.⁸ Nuovi esperimenti sembrano tuttavia confermare che la fallacia è

⁸Nel gergo quantistico, due proprietà a e b si dicono incompatibili se sono rappresentate da operatori di proiezione ortogonale M_a e M_b , rispettivamente, tali che $M_a M_b \neq M_b M_a$. Questa condizione implica che la probabilità di trovare a e b in una misura dipende in generale dall'ordine con cui a e b sono misurate.

indipendente da tali effetti di ordine [3, 18, 21].

Le osservazioni fatte sopra indicano che la fallacia della congiunzione è ancora un problema aperto in psicologia cognitiva.

Un esperimento più completo venne eseguito da Morier e Borgida nel 1984 [8]. Ai partecipanti veniva chiesto di giudicare la probabilità dei seguenti eventi:

A : Linda è una femminista;

B : Linda è una cassiera di banca;

' A e B ': Linda è una femminista e una cassiera di banca;

' A o B ': Linda è una femminista o una cassiera di banca.

Le rispettive probabilità erano ordinate nel modo seguente:

$$\mu(A) = 0.83 > 0.60 = \mu(A \text{ o } B), \quad (37)$$

$$\mu(A \text{ e } B) = 0.36 > 0.26 = \mu(B). \quad (38)$$

L'Equazione (37) esprime la fallacia della congiunzione, mentre l'Equazione (38) esprime una nuova fallacia, relativa alla disgiunzione e nota in letteratura con il nome di fallacia della disgiunzione [8]. Il test metteva quindi in evidenza l'esistenza simultanea di due fallacie, una per la congiunzione e una per la disgiunzione.

Nella prospettiva generale per i sistemi concettuali presentata nelle sezioni precedenti, la fallacia della congiunzione viene considerata come un effetto di sovra-estensione rispetto alla congiunzione di due concetti. Utilizzando la terminologia tipica della teoria dei concetti, infatti, lo studio della fallacia della congiunzione viene ricondotto al problema dell'appartenenza dell'esemplare Linda ai concetti Femminista (A), Cassiera di Banca (B) e alla loro congiunzione Femminista e Cassiera di Banca (' A e B '). Sia $\mu(A)$, $\mu(B)$ e $\mu(A \text{ e } B)$ la probabilità che Linda sia giudicata femminista, cassiera di banca e femminista e cassiera di banca, rispettivamente, in un test cognitivo. Appliciamo alla fallacia della congiunzione il modello quantistico in spazi di Hilbert per la congiunzione di due concetti, come segue [44, 45].

Si considerino le Equazioni (37) e (38). Poiché $\mu(A \text{ e } B) = 0.36$ e $\mu(A) + \mu(B) = 0.83 + 0.26 = 1.09 > 1$, si possono usare le equazioni (20)–(23),

(25), (28) and (29), ottenendo

$$\begin{aligned}\theta_c &= 121.44^\circ, \\ |A\rangle &= (0.91, 0, 0.41), \\ |B\rangle &= e^{i121.44^\circ} (0.39, 0.33, -0.86),\end{aligned}\quad (39)$$

nello spazio di Hilbert \mathbb{C}^3 , dove il vettore unitario $|A\rangle$ rappresenta il concetto Femminista, il vettore unitario $|B\rangle$ rappresenta il concetto Cassiera di Banca e θ_c è l'angolo di interferenza per la congiunzione.

Il modello quantistico per la congiunzione di due concetti permette di riprodurre fedelmente i dati di vari test sulle fallacie congiuntive [44, 55, 56]. Esso permette inoltre di spiegare la fallacia della congiunzione assumendo che, quando si considera l'esemplare Linda, insieme alla sua storia, i concetti Femminista e Cassiera di Banca interferiscono distruttivamente nella congiunzione Femminista e Cassiera di Banca, inteso come un nuovo emergente sistema concettuale. A differenza di altri modelli quantistici, questo modello predice anche l'esistenza di doppie fallacie della congiunzione, cioè di situazioni in cui $\mu(A \text{ e } B) > \max[\mu(A), \mu(B)]$, in analogia con gli effetti di doppia sovra-estensione osservati nella congiunzione di due concetti [18, 44].

Analogamente, la fallacia della disgiunzione viene considerata come un effetto di sotto-estensione rispetto alla disgiunzione di due concetti, e lo studio della fallacia viene ricondotto al problema dell'appartenenza dell'esemplare Linda ai concetti Femminista (A), Cassiera di Banca (B) e alla loro disgiunzione Femminista o Cassiera di Banca ($A \text{ o } B$). Siano $\mu(A)$, $\mu(B)$ e $\mu(A \text{ o } B)$ la probabilità che Linda sia giudicata femminista, cassiera di banca e femminista o cassiera di banca, rispettivamente, in un test cognitivo. Appliciamo alla fallacia della disgiunzione il modello quantistico in spazi di Hilbert per la disgiunzione di due concetti, come segue [44, 45].

Poiché $\mu(A \text{ o } B) = 0.60$ e $\mu(A) + \mu(B) = 0.83 + 0.26 = 1.09 > 1$, si possono utilizzare le equazioni (20), (21), (28) e (29) con θ_d al posto

di θ_c , (31) e (33), ottenendo:

$$\begin{aligned}\theta_d &= 81.08^\circ \\ |A\rangle &= (0.91, 0, 0.41) \\ |B\rangle &= e^{i81.08^\circ} (0.39, 0.33, -0.86)\end{aligned}\quad (40)$$

nello spazio di Hilbert \mathbb{C}^3 , dove il vettore unitario $|A\rangle$ rappresenta il concetto Femminista, il vettore unitario $|B\rangle$ rappresenta il concetto Cassiera di Banca e θ_d è l'angolo di interferenza per la disgiunzione.

Il modello quantistico per la disgiunzione di due concetti permette di riprodurre fedelmente i dati di vari test sulle fallacie disgiuntive [44, 55, 56]. Esso permette inoltre di spiegare la fallacia della disgiunzione assumendo che, quando si considera l'esemplare Linda, insieme alla sua storia, i concetti Femminista e Cassiera di Banca interferiscono costruttivamente nella disgiunzione Femminista o Cassiera di Banca, inteso come un nuovo emergente sistema concettuale. A differenza di altri modelli quantistici, questo modello predice anche l'esistenza di doppie fallacie della disgiunzione, cioè di situazioni in cui $\mu(A \text{ o } B) < \min[\mu(A), \mu(B)]$, in analogia con gli effetti di doppia sotto-estensione osservati nella disgiunzione di due concetti [18, 44].

L'analisi che precede mostra che strutture quantistiche emergono nella modellizzazione di effetti cognitivi che sono altamente problematici se si utilizzano strutture classiche. La medesima conclusione può essere ottenuta studiando l'effetto disgiunzione.

L'effetto disgiunzione

L'effetto disgiunzione è un esempio di errore decisionale (si veda il punto (b) nella sezione introduttiva) e mostra che le decisioni umane non sono in generale in accordo con il principio della cosa certa, uno degli assiomi fondamentali della teoria dell'utilità attesa di Leonard Savage [57].

Il principio della cosa certa venne suggerito a Savage dalla seguente storia.

“Un uomo d'affari considera l'acquisto di un determinato immobile. Egli ritiene che l'esito delle prossime elezioni presidenziali sia rilevante per la de-

cisione sull'acquisto. Quindi, [...], si chiede se comprerebbe se sapesse che il candidato democratico avrebbe vinto, e scopre che lo farebbe. Allo stesso modo, egli si chiede se comprerebbe se sapesse che il candidato repubblicano avrebbe vinto, e scopre che ancora lo farebbe. Vedendo che avrebbe comprato in entrambi i casi, decide che dovrebbe comprare, anche se non sa quale evento si verificherà." [57]

Tversky e Shafir testarono la validità del principio della cosa certa in un test che consisteva nel proporre ad un gruppo di studenti una scommessa in due fasi, cioè una scommessa che può essere giocata due volte [9]. Ad ogni stadio del test, la decisione consisteva nel giocare o meno una scommessa che ha le stesse probabilità di vincere, diciamo \$200, o di perdere, diciamo \$100. Il risultato chiave si basa sulla decisione riguardo alla seconda scommessa al termine della prima scommessa.

Il test includeva tre situazioni differenti:

(i) gli studenti venivano informati di aver vinto la prima scommessa;

(ii) gli studenti venivano informati di aver perso la prima scommessa;

(iii) gli studenti non conoscevano l'esito della prima scommessa.

Tversky e Shafir trovarono che il 69%, ovvero la maggioranza, degli studenti che sapevano di aver vinto la prima scommessa decisero di giocare nuovamente, il 59%, ancora la maggioranza, degli studenti che sapevano di aver perso la prima scommessa decisero di giocare nuovamente, mentre solo il 36% degli studenti che non sapevano se avessero vinto o perso decisero di giocare nuovamente (equivalentemente, il 64%, ovvero la maggioranza, decise di non giocare la seconda scommessa).

Il test della scommessa in due fasi è in disaccordo con il principio della cosa certa: gli studenti preferiscono in generale giocare di nuovo se sanno di aver vinto, e preferiscono giocare di nuovo anche se sanno di aver perso, ma generalmente preferiscono non giocare di nuovo quando non sanno se hanno vinto o perso. Tversky e Shafir ottennero lo stesso risultato in un altro test, simile alla scommessa in due fasi, oggi noto con il nome di problema Hawaii [9]. Inoltre, vio-

lazioni del principio della cosa certa sono state sistematicamente osservate in test simili, eseguiti più recentemente [58, 59].⁹

Più in generale, i risultati del test di Tversky e Shafir sono in disaccordo con la legge della probabilità totale (6) in probabilità Kolmogoroviana. Infatti, sia $p(P)$ la probabilità totale che lo studente decida di giocare nuovamente senza sapere se ha vinto o ha perso la prima scommessa, sia $p(W)$ (rispettivamente, $p(L) = 1 - p(W)$) la probabilità che lo studente vinca (rispettivamente, perda) la prima scommessa, e sia $p(P|W)$ (rispettivamente, $p(P|L)$) la probabilità condizionale che lo studente decida di giocare nuovamente sapendo di aver vinto (rispettivamente, perso) la prima scommessa. Si prova allora facilmente che non è possibile trovare una probabilità $p(W)$ in modo che le probabilità in [9] siano soddisfatte, cioè, $p(P) = 0.36$, $p(P|W) = 0.69$, $p(P|L) = 0.59$, e la legge della probabilità totale (6) sia soddisfatta, donde l'effetto disgiunzione. Come spiegazione dell'effetto, Tversky e Shafir suggerirono che le persone provano in genere avversione all'incertezza, ovvero preferiscono le scelte sicure a quelle insicure.

Nell'ambito della prospettiva generale che è stata presentata nelle sezioni precedenti e che adotta il formalismo quantistico in spazi di Hilbert, è possibile trattare l'effetto disgiunzione come un effetto di sotto-estensione rispetto alla disgiunzione di due concetti [43, 44]. A tal fine, sia A la situazione concettuale corrispondente a *Aver Vinto la Prima Scommessa*, sia B la situazione concettuale corrispondente a *Aver Perso la Prima Scommessa*, e sia " A o B " la situazione concettuale corrispondente a *Aver Vinto o Perso la Prima Scommessa*. Dai dati in [9], si ha $\mu(A \text{ o } B) = 0.36$ e $\mu(A) + \mu(B) = 0.69 + 0.59 = 1.28 > 1$. Quindi, si possono utilizzare le equazioni (20), (21), (28) e (29) con θ_d al posto di θ_c , (31) e (33), ottenendo:

$$\begin{aligned}\theta_d &= 141.76^\circ, \\ |A\rangle &= (0.83, 0, 0.56), \\ |B\rangle &= e^{i141.76^\circ} (0.43, 0.64, -0.64),\end{aligned}\quad (41)$$

nello spazio di Hilbert \mathbb{C}^3 , dove i vettori unitari e ortogonali $|A\rangle$ e $|B\rangle$ rappresentano le situazioni

⁹È interessante osservare che anche il dilemma del prigioniero, famoso in teoria dei giochi, può essere formulato nella forma di effetto disgiunzione [3].

concettuali A e B , rispettivamente, la combinazione lineare normalizzata $\frac{1}{\sqrt{2}}(|A\rangle + |B\rangle)$ rappresenta la disgiunzione " A o B ", e θ_d è l'angolo di interferenza per la disgiunzione.

Il modello quantistico per la disgiunzione di due concetti permette di riprodurre fedelmente i risultati di altri test sull'effetto disgiunzione [44, 58, 59]. Esso permette inoltre di spiegare l'effetto assumendo che le situazioni concettuali *Aver Vinto la Prima Scommessa* e *Aver Perso la Prima Scommessa* interferiscono distruttivamente nella disgiunzione *Aver Vinto o Perso la Prima Scommessa*, inteso come una nuova emergente situazione concettuale.

Una spiegazione unitaria

Si è visto nelle sezioni precedenti che, sia gli effetti di sovra-/sotto-estensione nella combinazione dei concetti, che le fallacie della congiunzione/disgiunzione e l'effetto disgiunzione rivelano la presenza di strutture genuinamente quantistiche in fenomeni cognitivi in cui è presente un certo tipo di incertezza. Tali strutture sono anche sistematicamente presenti in altri domini empirici collegati con le scienze cognitive, in particolare, la teoria della decisione e la *information retrieval*. A tal proposito, la prospettiva teorica sviluppata nell'ambito della collaborazione tra i gruppi di ricerca di Bruxelles e Leicester è stata applicata ad entrambi questi domini.

In teoria della decisione, è stato elaborato un modello generale che usa il formalismo quantistico in spazi di Hilbert per rappresentare preferenze e decisioni in condizioni di incertezza (si veda, per esempio, [25]). Secondo la già citata teoria dell'utilità attesa, gli individui dovrebbero prendere la decisione che massimizza la loro utilità rispetto ad una misura di probabilità Kolmogoroviana. Esistono situazioni concrete in cui le decisioni umane non sono invece in accordo con le previsioni della teoria dell'utilità attesa. Tale disaccordo è illustrato, per esempio, dai noti paradossi di Allais, Ellsberg e Machina [10, 11], i quali rivelano che specifici aspetti psicologici, quali l'avversione all'incertezza, guidano le decisioni in tali situazioni. Nel quadro della prospettiva quantistica generale, si è invece mostrato che gli individui prendono effettivamente la decisione che massimizza la lo-

ro utilità attesa, ma tale utilità è riferita rispetto ad una misura di probabilità quantistica, la quale incorpora gli aspetti psicologici di cui sopra. Tale prospettiva teorica permette di rappresentare i risultati di numerosi test su decisioni in presenza di incertezza, e costituisce il primo passo verso l'elaborazione di un'estensione di tipo quanto-meccanico della teoria dell'utilità attesa [25, 60].

Il successo del formalismo quantistico in spazi di Hilbert nella modellizzazione dei sistemi concettuali si può estendere al *World Wide Web* (WWW), quando quest'ultimo venga considerato come un sistema concettuale complessivo. È stata, in particolare, elaborata una descrizione in termini di stato, contesto e proprietà delle parole, dei concetti e dei documenti presenti nel WWW. Sulla base di tale descrizione, si è sviluppato un modello quantistico in spazi di Hilbert per rappresentare il significato di tali parole, concetti e documenti. Utilizzando tecniche di *information retrieval* e opportuni *corpora* di documenti, le fallacie della congiunzione e della disgiunzione ed altri effetti non-classici sono stati identificati anche nel WWW, ed una spiegazione in termini di strutture quantistiche è stata effettivamente proposta [61]. Il presente approccio estende a spazi di Hilbert complessi gli approcci tradizionali in *information retrieval* che utilizzano spazi vettoriali reali. Tale prospettiva evidenzia l'importanza dei numeri complessi nella rappresentazione dei processi informatici (si veda, per esempio, [23]).

Concludiamo questo articolo con un commento sulla possibilità di fornire una spiegazione unitaria per le strutture quantistiche identificate nei processi cognitivi e decisionali. Infatti, si può formulare la seguente ipotesi esplicativa sui meccanismi che sono alla base del ragionamento umano: il ragionamento è una sovrapposizione di due processi strutturati, un processo di tipo logico e un processo di tipo emergente. Il ragionamento logico combina sistemi concettuali, cioè, concetti, combinazioni di concetti, proposizioni, e anche sistemi decisionali complessi, applicando le regole della logica Booleana, in alcuni casi arricchita da una semantica probabilistica Kolmogoroviana. Il ragionamento emergente consente invece la formazione di nuovi sistemi concettuali, nel caso di concetti, nuovi concetti, nel caso di proposizioni, nuove proposizioni, ecc.

Il significato di tali sistemi concettuali emergenti è legato a quello dei sistemi concettuali componenti, ma non obbedisce alle regole di combinazione della logica Booleana. Questi due meccanismi agiscono in sovrapposizione nel pensiero umano durante un processo di ragionamento, il primo essendo guidato dalle regole della logica classica, il secondo seguendo un meccanismo non-classico di emergenza.

La struttura a due livelli del ragionamento può essere formalizzata nello spazio di Fock introdotto nelle sezioni precedenti, dove sono attivi solo i primi due settori, e gli stati dei sistemi concettuali sono rappresentati da vettori unitari di questo spazio di Fock a due settori. Più precisamente, il settore 1 dello spazio di Fock, formato da uno spazio di Hilbert individuale, modella l'emergenza concettuale, quindi la combinazione di due concetti è rappresentata da una sovrapposizione dei vettori che rappresentano i concetti componenti in questo spazio di Hilbert, il che spiega come fenomeni di interferenza quantistica possano essere coinvolti in un processo di emergenza. Il settore 2 dello spazio di Fock, cioè un prodotto tensoriale di due copie identiche di questo spazio di Hilbert, modella una combinazione di concetti a partire dai concetti componenti richiedendo che le regole della logica per il connettivo logico usato per la combinazione siano soddisfatte.

Il seguente esempio fornisce un significativo supporto a questa ipotesi [28, 44, 51].

Hampton misurò i pesi di appartenenza dell'esemplare Oliva rispetto ai concetti Frutta, Verdura e la loro disgiunzione Frutta o Verdura, trovando 0.50, 0.10 e 0.80, rispettivamente [6], come si è visto nelle sezioni precedenti. L'Equazione (18) è violata in questo caso, perciò i dati sono non-classici rispetto alla disgiunzione. Recentemente, i pesi di appartenenza dell'esemplare Oliva sono stati testati rispetto ai Frutta, Verdura e la loro congiunzione Frutta e Verdura, trovando 0.56, 0.63 and 0.65, rispettivamente [49]. L'Equazione (12) è violata in questo caso, e i dati sono non-classici rispetto alla congiunzione. Tuttavia, si noti che l'esemplare Oliva è doppiamente sovra-esteso in entrambi i casi, con simili pesi di appartenenza rispetto alla disgiunzione e alla congiunzione. Questo

risultato suggerisce che esistono situazioni in cui gli individui non considerano se si usa il connettivo 'o' oppure il connettivo 'e', ma essi giudicano se l'esemplare è membro del nuovo concetto emergente, ottenuto combinando i due concetti componenti, ma in un modo che è indipendente dal fatto che la combinazione è ottenuta mediante congiunzione o disgiunzione. Questo è un chiaro esempio di processo in cui l'emergenza concettuale domina nettamente il processo di combinazione logica.



- [1] J. R. Busemeyer, A. Diederich *Cognitive Modeling*, Sage Publications Inc., Los Angeles (2009).
- [2] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitrechnung*, Ergebnisse Der Mathematik (1933); translated as *Foundations of Probability*, Chelsea Publishing Company, New York (1950).
- [3] J. R. Busemeyer, P. D. Bruza, *Quantum Models of Cognition and Decision*, Cambridge University Press, Cambridge (2012).
- [4] E. Haven, A. Y. Khrennikov, *Quantum Social Science*, Cambridge University Press, Cambridge (2013).
- [5] J. A. Hampton, Overextension of conjunctive concepts: Evidence for a unitary model for concept typicality and class inclusion, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14 (1988a) 12–32.
- [6] J. A. Hampton, Disjunction of natural concepts, *Memory and Cognition* 16 (1988b) 579–591.
- [7] A. Tversky, D. Kahneman, Extension versus intuitive reasoning: The conjunction fallacy in probability judgment, *Psychological Review* 90 (1983) 293–315.
- [8] D. Morier, E. Borgida, The conjunction fallacy: A task specific phenomenon?, *Personality and Social Psychology Bulletin* 10 (1984) 243–252.
- [9] A. Tversky, E. Shafir, The disjunction effect in choice under uncertainty, *Psychological Science* 3 (1992) 305–309.
- [10] D. Ellsberg, Risk, ambiguity, and the Savage axioms, *Quarterly Journal of Economic* 75 (1961) 643–669.
- [11] M. J. Machina, Risk, ambiguity, and the dark-dependence axioms, *American Economic Review* 99 (2009) 385–392.
- [12] D. Aerts, A possible explanation for the probabilities of quantum mechanics, *Journal of Mathematical Physics* 27 (1986) 202–210.
- [13] I. Pitowsky, *Quantum Probability, Quantum Logic*, Lecture Notes in Physics vol. 321, Springer, Berlin (1989).
- [14] D. Aerts, S. Aerts, Applications of quantum statistics in psychological studies of decision processes, *Foundations of Science* 1 (1995) 85–97.

- [15] D. Aerts, L. Gabora, A theory of concepts and their combinations I: The structure of the sets of contexts and properties, *Kybernetes* 34 (2005a) 167–191.
- [16] D. Aerts, L. Gabora, A theory of concepts and their combinations II: A Hilbert space representation, *Kybernetes* 34 (2005b) 192–221.
- [17] D. Aerts, Quantum structure in cognition, *Journal of Mathematical Psychology* 53 (2009) 314–348.
- [18] J. R. Busemeyer, E. M. Pothos, R. Franco, J. S. Trueblood, A quantum theoretical explanation for probability judgment errors, *Psychological Review* 118 (2011) 193–218.
- [19] D. Aerts, J. Broekaert, L. Gabora, S. Sozzo, Quantum structure and human thought, *Behavioral and Brain Sciences* 36 (2013) 274–276.
- [20] D. Aerts, L. Gabora, S. Sozzo, Concepts and their dynamics: A quantum–theoretic modeling of human thought. *Topics in Cognitive Science* 5 (2013) 737–772.
- [21] E. M. Pothos, J. R. Busemeyer, Can quantum probability provide a new direction for cognitive modeling?, *Behavioral and Brain Sciences* 36 (2013) 255–274.
- [22] Z. Wang, T. Solloway, R. M. Shiffrin, J. R. Busemeyer, Context effects produced by question orders reveal quantum nature of human judgments, *Proceedings of the National Academy of Sciences* 111 (2014) 9431–9436.
- [23] M. Melucci, *Introduction to Information Retrieval and Quantum Mechanics*, Springer, Berlin (2015).
- [24] P. D. Kvam, T. J. Pleskac, S. Yu, J. R. Busemeyer, Interference effects of choice on confidence: Quantum characteristics of evidence accumulation, *Proceedings of the National Academy of Sciences* 112 (2016) 10645–10650.
- [25] D. Aerts, E. Haven, S. Sozzo, A proposal to extend expected utility in a quantum probabilistic framework, *Economic Theory* 65 (2018) 1079–1109.
- [26] A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases, *Science* 185 (1974) 1124–1131.
- [27] D. Aerts, S. Sozzo, T. Veloz, Quantum structure in cognition and the foundations of human reasoning, *International Journal of Theoretical Physics* 54 (2015a) 4557–4569.
- [28] D. Aerts, S. Sozzo, Quantum structure in cognition: Origins, developments, successes and expectations, *The Palgrave Handbook of Quantum Models in Social Science: Applications and Grand Challenges*, E. Haven, A. Y. Khrennikov Eds., Palgrave & Macmillan, London, pp. 157–193 (2016).
- [29] P. Samuelson, A quantum theory model of economics: Is the co-ordinating entrepreneur just worth his profit?, *The collected scientific papers of Paul A. Samuelson 4*, MIT Press, Cambridge, MA, pp. 104–110 (1979).
- [30] A. Sen, Maximization and the act of choice, *Econometrica* 65 (1997) 745–779.
- [31] D. Kahneman, A. Tversky, *Choice, Values and Frames*, Cambridge University Press, Cambridge (2000).
- [32] E. Rosch, Prototype classification and logical classification: The two systems, *New Trends in Conceptual Representation: Challenges to Piaget Theory?*, E. K. Scholnick Ed., Lawrence Erlbaum, New Jersey, pp. 133–159 (1983).
- [33] J. A. Hampton, Categories, prototypes and exemplars, *Routledge Handbook of Semantics*, N. Riemer Ed., Routledge, London, pp. 125–141 (2016).
- [34] L. Zadeh, A note on prototype theory and fuzzy sets, *Cognition* 12 (1982) 291–297.
- [35] D. Osherson, E. Smith, On the adequacy of prototype theory as a theory of concepts, *Cognition* 9 (1981) 35–58.
- [36] S. Alxatib, J. Pelletier, On the psychology of truth gaps, *Vagueness in Communication*, R. Nouwen, R. van Rooij, U. Sauerland, H.-C. Schmitz Eds., Springer-Verlag, Berlin, Heidelberg pp. 13–36 (2011).
- [37] S. Sozzo, A quantum probability explanation in Fock space for borderline contradictions, *Journal of Mathematical Psychology* 58 (2014) 1–12.
- [38] J. M. Jauch, *Foundations of Quantum Mechanics*, Addison Wesley, Reading, MA (1968).
- [39] C. Piron, *Foundations of Quantum Physics*, Reading, Reading, MA (1976).
- [40] D. Aerts, Foundations of quantum physics: A general realistic and operational approach, *International Journal of Theoretical Physics* 38 (1999) 289–358.
- [41] D. Aerts, M. Sassoli de Bianchi, S. Sozzo, On the foundations of the Brussels operational-realistic approach to cognition, *Frontiers in Physics* doi 10.3389/fphy.2016.00017 (2016).
- [42] P. A. M. Dirac, *Quantum mechanics, 4th Ed.*, Oxford University Press, Oxford (1958).
- [43] S. Sozzo, Effectiveness of the quantum-mechanical formalism in cognitive modeling, *Soft Computing* 21 (2017) 1455–1465.
- [44] D. Aerts, M. Sassoli de Bianchi, S. Sozzo, T. Veloz, Modeling Human Decision-making: An Overview of the Brussels Quantum Approach, *Foundations of Science* doi 10.1007/s10699-018-9559-x (2018).
- [45] R. Pisano, S. Sozzo, A unified theory of human judgements and decision-making under uncertainty, *Entropy* 22 (2020) 738: doi 10.3390/e22070738.
- [46] D. Aerts, S. Sozzo, Quantum structure in cognition. Why and how concepts are entangled. *Quantum Interaction. Lecture Notes in Computer Science* 7052 (2011) 116–127., Springer Berlin.
- [47] D. Aerts, J. Arguëlles, L. Beltran, S. Geriente, M. Sassoli de Bianchi, S. Sozzo, V. Veloz, Quantum entanglement in physical and cognitive systems: A conceptual analysis and a general representation, *European Physical Journal Plus* 134 (2019) 493.
- [48] D. Aerts, S. Sozzo, T. Veloz, Quantum nature of identity in human concepts: Bose-Einstein statistics for conceptual indistinguishability, *International Journal of Theoretical Physics* 54 (2015b) 4430–4443.

- [49] S. Sozzo, Conjunction and negation of natural concepts: A quantum-theoretic modeling, *Journal of Mathematical Psychology* 66 (2015) 83–102.
- [50] D. Aerts, S. Sozzo, T. Veloz, Quantum structure of negation and conjunction in human thought, *Frontiers in Psychology*, doi 10.3389/fpsyg.2015.01447 (2015c).
- [51] D. Aerts, S. Sozzo, T. Veloz, New fundamental evidence of non-classical structure in the combination of natural concepts, *Philosophical Transactions of the Royal Society A* 374 (2915d) 20150095.
- [52] A. K. Shah, D. M. Oppenheimer, Heuristics made easy: An effort-reduction framework, *Psychological Bulletin* 134 (2008) 207–222.
- [53] R. Moro, On the nature of the conjunction fallacy, *Synthese* 171 (2009) 1–24.
- [54] H. A. Simon, A behavioral model of rational choice, *The Quarterly Journal of Economics* 69 (1955) 99–118.
- [55] J. E. Fisk, Judgments under uncertainty: Representativeness or potential surprise?, *British Journal of Psychology* 93 (2002) 431–449.
- [56] Y. Lu, The conjunction and disjunction fallacies: Explanations of the Linda problem by the equate-to-differentiate model, *Integrative Psychological and Behavioral Science* 50 (2015) 507–531.
- [57] L. Savage, *The Foundations of Statistics*, John Wiley & Sons, New York (1954).
- [58] A. Kühberger, D. Kamunská, J. Perner, The disjunction effect: Does it exist for two-step gambles?, *Organization Behavior and Human Decision Processes* 85 (2001) 250–264.
- [59] C. Lambdin, C. Burdsal, The disjunction effect reexamined: Relevant methodological issues and the fallacy of unspecified percentage comparisons, *Organization Behavior and Human Decision Processes* 103 (2007) 268–276.
- [60] D. Aerts, S. Geriente, C. Moreira, S. Sozzo, Testing ambiguity and Machina preferences within a quantum-theoretic framework for decision-making, *Journal of Mathematical Economics* 78 (2018) 176–185.
- [61] D. Aerts, J. Arguëlles, L. Beltran, L. Beltran, M. Sassoli de Bianchi, S. Sozzo, T. Veloz, Towards a quantum World Wide Web, *Theoretical Computer Science* 752 (2018) 116–131.

e informatiche. Sandro Sozzo è autore di un centinaio di pubblicazioni su riviste internazionali e ha tenuto lezioni nel Regno Unito (Oxford), Unione Europea, Stati Uniti, Canada, Cina e Nuova Zelanda. È inoltre segretario dell'associazione *International Quantum Structures Association* e direttore editoriale della rivista *Foundations of Science* della *Springer Nature*. Nel 2014, Sandro Sozzo ha fondato il *Centre for Quantum Social and Cognitive Science* (IQSCS), che attualmente dirige. Nell'ambito delle attività del centro, è uno dei responsabili del consorzio QUARTZ, finanziato dal programma *Horizon 2020* dell'Unione Europea.



Sandro Sozzo: Fisico (laurea e dottorato conseguiti presso l'Università di Lecce), è *Associate Professor* presso la *School of Business* della *University of Leicester*. I suoi interessi di ricerca riguardano principalmente i fondamenti delle scienze naturali e cognitive. In particolare, ha fornito contributi noti a livello internazionale nello sviluppo di modelli fisico-matematici in psicologia cognitiva, con applicazioni nelle scienze economiche

Big Data, profezie, assicurazione: una prospettiva sociologica

Alberto Cevolini

Dipartimento di Comunicazione ed Economia,
Università degli Studi di Modena e Reggio Emilia

A cosa servono i *Big Data*?

I *Big Data* rappresentano senza dubbio uno dei temi cruciali dei nostri tempi, un tema sul quale si sta concentrando non solo la ricerca scientifica ma anche il dibattito sociale. Tanta attenzione si deve al fatto che i *Big Data* potrebbero cambiare in modo significativo le nostre vite e avere un impatto rilevante pressoché su ogni settore della società, dalla medicina al diritto, dal controllo dell'ordine pubblico all'assicurazione. Questa invadenza più o meno silenziosa solleva ovviamente questioni etiche e politiche, ma anche pedagogiche e giuridiche che rientrano poi nel sistema scientifico come oggetto di ricerca empirica. È giusto quindi tenere alta l'attenzione e cercare di impostare il dibattito pubblico su delle solide basi matematiche, logiche e filosofiche.

Quando si parla di *Big Data*, si possono porre innanzitutto due domande, entrambe legittime e pertinenti: che cosa sono? E a che cosa servono? Qui vorrei provare a dire qualcosa che possa servire a dare una risposta almeno parziale a entrambe le domande, partendo tuttavia dalla seconda e lasciando per il momento in sospeso la prima.

I *Big Data* possono servire a fare molte cose differenti. Gli usi vanno dal riconoscimento facciale al controllo delle operazioni finanziarie, dalla

personalizzazione degli inserti pubblicitari all'identificazione di forme, schemi, o patterns - in breve: ridondanze - che permettono di fare distinzioni estremamente accurate, sebbene non infallibili (per esempio fra tessuti tumorali e tessuti non tumorali). C'è tuttavia un impiego dei *Big Data* che sta attirando un'enorme attenzione (e investimenti altrettanto consistenti): a quanto pare, con i *Big Data* si può predire il futuro¹.

Ovviamente i dati da soli non bastano. Bisogna, come si dice nel gergo tecnico, elaborarli. Questa elaborazione, a sua volta, non può avvenire in modo arbitrario, ma deve seguire delle istruzioni ben precise. L'insieme di queste istruzioni è quello che si definisce algoritmo. Per questo quando si parla di *Big Data* si parla anche pressoché inevitabilmente di algoritmi. Qui di nuovo ci si può domandare che cosa sono gli algoritmi e che cosa fanno².

Senza invadere il campo della matematica o dell'informatica, si può dire in modo molto ge-

¹Mayer-Schönberger e Cukier [1], p. 23 affermano esplicitamente che "nella loro essenza, i Big Data hanno a che fare con le previsioni".

²Un importante saggio sugli algoritmi che non richiede particolari competenze per essere letto è quello di Cardon [2]. Un testo più divulgativo è quello di McCormick [3]. Le tecniche di *data mining* (scavo dei dati) sono illustrate in modo magistrale nel volume di Provost e Fawcett [4].

nerale che gli algoritmi servono a scomporre dei problemi complessi in fasi più semplici. Per ciascuna fase l'algoritmo dà alla macchina (oppure a un personale umano) le istruzioni da seguire per fare qualcosa e arrivare a una conclusione che servirà poi come punto di partenza per la fase successiva. La successione delle fasi è fondamentale: essa deve essere stabilita in anticipo e non può essere modificata in corso d'opera, spostando per esempio una fase da un punto all'altro della serie. Anche una ricetta di cucina è un algoritmo e chi vuole cuocere la pasta, per esempio, sa bene che prima deve far bollire l'acqua e solo dopo può gettare la pasta - non viceversa. In ogni singola fase quindi si hanno delle istruzioni ben precise che vanno seguite scrupolosamente fino a ottenere un risultato, passando da una fase all'altra per arrivare alla soluzione del problema.

Quando l'algoritmo serve a istruire una macchina (un calcolatore, un computer), la macchina presenta delle caratteristiche particolari che ricordano quelle che in cibernetica si chiamano macchine non *triviali*. Da un lato la macchina è completamente deterministica, in quanto non può decidere arbitrariamente di cambiare le istruzioni da seguire, oppure di non fare niente quando si dice alla macchina di fare qualcosa. Nonostante questo la macchina è però anche del tutto imprevedibile: una volta inserito l'input non si può prevedere quale sarà l'output, perché nel frattempo la macchina ha elaborato appunto dei dati e nessuno può conoscere in anticipo il risultato di questa elaborazione, anche nel caso in cui conosca le istruzioni che sono state impartite alla macchina. Infine la macchina è del tutto dipendente dai dati con cui viene addestrata. Anche per questo la macchina è imprevedibile: nessuno ha un accesso diretto ai singoli dati che vengono inseriti nella macchina, tanto meno quando questi dati sono prodotti continuamente e in quantità astronomiche a partire dai comportamenti digitali degli utenti.

Una definizione così astratta dell'algoritmo si lascia facilmente generalizzare. Uscendo per un attimo dall'argomento di questo articolo, si può dire che anche l'albero di Porfirio è un algoritmo. La dialettica platonica, così come viene presentata per esempio nel *Sofista*, doveva servire a risolvere un problema complesso, quello della definizione. Questo problema è comples-

so per il semplice fatto che il senso in quanto tale è complesso. Per venire a capo di questa complessità, Platone aveva escogitato delle istruzioni altrettanto semplici quanto brillanti: in definitiva tutta la dialettica si riduce a due sole istruzioni: "dividi per due e tieni sempre la destra". In questo modo, partendo da una unità (la *synagogé*) che doveva essere scelta con molta cautela, si poteva procedere alla divisione binaria (la *diaeresis*), passando via via a divisioni successive finché non si fosse giunti a ciò che è indivisibile per definizione, cioè l'individuo. A questo punto, raccogliendo assieme tutti i risultati delle diverse fasi seguendo l'ordine della successione, si poteva dare la definizione dell'individuo (il pescatore, il sofista ecc.) preso in considerazione.

Tornando ora al nostro tema, se è vero che con i *Big Data* si possono fare molte cose e che per questo sono indispensabili degli algoritmi, allora è evidente che non tutti gli algoritmi sono uguali, ma la loro natura dipende dallo scopo per il quale vengono scritti. Poiché qui noi siamo interessati all'uso che si può fare dei *Big Data* per predire il futuro, gli algoritmi con i quali ci confrontiamo in modo particolare sono i così detti algoritmi predittivi. A questo proposito c'è un settore sociale che è particolarmente interessato a questo tipo di algoritmi per delle ragioni abbastanza ovvie: l'assicurazione. L'assicurazione infatti è sempre stata particolarmente avida di dati, da un lato, e lavora con predizioni, dall'altro. Vorrei soffermarmi rapidamente su questi due punti.

L'importanza dei dati per l'industria assicurativa

L'assicurazione ha sempre avuto un appetito vorace per i dati [5]. Il materiale grezzo dell'industria assicurativa in effetti non è l'acciaio o il carbone, ma sono i dati. Perché? Per rispondere a questa domanda ci dobbiamo chiedere prima di tutto che cos'è un dato. Questo ci offre l'occasione per recuperare la domanda che avevamo lasciato in sospeso.

In modo molto generale, possiamo dire che un dato è un riferimento alla realtà, cioè all'ambiente esterno. Se dico che la temperatura di questa stanza è 20°, ho già prodotto un dato. Potrei poi misurare la temperatura in momenti diversi del-

la giornata, in giorni e mesi diversi dell'anno, e compilare così una serie storica che mi permette di capire come varia la temperatura nel corso del tempo. Il punto fondamentale è che il riferimento alla realtà non è mai la realtà alla quale si fa riferimento. Il dato "20°" non è né caldo, né freddo. Il dato riferito alla temperatura, in altri termini, non ha una temperatura propria. Questa distinzione fra riferimento alla realtà e realtà alla quale si fa riferimento coincide con la distinzione fra un riferimento ad altro e un riferimento a sé nei sistemi autoreferenziali, come le coscienze o i sistemi di comunicazione, e costituisce il punto di partenza per costruire una realtà con la quale il sistema possa confrontarsi.

I dati non sono unità ma differenze. 20° non è né 19°, né 21°. Se queste differenze fanno la differenza per l'osservatore, cioè se aggiungono qualcosa che l'osservatore prima non sapeva oppure non si aspettava, allora esse hanno valore di informazione, secondo la classica definizione di Bateson [6]. In questo senso le macchine che elaborano dati sono macchine che producono informazioni. Allo stesso tempo i dati sono un modo molto particolare per farsi un'idea della realtà con la quale abbiamo a che fare. Quello che oggi diamo per scontato - che la realtà sia ridotta a dati e che i dati siano comunicati per offrire una rappresentazione oggettiva della realtà - dovrebbe essere considerato piuttosto, per la sua elevata improbabilità, come un risultato dell'evoluzione socio-culturale. Soltanto a partire dal XVII-XVIII secolo ci siamo abituati a considerare la realtà in termini di dati, vale a dire in forma statistica. E solo quando questa particolare forma di rappresentazione della realtà sociale è stata accettata come qualcosa di plausibile, l'assicurazione moderna è decollata.³

La statistica parte dal presupposto che il caso singolo sia imprevedibile per definizione. Come per i contingenti futuri di Aristotele, l'unica cosa certa che possiamo affermare sul singolo caso è che non c'è alcuna certezza su come andrà a finire. Ma se molti casi singoli che condividono certe proprietà vengono aggregati, si scopre improvvisamente che a livello di aggregato emergono delle regolarità sulle quali si può contare.

³Sulla lenta affermazione della statistica e le difficoltà incontrate per essere accettata come qualcosa di normale si veda Porter [7] e Desrosières [8].

Si scopre in altri termini che c'è un ordine nella contingenza. Questa scoperta non era passata inosservata in ambito teologico. Da un lato essa poteva essere assunta come la prova che, nonostante la sua estrema turbolenza, nel mondo ci sia una Provvidenza; dall'altro lato dimostrava che quello che dal punto di vista dell'uomo è aleatorio, dal punto di vista di Dio è necessario [9].

Le compagnie di assicurazione avevano visto in questa possibilità di calcolare la necessità del caso un'opportunità fondamentale per fare i conti con un futuro imprevedibile. Paradossalmente però, quello che costituisce ancora oggi un requisito imprescindibile dell'industria assicurativa, aveva rappresentato nel XVII secolo un ostacolo alla sua istituzionalizzazione per via del fatto che, come tutti i contratti aleatori, anche l'assicurazione doveva garantire una parità di condizioni a tutti coloro che decidevano di partecipare alla scommessa [10, 11]. Nessuno giocherebbe d'azzardo se sapesse che i dadi sono truccati. Fra assicurato e assicuratore si pretendeva quindi un'equa esposizione all'alea che costituiva l'oggetto del contratto. Solo così il contratto poteva essere accettato come lecito. L'ironia stava nel fatto che proprio questa classificazione giuridica dell'assicurazione fra i contratti aleatori, come le scommesse e il gioco d'azzardo, ostacolava (non solo per ragioni morali) la sua accettazione e diffusione nella società. A questo si aggiungeva il fraintendimento secondo cui la statistica avrebbe offerto all'assicuratore un vantaggio sull'assicurato. Per la legge dei grandi numeri, l'assicuratore sapeva che nel lungo periodo e scommettendo su molti casi avrebbe avuto qualche probabilità di guadagnarci⁴, ma nel confronto diretto con l'assicurato l'alea restava incalcolabile.

Nella statistica, in ogni modo, quello che veniva obliterato era proprio l'individuo. Mentre l'assicuratore si confronta sempre con una probabilità, l'assicurato si confronta piuttosto con la possibilità [13]. Nel primo caso l'assicuratore cerca di prevedere quanto potrebbe costargli alla fine il rischio di garantire al gruppo di assicurati una copertura finanziaria per un evento parti-

⁴Questo lo si sapeva anche prima dell'avvento del calcolo della probabilità statistica. Si veda Cotrugli [12], p.75: all'assicuratore si consiglia di "assicurare al continuo, & sopra ogni nave, perché l'una ristora l'altra, & di molti [casi l'assicuratore] non può che guadagnare".

colare. Nel secondo caso l'assicurato vorrebbe sapere se la sfortuna toccherà proprio a lui, oppure se verrà risparmiato. Di fronte alla probabilità calcolata su base statistica, la domanda dell'assicurato è sempre: "è questo il mio caso?". Ma di fronte a questa domanda l'assicuratore può solo sollevare le braccia e dire: "Non lo so!".

Proprio qui gli algoritmi predittivi fanno in qualche modo la differenza. Essi promettono infatti una previsione personalizzata, tagliata su misura sul singolo individuo, a partire dallo scavo dei dati prodotti dai comportamenti di centinaia di migliaia di altri individui che in comune non hanno alcunché, a parte il fatto di produrre con i propri comportamenti, appunto, dei dati. Con questi algoritmi si affermerebbe poco alla volta quella che François Ewald ha definito una logica della singolarità [14]: anziché inserire l'individuo in un gruppo e annullarlo nella forma dell'uomo medio, l'elaborazione dei dati produce profili individuali che servono poi a personalizzare il pronostico, offrendo così a ciascun individuo una rappresentazione unica del suo destino.

Per raggiungere questo scopo gli algoritmi vanno a cercare nei dati delle correlazioni. Anziché limitarsi a indagare le variabili suggerite dalle ipotesi di partenza della ricerca, gli algoritmi predittivi ammettono qualsiasi tipo di dato che possa essere trattato dalla macchina e attraverso delle raffinatissime tecniche di tipo statistico cercano di isolare dei valori che siano fortemente correlati a uno stato futuro - ovvero, dei predittori. Accettando qualsiasi tipo di dato, questi algoritmi finiscono spesso per trovare correlazioni importanti a partire da quelli che vengono definiti dati vicarianti. In altri termini: le correlazioni più forti, spesso, sono paradossalmente quelle deboli.

Se per esempio si volesse acquistare un'auto sul mercato dell'usato sicuro, si vorrebbe sapere con un certo margine di incertezza residuale a che cosa si debba prestare maggiormente attenzione per evitare di comprare un'auto che dopo pochi chilometri (cioè nel futuro) ci potrebbe lasciare a piedi. La ricerca ha mostrato che in modo del tutto inaspettato, quindi altamente informativo, un ottimo predittore di un usato sicuro è il colore arancione e non, per esempio, la condizione del motore o l'usura dei freni [15]. Perché?

Questo gli algoritmi predittivi non lo dicono. Essi esplorano i dati alla ricerca di correlazioni, ma non spiegano perché certe correlazioni funzionano meglio di altre, e nemmeno perché in generale esse funzionino. Gli algoritmi predittivi dicono in breve se certi valori possono predire il futuro con una certa probabilità, ma non perché.

Ovviamente anche la statistica ha sempre usato dati vicarianti: il sesso del conducente, per esempio, è un ottimo predittore della sua prudenza alla guida. E il ritardo nel pagamento delle bollette è un ottimo predittore della sua imprudenza. Ma quello che si profila con l'uso di algoritmi che vengono allenati per scavare nei dati alla ricerca di predittori è che la scoperta e l'impiego di strane (deboli) correlazioni potrebbe diventare incontrollata, con degli effetti ancora poco chiari e difficilmente giustificabili, sia sul piano etico sia sul piano giuridico, soprattutto qualora queste correlazioni venissero impiegate non solo per prevedere il futuro, ma anche per prendere delle decisioni.

Il futuro dell'assicurazione

Questi risultati costituiscono in realtà soltanto l'inizio di una ricerca che è ancora in buona parte da realizzare. Da un punto di vista sociologico l'ipotesi è che se cambia il modo di predire il futuro, cambia per forza anche il modo di assicurare. Con quali conseguenze sulla società complessiva, tuttavia, è ancora difficile da dire. Qui mi limito a proporre alcune considerazioni particolari che dovrebbero più che altro dare un'idea della complessità del problema in questione.

Per molti osservatori, quello che si delinea all'orizzonte con l'uso dei *Big Data* a scopi predittivi è un nuovo mondo che nel caso dell'assicurazione sarebbe caratterizzato soprattutto dall'opportunità inedita di personalizzare il premio della polizza assicurativa per garantire, così, a ciascun assicurato il giusto prezzo [14]. Già solo questo solleva molti dubbi su quello che dovrebbe essere considerato giusto in un meccanismo come quello assicurativo che è ancora essenzialmente basato sul principio di mutualità, per cui i casi più fortunati compensano sul piano finanziario quelli meno fortunati, con il vantaggio, dal lato degli assicurati, che i danni possono essere risarciti, e dal lato dell'assicurato-

re, che alla fine di un certo periodo la compagnia ci può guadagnare qualcosa. Questo meccanismo funziona finché la segmentazione della popolazione di riferimento garantisce allo stesso tempo la competitività dei premi di polizza e la compensazione finanziaria. Ma apre dei dilemmi difficilmente risolvibili quando il segmento si avvicina alla soglia di 1 (cioè il singolo individuo) e rinuncia al principio di mutualità.

I premi possono per esempio essere adeguati alla differenza statistica fra maschi e femmine, per cui le femmine pagano meno dei maschi per una copertura Rc auto perché in media guidano più prudentemente. Questo garantisce alla compagnia, come abbiamo visto, un certo controllo sull'alea dei singoli casi e può considerarsi un modo equo di stabilire il prezzo dell'assicurazione. Ma se si afferma una logica della singolarità, il maschio che guida molto prudentemente potrebbe sentirsi giustamente discriminato da una statistica che generalizza per classi e stigmatizza i maschi, obliterando le peculiarità di ciascun individuo. L'obiezione sarebbe in altri termini: "è vero che sono maschio, e che statisticamente dovrei guidare meno prudentemente delle femmine, ma in quanto individuo non mi riconosco nella classe in cui la statistica mi inserisce forzatamente in quanto, pur essendo maschio, sono un individuo assai prudente". In questo senso la statistica è discriminante per definizione, di conseguenza le differenze che essa impone non sono eque.

Le tecnologie digitali offrono appunto questo: la possibilità di profilare ciascun individuo in modo da assegnare, per così dire, "a ciascuno il suo" - il suo premio, la sua affidabilità creditizia, la sua pubblicità. La telematica, per esempio, consente di monitorare con estrema precisione il comportamento dell'individuo alla guida (in ambito assicurativo si parla a questo proposito di tariffe comportamentali). In questo modo dovrebbe essere possibile fissare un prezzo di polizza che sia proporzionale all'esposizione individuale al pericolo. A ciascun conducente si potrebbe quindi offrire un premio tagliato su misura sul comportamento effettivo e non sulle differenze più o meno arbitrariamente fissate dalla statistica tradizionale (età, sesso, luogo di residenza e così via).

Tutto questo suona di primo acchito molto

equo. Dopo tutto il ragionamento è: "se guido poco e prudentemente, perché dovrei pagare anche per quelli che guidano molto e imprudentemente?" Ma questa equità nasconde parecchie insidie (più estesamente su questo punto si veda Ref.[16]).

Da un lato, come detto, non è chiaro come l'industria assicurativa potrebbe accettare questo metodo di tariffazione senza mettere a repentaglio il meccanismo della mutualità. Qui tra l'altro si vede come l'incertezza, che è uno dei grandi problemi esistenziali dell'uomo, costituisca sul piano sociale anche una risorsa senza la quale, appunto, una certa solidarietà basata sull'opacità del destino individuale non sarebbe possibile. Dall'altro lato si contraddice in qualche modo quella che da sempre è stata la funzione sociale dell'assicurazione, ovvero incoraggiare gli individui a esporsi a dei pericoli per approfittare delle opportunità ad essi connesse. Portata all'estremo, la logica della profilazione individuale indurrebbe a praticare una forma di auto-inibizione che minerebbe i presupposti su cui si basa buona parte della società moderna, soprattutto quando essa viene intesa come una società orientata al rischio.

In conclusione vorrei tornare brevemente sul tema della previsione. Nella misura in cui essa viene personalizzata, il pronostico finisce per assomigliare più che altro a una profezia. Il destino di Edipo non viene anticipato perché Edipo aveva delle proprietà statisticamente rilevanti come "essere maschio" e "figlio di una famiglia benestante", bensì solo in quanto Edipo. E la profezia funziona, cioè si auto-adempie, soltanto in quanto viene comunicata. Questo solleva alcuni interrogativi anche in merito all'uso di algoritmi predittivi che lavorano con dati digitali.

Una prima questione, anche qui tipicamente sociologica, è che il problema delle profezie (incluse le predizioni algoritmiche) non è tanto se siano vere o false. La vera domanda è che cosa succede se la gente ci crede? Edipo avrebbe avuto un destino diverso se i suoi genitori non avessero creduto all'oracolo che aveva profetizzato la sciagura.

Una seconda questione è che la profezia funziona appunto solo se viene comunicata. I suoi effetti dipendono dal fatto che quando viene resa nota, la profezia entra automaticamente nel cir-

colo ricorsivo della comunicazione, innescando delle conseguenze che spesso sono non solo non-intenzionali, ma anche imprevedibili. Viene da chiedersi dunque se sia opportuno comunicare le predizioni ricavate dalla elaborazione dei dati digitali, ma anche se sia lecito tacerle, soprattutto quando sulla base di queste predizioni vengono prese delle decisioni drasticamente discriminanti sul piano sociale, come la libertà vigilata per un carcerato, l'assunzione per un neo-laureato o il credito per un imprenditore.

Il problema poi non è solo se sia opportuno comunicare le predizioni, ma anche se sia opportuno comunicare i predittori. Chi sapesse in anticipo che il colore arancione è un ottimo predittore di un usato sicuro sul mercato delle automobili, potrebbe far dipingere la propria auto di arancione prima di venderla. In questo modo la trasparenza nell'impiego di algoritmi predittivi darebbe agli individui la possibilità di fregare il sistema⁵. Ma qui di nuovo si apre il dilemma etico e giuridico di come non comunicare predittori che possono essere determinanti nel prendere delle decisioni che non solo dipendono dal futuro, ma condizionano anche il futuro degli individui coinvolti. Di questo si dovrà occupare la società perché anche se il destino è sempre individuale, le conseguenze sono sempre sociali.

Questo articolo costituisce la versione ampliata della conferenza tenuta al convegno "Conoscenza e libertà nell'epoca dei Big Data (Seconda parte: Big Data: libertà, profezia o predestinazione?)" organizzato dall'Istituto Filosofico di Studi Tomistici e svoltosi presso la Camera di Commercio di Modena il 10 ottobre 2020. Colgo l'occasione per ringraziare il Prof. Claudio Testi per l'invito e i partecipanti al convegno per le domande con cui hanno animato il dibattito.



[1] V. Mayer-Schönberger, K. Cukier: *Big Data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Milano, 2013 (Garzanti).

[2] D. Cardon: *Che cosa sognano gli algoritmi. Le nostre vite al tempo dei big data*, Mondadori, Milano (2016).

⁵Sul vasto dibattito etico e giuridico che ruota attorno a questo problema (ma senza una chiara comprensione del punto cruciale) si veda [17].

- [3] J. MacCormick: *9 algoritmi che hanno cambiato il futuro*, Apogeo, Milano (2012).
- [4] F. Provost, T. Fawcett: *Data Science for Business. What You Need to Know About Data Mining and Data-Analytics Thinking*, O'Reilly, Sebastopol (CA-USA) (2013).
- [5] R. Swedloff: *Risk Classification's Big Data (R)evolution*, Connecticut Insurance Law Journal, 21 (2014) 339.
- [6] G. Bateson: *Forma, sostanza e differenza*. In: G. Bateson: *Verso un'ecologia della mente*, Adelphi, Milano (2000).
- [7] T. Porter: *The Rise of Statistical Thinking 1820-1900*, Princeton University Press, Princeton (1986).
- [8] A. Desrosières: *La politique des grands nombres. Histoire de la raison statistique*, Éditions La Découverte, Paris (1993).
- [9] J. P. Süßmilch: *Die Göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen*, J. C. Spener, Berlino (1741).
- [10] L. Daston: *Rational Individual Versus Laws of Society: From Probability to Statistics*. In: M. Heidelberger et al. (a cura di): *Probability Since 1800. Interdisciplinary Studies of Scientific Development* BVA Bielefelder Verlag GmbH, Bielefeld (1983).
- [11] L. Daston: *The Domestication of Risk: Mathematical Probability and Insurance, 1650-1830*. In: Lorenz Krüger et al. (a cura di): *The Probabilistic Revolution I*, The MIT Press, Cambridge (MA) (1987).
- [12] B. Cotrugli: *Della mercatura et del mercante perfetto [ediz. orig. 1458]*, Alla Libreria del Bozzola, Brescia (1602)..
- [13] N. Luhmann: *Das Risiko der Versicherung gegen Gefahren*, Soziale Welt, 43 (1996) 273.
- [14] F. Ewald: *Assurance, prévention, prédiction ... dans l'univers du Big Data*, Institut Montparnasse, Paris (2012).
- [15] Q. Hardy: *Bizarre Insights From Big Data*, The New York Times, 28 marzo. <https://bits.blogs.nytimes.com/2012/03/28/bizarre-insights-from-big-data/> [ultimo accesso 21 ottobre 2020].
- [16] A. Cevolini, E. Esposito: *From Pool to Profile: Social Consequences of Algorithmic Prediction in Insurance*, Big Data & Society, 7 (2020) 1.
- [17] F. Pasquale: *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, MA (2015)..



Alberto Cevolini: è professore associato di Sociologia presso l'Università di Modena e Reggio Emilia. Attualmente è impegnato nel progetto ERC PREDICT (ADG 2018-Nr. 833749, PI Elena Esposito) per la parte dedicata all'assicurazione.

Linguaggio nei numeri e numeri nel linguaggio.

Linguistica, Matematica e Cryptonalisi

Thomas Christiansen

Dipartimento Studi Umanistici, Università del Salento, Lecce, Italia

1 Introduzione: la scienza della linguistica

Ai tradizionalisti, gli studi di matematica e di linguistica, potrebbero sembrare avere poco in comune, poiché costituiscono due modi di apprendimento molto differenti. Nelle università medioevali le sette arti cosiddette liberali, erano separate in una divisione inferiore o *trivium* (grammatica, retorica, e logica) e una superiore, *quadrivium* (aritmetica, musica, geometria e astronomia). Oggi la matematica è trattata come una scienza. Lo studio del linguaggio è ancora, ma sempre meno, unito ad argomenti come letteratura e traduzione e classificato come una delle arti¹. Ulteriore confusione emerge dal fatto che il termine linguista viene utilizzato in modi piuttosto differenti: da una parte descrive chi sa parlare bene lingue diverse (cioè un poliglotta o

¹In maniera analoga, la logica è uscita dagli stretti confini del *trivium* in quattro campi differenti: filosofico, formale, informale e matematico.

1 Introduction: the Science of Linguistics

To the more traditionally minded, the studies of mathematics and linguistics may seem to have little in common, constituting two very different types of learning. In medieval universities, the seven so-called liberal arts were divided into a lower division or *trivium* (grammar, rhetoric, and logic) and a higher *quadrivium* (arithmetic, music, geometry, and astronomy). Today, maths is treated as a science. Still, but increasingly less so, language is sometimes lumped together with things like literature and translation and classed as one of the arts¹. Added confusion comes from the fact that the term linguist has quite different uses: on the one hand, it can describe someone who is good at speaking different languages (i.e. a polyglot or plurilingual person); on the other,

¹Similarly, logic has also moved out of the strict confines of the *trivium* into four separate fields: philosophical, formal, informal, and mathematical.

una persona plurilingua): dall'altra parte denota lo studioso che analizza i linguaggi e le loro strutture, spesso ad un alto livello di astrazione. Quest'ultimo è uno scienziato, specialmente nella sua incarnazione moderna, poiché osserva, misura, costruisce e verifica ipotesi.

L'incremento di un approccio scientifico alla linguistica è emerso gradualmente a partire dalla metà del ventesimo secolo fino a diventare la corrente principale. Questo è avvenuto rendendosi conto che il linguaggio è un fenomeno naturale, un prodotto della biologia umana (come l'abilità di ridere o di camminare su due gambe), non è un artefatto culturale creato dall'uomo (come la birra, il pane o la poesia). Per questo il linguaggio ha acquisito un significato speciale nei contesti della biologia, delle scienze cognitive e della psicologia poiché oggi è riconosciuto come qualcosa che si è evoluto, come il resto dei nostri corpi e menti, in centinaia, migliaia e possibilmente milioni di anni. Conseguentemente, è qualcosa di molto più complesso e complicato di ogni cosa che possa essere prodotta dall'uomo.

L'argomento che l'abilità di usare il linguaggio (sebbene non si parli di una lingua specifica) sia istintivo è comunemente associata a Chomsky [1], sebbene ci siano indicazioni di questa visione in Darwin (Ref. [2], Chapter III):

"Come osserva Horne Tooke, uno dei fondatori della nobile scienza della filologia il linguaggio è un'arte come fare la birra o il panificare; ma scrivere sarebbe simile. Non è un vero istinto, poiché ogni lingua deve essere appresa. È tuttavia molto diversa da tutte le arti comuni dato che l'uomo ha una tendenza istintiva a parlare, come vediamo nel balbettare degli infanti; mentre nessun bimbo ha la tendenza a produrre birra, panificare o scrivere." [Sottolineatura nostra].

Prima di rendersi conto che il linguaggio fosse istintivo, la cosiddetta teoria comportamentista della psicologia considerava che l'acquisizione potesse essere basata puramente su stimoli e risposte. Chomsky, e molti altri dopo lui, tra i quali lo psicologo cognitivo Steven Pinker [3], controbatterono che questi meccanismi non riuscivano

it denotes the type of scholar who analyses languages and their structures, often at highly abstract levels. The latter is a scientist, especially in their modern incarnation, because they observe, measure, make and test hypotheses.

The increasingly scientific approach to linguistics gradually came to dominate the mainstream from the mid-20th century onwards. It arose from the realisation that language is a natural phenomenon, a product of human biology (like the ability to laugh or to walk on two legs), not a cultural artefact created by humans (like beer, bread, or poetry). As such, language is something that has acquired special significance in the contexts of biology, cognitive science and psychology because it is now recognised as something that has evolved, like the rest of our bodies and minds, over hundreds of thousands, possibly millions, of years. Consequently, it is something far more complex and complicated than anything human-made could ever be.

The argument that the ability to use language (though not to speak a specific language) is instinctive is most readily associated with Chomsky [1] although there are hints of this view in Darwin (Ref. [2], Chapter III):

"As Horne Tooke, one of the founders of the noble science of philology, observes, language is an art, like brewing or baking; but writing would have been a better simile. It certainly is not a true instinct, for every language has to be learnt. It differs, however, widely from all ordinary arts, for *man has an instinctive tendency to speak*, as we see in the babble of our young children; whilst no child has an instinctive tendency to brew, bake, or write." [Emphasis ours]

Prior to the realisation that language was instinctive, the so-called behaviourist theory of psychology had held that acquisition could be based purely on stimulus and response. Chomsky, and many others since, including the cognitive psychologist, Steven Pinker [3], counter that such mechanisms cannot account for the

a tener conto della velocità con la quale ogni bambino acquisisce il linguaggio o la complessità delle strutture sintattiche che produce, anche ad una giovane età, alcune di queste sono generate autonomamente da regole astratte e non sono prodotte dall'imitazione di quello che sente attorno a sé. Secondo Chomsky [4], i bambini nascono con una grammatica universale che si può descrivere come intrinsecamente connessa nel cervello. Questo è evoluto come meccanismo con cui l'infante può analizzare un numero limitato di *input* di una lingua specifica e convertirlo in un modello mentale funzionante del linguaggio stesso. Questo modello può essere usato per generare strutture permettendo, in effetti, al bambino di cominciare a parlare la lingua in questione². I linguaggi non sono quindi scaricati o appresi. Invece, sono acquisiti per contatto con *input* adeguati, e ricostruiti attraverso la mappatura degli schemi e delle forme incontrati nella cornice fornita dalle grammatica universale che è innata.

Alcuni dei primi studi di linguistica nell'era moderna erano focalizzati sull'origine dei linguaggi e la connessione tra loro: per esempio, nei secoli diciottesimo e diciannovesimo, l'osservazione di William Jones che le lingue classiche dell'antica Grecia e Roma sembravano condividere delle strutture simili con lingue come il Farsi ed il Sanscrito, e la conclusione che avevano radici comuni; oppure la legge di Jakob Grimm riguardante i cambiamenti sistematici del suono nelle consonanti occlusive / plosive, a cui furono sottoposte quando evolsero nel tardo Proto-Germanico. Nel ventesimo secolo, i linguisti (ad esempio Ferdinand de Saussure, Otto Jespersen, Leonard Bloomfield, Noam Chomsky, Michael Halliday) rivolsero la loro attenzione alla struttura del linguaggio (in astratto) studiando cose come il meccanismo che permette ai concetti di essere condensati in un messaggio e quindi codificati in suoni, parole o strutture sintattiche. In questo modo, l'analisi di tutti i livelli del linguaggio (ad esempio fonologia, morfologia, lessico, grammatica e sintassi) fu vista come qualcosa

²In Inglese, con una certa confusione, lo stesso termine *language* è usato sia per le regole e i principi che governano il linguaggio sia per indicare la produzione degli elementi e delle strutture linguistiche: quello che de Saussure [5], in francese, chiamava *langue* e *parole* e che Chomsky [6] ha definito come *competence* e *performance*.

speed at which any child acquires language or the complexity of the syntactic structures that they produce even at a young age, some of which they can be seen to generate for themselves out of abstract rules and not merely through imitation of what they hear around them. According to Chomsky [4], children are born with a universal grammar which one could describe as "hard-wired" into the brain. This has evolved to serve as a mechanism whereby infants can analyse a limited amount of input from a specific language and convert it into a functional mental model of that same language. This model can then be used to generate structures, in effect allowing the child to start "speaking" the language in question². Languages therefore are not so much "handed down" or learnt. Rather, they are acquired through contact with adequate input, and reconstructed through the mapping of the patterns and forms encountered onto the framework provided by the innate universal grammar.

Some of the earliest treatments of linguistics in the modern era were focused on the origin of languages and the connections between them: for example, in the eighteenth and nineteenth centuries, William Jones's observation that the classical Languages of ancient Greece and Rome seemed to share common features with languages such as Farsi and Sanskrit, and his conclusion that they must share common roots; or Jakob Grimm's laws regarding the systematic sound changes which the Proto-Indo-European stop consonants underwent as they developed in later Proto-Germanic. In the twentieth century, linguists (e.g. Ferdinand de Saussure, Otto Jespersen, Leonard Bloomfield, Noam Chomsky, Michael Halliday) turned their attention to the structure of language (in the abstract), looking at such things as the mechanisms that allow concepts to be put together into messages and then encoded as sounds, words or syntactic structures. In this way, analysis of all levels of language (e.g. phonology, morphology, lexis, grammar, and syntax) came to be seen as something descriptive.

²In English, confusingly, the same term, *language*, is used for both the rules and principles governing language, and for the actual production of linguistic elements and structures: what de Saussure [5], in French, called *langue* and *parole* or what Chomsky [6] named *competence* and *performance*.

di descrittivo. Tradizionalmente, le regole del linguaggio sono state considerate come prescritte: non emergenti naturalmente attraverso l'evoluzione del sistema, ma determinate da una convenzione tra gli utilizzatori delle lingue e le generazioni che li hanno preceduti. Per questo motivo i grammatici tradizionalisti e quelli che hanno ancora propensioni pedagogiche, possono scartare con leggerezza quelle cose che non riescono a spiegare come fossero eccezioni, senza rendersi conto che stanno falsificando le loro stesse ipotesi.

Oggi, la linguistica è vista come uno studio scientifico di un fenomeno naturale: essenzialmente un'indagine del comportamento, allo stesso modo in cui uno zoologo può osservare un animale oggettivamente e spassionatamente senza ricorrere ad alcuna idea preconcepita di giusto o sbagliato, cioè, come la creatura in questione dovrebbe, o non dovrebbe, comportarsi. Secondo il pluritalentoso Chomsky³, che nella prima parte della sua carriera fu all'avanguardia collaborando in vari articoli di linguistica matematica, l'importanza del linguaggio risiede nel fatto che costituisce una "finestra della mente"⁴. Il linguaggio è così centrale per la condizione umana e per la società che non dovrebbe essere una sorpresa che oggi esista una miriade di filoni della linguistica che vanno da modelli astratti fino a studi del linguaggio applicato ad aree specifiche⁵. Ci sono anche numerose e molto diverse maniere di descrivere il linguaggio e le sue strutture, ad esempio, semplicemente collegando le strutture sintattiche: grammatica dei casi, prospettiva funzionale delle frasi, grammatica funzionale al lessico, teoria ottimale, grammatica sistemica, e, in ultimo, ma non meno importante, la grammatica

Traditionally, the "rules" of language had been seen as prescribed: not naturally occurring through the evolution of a system, but determined by convention among the users of the languages and the generations that had preceded them - and this is why traditional grammarians, and still some of those today of a pedagogical bent, can glibly dismiss those things that they cannot explain as "exceptions" without realising that they were thereby falsifying their own hypotheses.

Today, linguistics is seen as a scientific study of a naturally occurring phenomenon: essentially an enquiry into behaviour, in the same way that a zoologist might observe an animal objectively and dispassionately without recourse to any preconceived idea of correct and incorrect, that is, how the creature in question should or should not behave. According to the multitalented Chomsky³, who in his early career was trailblazing in that he collaborated on various papers specifically on mathematical linguistics, the importance of language stems from the fact that it constitutes a "window on the mind"⁴. Language is so central to the human condition and society that it should come as no surprise that today there exist myriad strands of linguistics ranging from abstract models to the study of language applied to specific areas⁵. They are also numerous and very diverse ways of describing language and its structures, for example, just relating to syntactic structures: case grammar; functional sentence perspective; lexical functional grammar; optimality theory; systemic grammar, and, last but not least, Chomskyan transformational-generative grammar.

³Anche se è spesso identificato come "padre della linguistica moderna" è anche noto come filosofo, scienziato cognitivo, storico e attivista politico.

⁴Chomsky ritiene che la linguistica sia una branca della psicologia "Sono primariamente interessato alla possibilità di imparare qualche cosa, dallo studio del linguaggio, che porterà a chiarire alcune proprietà inerenti la mente umana." ([7], p. 90).

⁵Esempi del secondo questo ricadrebbero sotto la generica definizione di linguistica (ad esempio, lo studio dell'insegnamento del linguaggio); esempi del primo includerebbero approcci molto diversi come: linguistica clinica, linguistica computazionale, linguistica dello sviluppo, ecolinguistica, linguistica evolutiva, linguistica forense, linguistica storica, neurolinguistica, sociolinguistica.

³As well as being often referred to as "the father of modern linguistics", he is also a noted philosopher, cognitive scientist, historian, and political activist.

⁴Chomsky argues that linguistics is a branch of psychology: "I am primarily intrigued by the possibility of learning something, from the study of language, that will bring to light inherent properties of the human mind" ([7], p. 90).

⁵Examples of the latter, would fall under the general banner of applied linguistics (e.g. the study of language teaching); examples of the former would include such diverse approaches as: clinical linguistics; computational linguistics; developmental linguistics; ecolinguistics; evolutionary linguistics; forensic linguistics; historical linguistics; neurolinguistics; sociolinguistics.

chomskiana generativo-trasformazionale.

In questo articolo, fornirò degli esempi di come alcuni aspetti della linguistica si basino su principi matematici. La matematica può essere vista come rilevante in ogni area e ad ogni livello della moderna analisi linguistica. Di conseguenza la sfida, in un breve articolo come questo risiede non tanto nel trovare esempi ma piuttosto nell'evitare di rimanere impantanati dall'evidenza. Per questa ragione strutturerò la presentazione attorno ad una specifica area che illustra, anche a coloro che potrebbero nutrire scarso interesse in problemi di linguistica astratta, l'applicazione pratica della matematica alla linguistica, specificamente la criptoanalisi o la decodifica dei codici⁶. Questa è un'area che mostra ampiamente l'alto livello che può raggiungere lo sforzo dell'uomo quando studiosi con diverse conoscenze, tra i quali matematici e linguisti, condividono le loro risorse molto differenti e lavorano insieme per uno scopo comune.

2 Criptoanalisi: decodifica dei codici

Una delle aree più note in cui i percorsi dei matematici e dei linguisti si sono incrociati è nel mastodontico sforzo fatto nel periodo della Seconda Guerra Mondiale dagli Alleati per decifrare le comunicazioni radio criptate delle potenze dell'Asse all'interno dell'area che i militari chiamano *Signal Intelligence* (SIGINT). Questo fu sviluppato non solo dalla britannica *Government Code and Cypher School* (Stazione X), ma anche: dal prebellico polacco *Biuro Szyfrów* (Ufficio codici) (fino al 1939); dal francese *Deuxième Bureau* (fino al 1940); o, negli USA, dal *Army Signal Intelligence Service* del OP-20-G della marina militare. In queste strutture, esperti in vari campi, anche lingue (specialmente quelle usate dal nemico) erano affiancati da matematici. Per ironia del destino questo era in aperta contraddizione con Godfrey Hardy, il pacifista e matematico, che nel 1940 dichiarò pubblicamente che "la vera matematica non ha effetti sulla guerra." [8].

Lavorando insieme come squadre interdisciplinari furono capaci di decifrare codici che erano

⁶Da non confondere con la criptografia che è la scienza di produrre codici.

In this article, we will provide some examples of how aspects of linguistics rely on mathematical principles. Mathematics can be seen as relevant to every area and every level of linguistic analysis in modern times. Consequently, the challenge in a brief article such as this lies not in finding examples but rather in avoiding being swamped by the evidence. For this reason, we will structure our discussion around one specific area which illustrates, even to those who may otherwise show little interest in abstract linguistic matters, the practical applications of mathematics to linguistics, namely cryptanalysis⁶ or codebreaking. This is an area which amply shows the heights of human endeavour that can be achieved when scholars of different backgrounds, among whom mathematicians and linguists, pool their very different resources and work towards a common goal.

2 Cryptanalysis: Codebreaking

One of the best known areas where the paths of mathematicians and linguists have crossed is in the mammoth efforts made in the general period of the Second World War by the Allies to decipher the encrypted radio communications of the Axis Powers within the area of what the military call *Signals Intelligence* (SIGINT). This was carried out not just in Britain's *Government Code and Cypher [sic] School* ("Station X") at Bletchley Park, where Alan Turing worked, but also: the pre-war Polish *Biuro Szyfrów* "Cipher Bureau" (until 1939); the French *Deuxième Bureau* (until 1940); or in the USA, the Navy "OP-20-G", or *Army Signal Intelligence Service*. In such establishments, experts in various fields, including languages (especially those used by the enemy) were paired with mathematicians. In an irony of destiny, this was in direct contradiction of Godfrey Hardy, the pacifist and mathematician, who in 1940 famously declared that "real mathematics has no effect on war." [8].

Working together, such interdisciplinary teams were able to crack codes that were thought by their creators and users to be all but indecipher-

⁶Not to be confused with cryptography - code making.

stati pensati dai loro creatori come indecifrabili. Ci riuscirono identificando delle strutture sempre più complesse nei sempre più sofisticati codici generati dalle macchine usate dalle potenze dell'Asse. Per fare questo, e alla velocità richiesta dalle operazioni di guerra, dovettero progettare e costruire (nessuno di questi era un compito facile) della macchine di decifrazione come le bombe (uno sviluppo delle polacche *bomba kryptologiczna*) a Bletchley Park e anche il primo computer elettronico programmabile: Colossus (si veda la Sez. 3.2).

È possibile decifrare i codici di linguaggio proprio perché i messaggi, il modo in cui il linguaggio si manifesta, contengono, inevitabilmente, delle caratteristiche che sono ripetute e che possono essere contate e osservate a diversi livelli di frequenza in relazione tra loro. Nelle prossime due sezioni introdurrò le regole fondamentali del linguaggio per cifrare (criptografia) e decifrare (criptoanalisi) come preliminari necessari per una discussione più dettagliata della criptoanalisi dei codici altamente sofisticati usati nella Seconda Guerra Mondiale (Sez. 3.0).

2.1 Criptografia

I primi codici conosciuti risalgono ai tempi antichi, ed erano estremamente elementari rispetto agli standard moderni. Per questo motivo lo sforzo di nascondere il messaggio era paragonabile a quello per criptarlo. Nella moderna era delle telecomunicazioni, nascondere un messaggio è sempre più difficile. Prima della Seconda Guerra Mondiale, gli inglesi avevano investito fortemente nella tecnologia per cogliere, ricevere e leggere segnali radio, ed erano capaci di sentire le comunicazioni dell'Asse molto più lontano di quanto i Tedeschi e gli Italiani sospettassero, anche nel fronte orientale, che i Tedeschi immaginavano fosse al sicuro al di là del loro raggio d'azione.

In realtà, sfruttando il fatto che i segnali radio viaggiano su distanze estremamente lunghe, durante la Guerra Fredda le cosiddette *numbers stations* erano allestite in modo da trasmettere sequenze di numeri apparentemente casuali letti ad alta voce (soprattutto da voci sintetiche). Si pensa che questo metodo sia stato utilizzato da varie agenzie di spionaggio [8] per comunicare con agenti sul campo (a cui veniva indicato un

abile. They did this by identifying increasingly complex patterns in the ever more sophisticated machine-generated codes used by the Axis powers. To do this, and at the speed that wartime operations required, they had to design and build (neither easy tasks) decryption machines, such as, at Bletchley Park, the *bombes* (a development of the Polish *bomba kryptologiczna*) as well as the world's first programmable electronic computer: Colossus (see § 3.2).

Decryption of coded language is possible precisely because messages, the way that language manifests itself, inevitably consist of repeated features that may be counted and observed in different levels of frequency in relation to each other. In the next two sections, we will introduce the basics of language encryption (cryptography), and decryption (cryptanalysis) as a necessary preliminary for a more detailed discussion of the cryptanalysis of the highly sophisticated ciphers employed in WW2 in § 3.0.

2.1 Cryptography

The earliest known codes go back to ancient times, and they were extremely simple by modern standard. For this reason, as much effort went into hiding the message, inventing secret compartments etc., as into the actual encryption. In the modern era with the use of telecommunications, hiding a message is increasingly difficult. Before WW2, the British had invested heavily in technology to pick up, receive and read radio signals, which meant that they were able to listen in to Axis communications from much further away than the Germans or Italians suspected, even from the Eastern Front, which the Germans had thought were safely out of their range.

Indeed, exploiting the fact that radio signals could travel extremely long distances, during the Cold War so-called *numbers stations*, were set up openly transmitting apparently random series of numbers read aloud (mostly by synthetic voices). This method is widely thought to have been used by some agencies [8] to communicate with agents in the field (who would be given specific times at which to listen for messages),

istante preciso in cui ascoltare i messaggi), e forse qualcuna di queste stazioni trasmette ancora. La grande maggioranza dei numeri trasmessi sono casuali, ma altri, presumibilmente, formano un messaggio cifrato. Gli operatori SIGINT, si presume, sono obbligati a cercare di decifrare tutto, senza sapere quale sia la parte interessante⁷.

Al livello più semplice, il criptaggio implica un metodo detto *shift* o *cifrario di Cesare* dal nome di uno dei primi che lo ha adottato: lo stesso Giulio Cesare. Il metodo implica di rimpiazzare una lettera del testo base (quello non cifrato) con un altro carattere o simbolo nel testo cifrato. Ad esempio, A diventa D, B diventa E, C diventa F, e così via. "We attack at dawn" quindi diventa "zh dwwdfn dw gdzq". Questo semplice metodo fu utilizzato per secoli aumentando la sofisticazione per renderlo più difficile da decifrare. Ad esempio, si può mescolare, o idealmente randomizzare, lo *shift* (A diventa C, B diventa Z, C diventa P.)

Al di là del sistema di base, il *cifrario di Vigenère* fu sviluppato nel XVI secolo. Questo usa una sostituzione polialfabetica e alfabeti con sostituzione multipla. La codifica del testo originale è fatta usando il quadrato, o tavola, di Vigenère dove diverse colonne indicano diversi *shift* corrispondenti a diverse lettere dell'alfabeto. Tipicamente una parola chiave (ad esempio *water*) indica in quale ordine i diversi *shift* devono essere applicati (ad esempio, nella prima colonna A=W, in quella successiva A=A, così via, e nella terza A=T ecc.). Spesso la parola chiave è ripetuta in tutte le colonne (W A T E R W A T E R W A T E R). In caso di parole chiave più corte, questo è poco sicuro poiché la ripetizione di ogni chiave è frequente, creando sequenze identificabili a criptoanalisti allenati. In alternativa, il resto dell'alfabeto può essere presentato dopo la parola chiave, ad esempio W A T E R B C D F G H I J K L M N O P Q S U V X Y Z. In quest'ultimo caso, la ripetizione è garantita solo dopo 26 lettere, ma questo, essendo la lunghezza esatta dell'alfabeto, può essere un intervallo piuttosto ovvio per il criptoanalista.

⁷Il ministro Cecoslovacco degli interni ha dichiarato che la Cecoslovacchia usava questo sistema per lo spionaggio. Bletchley Park e la base britannica a Cipro sono state entrambe citate come sorgenti di ben note *number station*.

and perhaps still is as some of these stations remain on air. Most of the numbers transmitted are purely random, but others presumably form encrypted messages. SIGINT operators, one presumes, are obliged to try to decipher everything without knowing which parts are of interest⁷.

At the basic level, encryption comprises the method known as the *shift* or *Caesar cipher*, after one of its early adopters: Julius Caesar himself. This involves replacing a letter in the plaintext (unencrypted message) with another character or symbol in the ciphertext. For example, A becomes D, B becomes E, C becomes F, and so on. "We attack at dawn" thus becomes "zh dwwdfn dw gdzq". This basic method was used for centuries with increasing sophistications to make it more difficult to crack. For example, one can use scrambled, or ideally randomised, shifts (A becomes C, B becomes Z, C becomes P).

Out of this basic system, the *Vigenère cipher* was developed in the 16th century. This uses polyalphabetic substitution and multiple substitution alphabets. The encryption of the original text is done using a *Vigenère square* or *table* - where different columns give different shifts corresponding to different letters of the alphabet. Typically, a key word (e.g. *water*) indicates in which order the various shifts are to be applied (e.g. in the first column A = W; in the next, A = A and so on; and in the third, A = T etc.). Often the keyword is repeated in all the columns (W A T E R W A T E R W A T E R). In the case of shorter keywords, this is less secure because the repetition of each key is frequent, creating patterns discernible to a trained cryptanalyst. Alternatively, the rest of the alphabet may be written out the after keyword, e.g. W A T E R B C D F G H I J K L M N O P Q S U V X Y Z. In the latter case, a repetition only after every 26 letters is guaranteed, but this, being the precise length of the alphabet, may be a rather obvious interval for a cryptanalyst.

⁷The Czech Ministry of the Interior has declared that Czechoslovakia used them for espionage. Bletchley Park and the British base of in Cyprus have both been cited as the source of one well-known *number station*.

Ci sono cifrari poligrafici che coinvolgono diverse combinazioni di lettere (ad esempio, "th", "er", "wh", "en" o "ing" in inglese) o anche suoni (ad esempio il suono /ð/ in *that, those, thing* o /θ/ in *north, thorough, three* che sono sostituiti da caratteri separati. Alcuni codici hanno mescolato diversi metodi per rendere più difficile la decifrazione, ad esempio i *nomenclator*⁸ usati nell'alto Medio Evo fino ai primi dell'800 che usavano *shift* e un sistema di simboli con combinazioni di lettere, suoni, parole comuni (ad esempio *he, who, is, have, that*) e nomi propri. Ad esempio il nostro messaggio cifrato "We attack at dawn", "zh dwwdfn dw gdzq" può essere reso più complesso come "£ dw*d €+ gr %", dove £ è il pronome *we*, * indica che la lettera precedente deve essere ripetuta, € è la comune combinazione di lettere *ck*, % sta per *wn* e + è la preposizione *at*.

La conoscenza della lingua codificata⁹ permette di indovinare i possibili significati di certi gruppi di caratteri nel testo cifrato. Si può essere guidati da cose semplici come la lunghezza della parola. Ad esempio nel nostro testo cifrato "zh dwwdfn dw gdzq" si può fare una ragionevole ipotesi sul possibile significato di "zh" e "dw" dato che c'è un numero limitato di parole di due lettere in inglese (ad es. *at, ax, in, so, to, we, of*) e alcune sono molto più frequenti di altre (*at* e *ax*). Ogni gruppo di tre lettere in un testo cifrato in inglese ha altissima probabilità di essere *the* oppure *and*, dato che sono molto più frequentemente usate di altre parole di tre lettere (*for, not, you* ecc.)¹⁰. La lunghezza delle diverse parole può essere mascherata raggruppandole in gruppi, ignorando le spaziature. In questo modo il nostro esempio potrebbe diventare "zhdww dfndw gdzqy" (la *y* è un carattere casuale aggiunto proprio per completare il gruppo di 5 caratteri)

⁸Così chiamati perché usati come pizzini dal *nomenclator* (l'ufficiale che declamava i nomi agli ospiti onorevoli che arrivavano per funzioni importanti).

⁹A volte, per aumentare la sicurezza, varie lingue poco comuni possono essere usate: ad esempio, le forze armate americane impiegavano *speaker* di lingue poco studiate, come Comanche, Navajo o Basco (Euskara), come telefonisti e operatori radio nella prima e seconda guerra mondiale. Più recentemente l'esercito Britannico ha usato il gallese per comunicazioni non vitali nelle operazioni NATO di pace nella ex-Yugoslavia.

¹⁰Si veda, ad esempio, Oxford English Corpus (www.sketchengine.eu/oxford-english-corpus/).

There are also polygraphic ciphers that involve also different combinations of letters (for example, "th", "er", "wh", "en" or "ing" in English) or even sounds (e.g. the /ð/ sound in *that, those, thing*, or the /θ/ in *north, thorough, three*) being substituted by separate characters. Some codes have also used mixtures of different methods to make deciphering more difficult, such as the *nomenclators*⁸ used in the later Middle Ages until the 1800s, which used shifts and a system of symbols for common combinations of letters, sounds, and common words (e.g. *the, who, is, have, that*) and proper names. For example, our cipher of "We attack at dawn", "zh dwwdfn dw gdzq" could be enhanced as "£ dw*d €+ gr %", where £ stands for the pronoun *we*, * indicates that the previous letter is doubled, € stands for the common letter combination of *ck*, % for *wn*, and + for the preposition *at*.

Knowledge of the features of the language encoded⁹ also allows one to guess at the possible meanings of certain groupings of characters in the ciphertext. One may be guided by such simple things as the length of the word. For example, in the ciphertext "zh dwwdfn dw gdzq" above, one may take an educated guess at the meaning of "zh" and "dw" merely because there are a limited number of two letter words in English (e.g. *at, ax, in, so, to, we, of*) and some are far more frequent than others (cfr. *at* and *ax*). Any three letter word in a ciphertext based on English has a very high probability of being either *the* or *and*, so frequently are they used in relation to the other three letter words (*for, not, you* etc.)¹⁰. The length of different words can be disguised by grouping the symbols in the ciphertext into uniform groupings, ignoring word-spaces. In this way, our example could become "zhdww dfndw gdzqy" (*y* being a random character added just to

⁸So called because they were often used as crib sheets by *nomenclator* (the official who called out the names of honoured guests arriving at important functions).

⁹Sometimes to increase security, various unfamiliar languages could be used: for example, the US military employed code talkers, speakers of lesser-studied languages, such as Comanche, Navajo or Basque (Euskara), as telephonists and radio operators in WW1 and WW2. More recently, the British army is said to have sometimes used Welsh for non-vital communications in its NATO peacekeeping operations in Ex-Yugoslavia.

¹⁰See, for example, Oxford English Corpus (www.sketchengine.eu/oxford-english-corpus/).

o "£ dw*d €+ gd %".

2.2 Criptoanalisi

Tutti i metodi di criptaggio citati nella sezione precedente sono stati resi notevolmente meno sicuri con l'applicazione di quello che oggi è denominata *frequency analysis*. Questo metodo risale almeno al IX secolo e ad Al-Kindi nel suo "Manoscritto per decifrare messaggi criptati" [9]. Nel tardo XV secolo, lo stesso approccio ha raggiunto l'Europa, con Cicco Simonetta che ha scritto un manuale che utilizza metodi simili.

La *Frequency analysis* è molto efficace con i classici codici descritti nella sez. 2.1 perché, nelle sequenze del linguaggio scritto, alcune lettere e loro combinazioni appaiono con varia frequenza. Ad esempio, la lettera e è più comune delle lettere x (che è il motivo perché a Scrabble ci sono più tessere di certe lettere comuni, e sono maggiori i valori delle lettere più rare). Inoltre, alcune combinazioni di lettere appaiono più frequentemente di altre (ad esempio i digrammi *an*, *er*, *on* e *th* sono alcune delle coppie di lettere più comuni in Inglese). Eccetto in alcuni nomi propri stranieri o parole derivate da altre lingue, alcune combinazioni non appaiono affatto, ad esempio in Inglese: *hj*, *bq*, *xz*. In modo simile alcune lettere sono spesso raddoppiate (ad esempio *ee*, *ff*, *ll*, *oo*, *ss*, *tt*) mentre altre non lo sono mai (ad esempio *aa*, *ii*, *hh*, *jj*, *qq*, *vv*, *xx*). Inoltre, alcune lettere si trovano in precise posizioni, ad esempio, in Inglese, è estremamente raro trovare più di tre consonanti in sequenza; la lettera y è più frequentemente trovata alla fine di una parola che all'inizio.; solo una manciata di parole termina con i¹¹.

Il fatto che parole e lettere siano combinate in maniera strutturata e prevedibile è stato sfruttato dal famoso scrittore per l'infanzia Charles Lutwidge Dodgson (noto come Lewis Carroll), autore dei libri di "Alice nel paese delle meraviglie". Non è una coincidenza che Dodgson fosse assunto come insegnante di matematica al Christchurch Colleague di Oxford (specializzato nei settori della geometria, logica matematica, algebra li-

¹¹La stessa analisi può essere applicata ai suoni: il suono distintivo *ng* / *ŋ* / come in *thing*, non appare mai alla fine di una parola.

complete the last 5-character group), or "£ dw*d €+ gd %".

2.2 Cryptanalysis

All the encryption methods cited in the previous section were rendered notably less secure with the application of what is now called *frequency analysis*. This dates back to at least the 9th century and Al-Kindi in his "A Manuscript on Deciphering Cryptographic Messages" [9]. In the later 15th century, the same approach had reached Europe, with Cicco Simonetta also writing a manual employing similar methods.

Frequency analysis proves effective with the classical ciphers described in §2.1 because, in a stretch of written language, certain letters and combinations of letters occur with varying frequencies. For example, the letter e is more common than the letter x (which is precisely why in a Scrabble set there are more of tiles of certain common letters, and the values of rarer letters are higher). Furthermore, certain combinations of letters occur more frequently than others (e.g. the bigrams or digraphs of *an*, *er*, *on* and *th* are some of the most common pairs of letters in English). Except perhaps in proper names or words derived from other languages, some combinations do not occur at all, e.g. in English: *hj*, *bq*, *xz*. Similarly, some letters are often doubled (e.g. *ee*, *ff*, *ll*, *oo*, *ss*, *tt*) while others never are (e.g. *aa*, *ii*, *hh*, *jj*, *qq*, *vv*, *xx*). Furthermore, certain letters are more likely in certain positions, for example, in English, it is extremely rare to find more than three consonants in sequence; the letter y is more frequently found at the end of a word than at the beginning; and only a handful of words end in i¹¹.

The fact that words and letters are patterned in a structured, predictable manner was exploited by the well-known children's writer Charles Lutwidge Dodgson (a.k.a. Lewis Carroll), author of the "Alice in Wonderland" books. Not coincidentally, Dodgson was employed as a lecturer in mathematics at Christchurch Colleague, Oxford (specialising in the fields of geometry, mathematical logic, linear and matrix algebra, and recreational mathematics). One of Dodgson's / Car-

¹¹The same is true of sounds: the distinctive *ng* sound / *ŋ* / as in *thing*, never occurs at the beginning of a word.

neare e delle matrici, e matematica ricreativa). Una delle creazioni più famose di Dodgson / Carroll è il poema "Jabberwocky" una parodia della letteratura in Inglese Antico, che appare in "Attraverso lo Specchio" (1871), e i cui primi versi sono riprodotti qui sotto¹²

"Era brillosto, e gli alacridi tossi
succhiellavano scabbi nel pantùle:
Méstili eran tutti i paparossi,
e strombavan musando i tartarocchi."

Questa è normalmente classificata come letteratura nonsense, ma si può vedere che fornisce qualche senso sebbene molte delle parole non sono affatto convenzionali in Italiano: ad esempio brillosto, alacridi, tossi, succhiellavano, scabbi, pantùle, méstili, paparossi, strombavan, musando, tartarocchi. Questo perché la grammatica e la sintassi sono ortodosse, in maniera tale che si può identificare senso (o funzione) grammaticale allo strano vocabolario: ad esempio brillosto, alacridi, scabbi, méstili sembra funzionino come aggettivi; succhiellavano, strombavan, musando sono verbi, e tossi, pantùle, paparossi e tartarocchi nomi. Il fenomeno psicologico detto *effort after meaning* [10] gioca un ruolo importante. Gli esseri umani hanno la tendenza nel ricercare senso in qualsiasi cosa porti un messaggio. Questa necessità fondamentale è visibile anche nel ben attestato fenomeno detto *pareidolia*, che si manifesta molto spesso in cose come vedere delle facce e oggetti in nuvole, o sentire delle parole in suoni casuali come nei fenomeni di voci elettroniche volgarizzato dagli investigatori del paranormale.

Un maggiore fattore di rassomiglianza del senso è che, sebbene parole come brillosto non esistano nel lessico italiano, sembrano parole italiane. Questo perché sono costruite con combinazioni di lettere come br, ll, sl, osto ecc. (si confronti ad esempio con parole come *enough* *which* o *thought* che non usano sequenze presenti in italiano). Infine, e molto ingegnoso da parte di Carroll, è la fonostesia o il simbolismo fonetico che suggerisce quale possa essere il significato di qualche parola, e quindi implica la codifica ad

¹²Qui adotto la traduzione italiana intitolata "Il Ciarlestrone" di Adriana Crespi (1974): <http://www76.pair.com/keithlim/jabberwocky/translations/italian1.html> (N.d.T.)

roll's most famous creations was the poem "Jabberwocky", a parody of Old English literature, which appeared in "Through the Looking Glass" (1871), and whose first verse is reproduced below:

"Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe."

This is usually classed as nonsense literature, but it can be seen that it does convey some sense even though many of the words are not conventionally English at all: e.g. brillig, slithy, toves, gyre, gimble, wabe, mimsy, borogoves, mome, raths, outgrabe. This is because the grammar and syntax of the text is orthodox, so one can at least ascribe some grammatical sense (or function) to the strange vocabulary: for example brillig, slithy, mimsy and mome all appear to function as adjectives; gyre, gimble and outgrabe are verbs; and toves, wabe, borogoves and raths are nouns. The psychological phenomenon of what has been called the *effort after meaning* [10] also plays a part. Humans have a tendency to strive to see sense in anything that is presented as a message. This basic urge is also seen in well-attested phenomenon such as *pareidolia*, which often manifests itself in such things as seeing faces or objects in clouds, or hearing words in random sounds as in the Electronic Voice Phenomena popularised by "paranormal investigators".

A major factor in this semblance of sense is that, even though such words as brillig etc. do not exist in the lexicon of English, they do look like English words. This is because they are made up of possible letter combinations in English such as br, ll, sl, ith, oves etc. (compare for example, Italian words like *gnocchi*, *ahimé* or *sforza* which do not use English letter patterns). Finally, and most ingeniously on the part of Carroll perhaps, the phonesthesia or phonetic symbolism of some of the words themselves hints at what their meaning might be and thus entails encoding at a whole new level compared to what we discuss elsewhere in this article. For example, slithy is reminiscent of slither. Mimsy, with its

un completo nuovo livello confrontato con quello che abbiamo discusso in altre parti di questo articolo. Nel testo originario, *mimsy*, con il suono corto /ɪ/ (prodotto quando la lingua è alta e nella posizione avanzata della bocca, che crea una piccola cavità risonante che amplifica alcune alte frequenze) fa pensare, almeno in Inglese, a qualche cosa di piccolo o delicato (come *bit*, *little*, *twitter*). *Mome*, in contrasto con il suono /əʊ/ un dittongo, o combinazione del suono di due vocali, il secondo dei quali (/ʊ/) prodotto con la lingua abbassata nel retro della bocca (risultando in una grande cavità che amplifica le basse frequenze) dà l'impressione di qualche cosa di più grande (ad esempio *row*, *tome*, *stone*) ([11], pag. 166-167).

L'analisi delle molte e varie sequenze che possono essere rilevate nel linguaggio naturale è diventata una parte importante della criptoanalisi. È anche un modo utile per distinguere testi realmente codificati da linguaggi di puro *nonsense*, pseudotesti, come quelli trasmessi da numerose stazioni utilizzate per confondere gli operatori e i criptonalisti del SIGINT (si veda la sez. 2.1). Forse il più noto esempio di questo esercizio nel distinguere senso e non-senso linguistico sono i tentativi di analizzare il cosiddetto Manoscritto di Voynich. Battezzato così dal nome di un libraio polacco, è la pergamena di un codice illustrato risalente al XV secolo, come indica una datazione al radiocarbonio¹³. Comprende più di 170 mila caratteri scritti a mano in alfabeto e linguaggio sconosciuti. Questo fornisce ai criptoanalisti un enorme mole di lavoro. Nonostante questo, nessuno è ancora riuscito a decodificarne in maniera conclusiva, anche una minima parte. La teoria che sia una burla, contenente un linguaggio privo di senso, è stata presentata da molti, ma il fatto che mostri caratteristiche simili a quelle di una lingua nella consistenza e frequenza nella sequenza di certe parole e lettere ha incoraggiato molti studiosi nel credere che si tratti di una genuina lingua. Negli anni '40 e '50 del secolo scorso, i criptoanalisti della US National Security Agency guidati dal leggendario William F. Friedman, ex capo del US Signals Intelligence Services

short /ɪ/ sound (produced when the tongue is high and at the front of the mouth, which creates a small resonant cavity that amplifies some higher frequencies) makes, in English at least, one think of something small or delicate (e.g. *bit*, *little*, *twitter*). *Mome*, by contrast, with its /əʊ/ sound, a diphthong (or combination of two vowels sounds), the second of which (/ʊ/) produced with the tongue low down to the back of the mouth (resulting in a large resonant cavity that amplifies low frequencies) gives the impression of something larger (e.g. *crow*, *tome*, *stone*) ([11], pag. 166-167).

Analysis of the many and diverse patterns that can be detected in natural language has become an important part of cryptanalysis. It is also a useful way of distinguishing between texts in real unknown codes or languages from real nonsense, pseudo-texts, such as those transmitted by some numbers stations used to confuse SIGINT operatives and cryptanalysts (see § 2.1). Perhaps the best-known example of such an exercise in telling the difference between linguistic sense and nonsense are the attempts at analysis of the Voynich Manuscript. Named after a Polish book dealer, this is an illustrated codex whose parchment originates, so radiocarbon dating shows, in the early 15th century¹². It comprises more than 170,000 handwritten characters in an unknown alphabet and language. It thus gives cryptanalysts a huge amount of material to work with. Despite this, no one has succeeded in decoding even a small portion of it conclusively. The theory that it is an elaborate hoax, constituting mere gibberish, has been put forward by many, but the fact that it does show language-like characteristics in its consistency and in the frequency of certain "words" and letter patterns has encouraged many scholars to believe that it is an example of a genuine language. In the 1940s and 50s, cryptanalysts working at the US National Security Agency led by the legendary William F. Friedman, former head of the US Signals Intelligence Services

¹³Una riproduzione digitalizzata ad alta risoluzione dell'opera completa, conservata nella libreria Beinecke di libri rari della Università di Yale, può essere consultata al sito: <https://archive.org/details/voynich>

¹²High resolution scans of the complete work, kept in the Yale University Beinecke Rare Book And Manuscript Library, can be viewed at: <https://archive.org/details/voynich>.

fallirono nel decifrarne anche una piccola parte, ma non riuscirono a scartarlo come una burla¹⁴.

Frequenze e sequenze possono essere usate per distinguere tra ricorrenze casuali e significative in senso statistico¹⁵, come nell'analisi dei segnali radio che arrivano da galassie lontane nella ricerca di vita extra-terrestre. Possono essere inerenti alla lingua e possono fornire importanti indizi ai crittoanalisti (almeno possono indicare in quale lingua sia scritto il testo originale cosa che sono stati incapaci di individuare con il codice Voynich). I crittografi si sono gradualmente resi conto di questo, e hanno concepito metodi in modo che queste sequenze, osservabili anche nel manoscritto di Voynich, possano essere mascherate. I principi soggiacenti nomenclatura e poligrafia possono essere rimodulati per questo scopo. Un notevole sviluppo fu un'invenzione, attribuita al Duca di Mantova, nel XV secolo, dell'omonimo codice dove le comuni lettere, o loro combinazioni, nel testo originale erano indicate da diversi caratteri o simboli nel testo cifrato secondo la loro frequenza, in questo modo, rendendoli più difficili da decifrare. Ad esempio, si può decidere di codificare la lettera a in due diverse maniere, sia come d o come £, e t come w e &. In questo modo, il nostro esempio "We attack at dawn", può diventare "zh dw&£fn dw gdzq" o "zh £&wdfn d& gdzq" e così via.

Nel XX secolo, i metodi di criptaggio implicano anche l'uso di diversi codici aleatori per ogni lettera o suono (o una loro combinazione) con cose

failed to decipher any of it, but could not dismiss it as a hoax¹³.

Frequencies and patterns can be used to differentiate between random occurrences and significance (in a statistical sense)¹⁴, as in the analysis of radio signals coming from distant galaxies in the search for extra-terrestrial life. They are also integral to language and can provide valuable clues to cryptanalysts (at least when they can guess which language the plaintext is in, which they have been unable to do with the Voynich manuscript). Cryptographers gradually realised this and devised ways so that such patterns, observable even in the Voynich manuscript, can be camouflaged. The principles underlying nomenclators and polygraphs could be repurposed for this purpose. A notable development was the invention, attributed to the Duca di Mantova, in the 15th century of the homophone cipher whereby common letter or letter combinations in the plaintext were subtitled by different characters or symbols in the ciphertext according to their frequency, in this way, making them harder to identify. For example, one may decide to encode the letter a in two different ways, either as a d or a £, and t as w and &. In this way, our example "We attack at dawn", could become "zh dw&£fn dw gdzq" or "zh £&wdfn d& gdzq" and so on.

In the 20th century, encryption methods also involved the use of different randomised ciphers for each letter or sound (or combinations thereof),

¹⁴Anche il famoso crittoanalista britannico John Tiltman (si veda la sez. 3.0) ha aiutato la squadra del NSA.

¹⁵Nella prima categoria possiamo classificare alcune discutibili applicazioni della matematica nell'analisi di sequenze in testi come l'ampiamente pubblicizzato codice della Torah [12]. Si afferma di aver identificato un insieme di parole codificate nascoste nel libro ebraico della Torah secondo Drosnin [12] che messe insieme hanno predetto importanti eventi storici. Altri analisti, tuttavia, hanno messo in evidenza che scoperte simili a quelle di Drosnin possono essere ritrovate in qualunque testo molto lungo, e quindi i suoi risultati sono lontani dalla significatività statistica [13].

¹³The famous British cryptanalyst, John Tiltman (see § 3.0) also assisted the NSA team.

¹⁴And in the former category, we can class some rather dubious applications of mathematics analysis of patterns in texts, such as the much publicised Bible of Torah code [12]. This claims to have identified a set of encoded words hidden within the Hebrew text of the Torah according to Drosnin [12], put together, these words have predicted significant historical events. Other analysts however have pointed out that similar findings to those of Drosnin may be found in any long text, and thus his results are far from statistically significant [13].

come one-time pad¹⁶, o macchine capaci di generare un numero quasi infinito di diverse combinazioni rendendo inutili analisi frequenziali fatte con carta e penna.

In queste sezioni (2.0-2.2), ci siamo limitati ad alcuni principi di base di codici relativamente semplici, quelli che esistevano prima del ventesimo secolo e l'avvento di congegni automatici di decifrazione ad alta velocità. Nella sezione 3.0, tratteremo uno dei codici più sofisticati ed importanti usati nella Seconda Guerra Mondiale allo scopo di illustrare come la matematica e la linguistica sono, insieme, state usate nella criptonalisi.

3 Il cifrario di Lorenz (Tunny)

Al tempo della Seconda Guerra Mondiale, che fu la prima guerra in cui le forze armate si affidavano alle telecomunicazioni su grande distanza usando la radio, l'arte della cifratura si è evoluta in una scienza. Decifrare messaggi quindi diventò molto più difficile e al di là delle capacità di parolieri, poliglotti, entusiasti enigmisti, giocatori di carte ed eccentrici assortiti che i militari avevano tradizionalmente trasformato in criptoanalisti¹⁷.

In questa sezione, mi concentrerò su un esempio, quello che ha condotto alla costruzione di Colossus, menzionato nella sezione 2.0, e analizzerò il processo che ha condotto alla decifrazione del codice tedesco Lorenz (Tunny per gli alleati). Questo è meno noto di Enigma, lo spionaggio lo

¹⁶Il moderno sviluppo del cifrario di Vigenère (sez. 2.1), questo è un metodo ampiamente usato consistente una chiave fatta di stringhe di caratteri o simboli selezionati aleatoriamente che si combinano con le lettere nel messaggio originale con un'addizione modulata (come le ore del giorno). Se la chiave è: tenuta segreta; lunga almeno come il messaggio stesso (quindi evita le ripetizioni trovate nel cifrario di Vigenère); è usata solo una volta; ed è veramente aleatoria; si ritiene quindi indecifrabile.

¹⁷Tra questi c'era una grande proporzione di donne anche all'inizio del moderno spionaggio dei segnali. Nella Prima Guerra Mondiale, fino ad un terzo del personale impiegato nel britannico SIGNINT era femminile; questa percentuale è salita alla metà nella Seconda Guerra Mondiale. Le donne superarono in numero gli uomini in molti contesti lontani dalle linee del fronte [14]. A Bletchley Park il 75% dei criptoanalisti era composto da donne tra le quali spiccano figure come Mavis Batey, Joan Clarke, Jane Fawcett, e Jean Valentine [15, 16]

with things like one-time pads¹⁵, or machines able to generate an almost infinite number of different combinations making frequency analysis of the pen and paper kind useless.

In these sections (§2.0-2.2), we have limited ourselves to some of the basic principles of relatively simple codes, those that existed before the twentieth century and the advent of high speed automated encryption devices. In Section 3.0, we will turn to one of most sophisticated and important codes used in WW2 as a way of illustrating how mathematics and linguistics were used in conjunction with each other in cryptanalysis.

3 The Lorenz (Tunny) cipher

By the time of WW2, which was the first war where armed forces were relying on telecommunications over vast distances using radio, the art of encryption had evolved into a science. Decrypting messages thus became much more difficult and beyond the ability of the sundry word-smiths, polyglots, word puzzle enthusiasts, card players, and sundry eccentrics who the military had traditionally turned to for cryptanalysis.¹⁶.

In this section, we will concentrate on one example, the same which led to the building of Colossus, mentioned in §2.0, and look at the processes leading to the cracking of one German cipher, Lorenz (Tunny to the Allies). This is less known than Enigma, and the intelligence labelled Ultra which it provided, but it was a far more complex code using quicker, more powerful en-

¹⁵A modern development of the Vigenère cipher (§2.1), this is a widely used method consisting of a key made up of a string of randomly selected characters or symbols which combine with letters in the original message by modulation addition (like hours in a day). If the key is: kept secret; is at least as long as the message itself (thus avoiding the repetition found in the Vigenère cipher); is only used once; and is truly random; then it is thought to be undecipherable.

¹⁶Among these were a high proportion of women even from the beginnings of modern signals intelligence. In WW1, up to a third of those engaged in British SIGNINT were female; by WW2, this had risen to a half. Women actually outnumbered men in many contexts away from the front line [14]. At Bletchley Park, 75% of the cryptanalysts were female among whom such key figures as Mavis Batey, Joan Clarke, Jane Fawcett, and Jean Valentine [15, 16]

ha etichettato Ultra ed era un codice molto più complesso che usava macchine di decifrazione più veloci e più potenti che richiedevano un minor numero di operatori (la macchina Enigma richiedeva, ad ogni terminale, tre operatori istruiti per inviare e ricevere).

Diversamente da Enigma (che usava un codice Morse) Tunny era basato su un sistema di telescriventi facile da usare che incorporava il codice di tipo binario delle telescriventi (anche noto come codice Baudot-Murray). Questo implicava la perforazione di un nastro di notevole spessore¹⁸. Il mittente digitava il testo non criptato e anche l'operatore dall'altra parte riceveva il messaggio non criptato. Dietro le quinte, c'erano due diversi strati di criptazione. La prima usava cinque ruote, e poi una seconda chiave era applicata ad altre cinque ruote. Due altre ruote generavano un "balletto", introducevano irregolarmente una serie di caratteri che avrebbero introdotto un'aleatorietà apparente nella chiave¹⁹. Tunny era particolarmente difficile da gestire perché, a differenza di Enigma, nessuno, dalla parte degli alleati, aveva mai visto l'apparato che lo produceva, il Lorenz-Schlüsselzusatz SZ40/42, e questo avvenne solo al termine delle ostilità.

¹⁸Da questo il nome Tunny (pesce tonno): i britannici chiamarono con questo nome tutti i codici diversi dal Morse

¹⁹Il numero totale di possibili configurazioni di criptazione che il Lorenz SZ42 avrebbe potuto generare era 1.6×10^{19} ($23 \times 26 \times 29 \times 31 \times 37 \times 41 \times 43 \times 47 \times 51 \times 53 \times 59 \times 61$)

cryption machines that required fewer operators (the Enigma machine required three trained operators at each end to send and receive).

Unlike Enigma (which used Morse Code), Tunny was based on a simple-to-use wireless teleprinter system that incorporated the binary-type code of the teleprinter (a.k.a. Baudot-Murray Code). This involved punching holes into ticker tape¹⁷ the sender typed in plaintext and the operator at the other end received a plaintext message. Behind the scenes, it used two different layers of enciphering. The first cipher used five wheels and then a second key was applied using another five wheels. An additional two wheels would generate a "stutter", an irregularly inserted character that would introduce apparent randomness into the key¹⁸. Tunny was particularly difficult to deal with because, unlike the Enigma machine, no one on the Allied side had even seen the device that produced it, the Lorenz-Schlüsselzusatz SZ40/42, and were only to do so at the very end of hostilities.

¹⁷Hence, its codename Tunny (tuna fish): the British codenamed all non-Morse ciphers after types of fish.

¹⁸The total number of enciphering possibilities that the Lorenz SZ42 could generate was 1.6×10^{19} ($23 \times 26 \times 29 \times 31 \times 37 \times 41 \times 43 \times 47 \times 51 \times 53 \times 59 \times 61$)

Testo originale		Carattere casuale		Testo cifrato		Carattere casuale		Testo originale
M		Q		J		Q		M
•	+	X	=	X	+	X	=	•
•	+	X	=	X	+	X	=	•
X	+	X	=	•	+	X	=	X
X	+	•	=	X	+	•	=	X
X	+	X	=	•	+	X	=	X

Tabella 1: Processo di criptazione additivo per il cifrario Tunny.
Basic Additive Encryption Process for Tunny cipher.

3.1 Progressi iniziali nel decifrare Tunny

Tunny usava un cifrario additivo, in cui una lettera casuale era aggiunta alla lettera del testo originale. Ad esempio, la lettera M nel testo originale (nel codice della telescrivente ••XXX) può essere combinata (per mezzo di una tavola della verità / una operazione XOR) con un carattere casuale Q (XXX•X), dando XX•X²⁰, che è il codice di telescrivente per J. La lettera M è quindi criptata come J nel testo cifrato. Per decifrare il testo, la macchina deve aggiungere lo stesso carattere casuale (Q) a J risultando ••XXX: il codice da telescrivente per M come nella tabella 1²¹.

Per la loro complessità le macchine Lorenz - Schlüsselzusatz SZ40/42, erano riservate ai più altri gradi del *Oberkommando der Wehrmacht* (OHW) Alto Comando Tedesco. Solo 30 erano in operazione (da confrontare con le circa mille macchine Enigma nelle varie versioni).

I messaggi intercettati venivano dati al leggendario decifratore e brillante linguista (poliglotta) brigadiere John Hessell Tiltman. Lui dedusse, correttamente, che Tunny era basato su una specie di codice Vennam (che usa un passaggio come operazione XOR come nella tabella 1). Inizialmente, affinché il ricevente possa sapere come regolare la sua macchina, il mittente deve trasmettere il settaggio della macchina con il messaggio (cosa che compromette la sicurezza). Questo fu fatto ponendo quello che i criptonalisti del Bletchley Park chiamarono un indicatore all'inizio della trasmissione. Dal 1942, quando Lorenz SZ40/42 divenne pienamente operativo, il sistema di indicatore consistette nella trasmissione in codice non criptato delle lettere "QEP" seguite da un numero di due cifre. Questo poteva essere identificato consultando un libro di codici noto ad entrambi gli operatori e mostrava il settaggio delle dodici ruote.

Infatti, quando due messaggi usano la stessa chiave, se sono addizionati o confrontati con un'operazione XOR (come nella tabella 1), allora gli effetti della chiave sono eliminati [17]. Que-

²⁰Il carattere X significa che due simboli confrontati sono differenti; • significa che sono gli stessi: X + X o • + • producono entrambi • e X + • o • + X risulta in X.

²¹Adattata dal sito Bill Tutte Memorial Fund (<https://billtuttememorial.org.uk/codebreaking/teleprinter-code>)

3.1 Initial progress in cracking Tunny

Tunny used an additive cipher, whereby a random letter was added to the letter in the plaintext. For example, the letter M in the plaintext (in teleprinter code, ••XXX) can be combined (by means of a truth table / XOR operation) with the random character Q (XXX•X), giving XX•X¹⁹, which is teleprinter code for J. M would thus be encrypted as J in the ciphertext. To decrypt the text, the machine would add the same random character (Q) to J resulting in ••XXX: the teleprinter code for M (as in table 1)²⁰.

Given their complexity, Lorenz - Schlüsselzusatz SZ40/42, machines were reserved for the upper echelons of the German *Oberkommando der Wehrmacht* (OHW) High Command. Only 30 were in operation (compared to thousands of Enigma machines of different versions).

Intercepted messages were given to the legendary codebreaker and brilliant linguist (polyglot) Brigadier John Hessell Tiltman. He correctly surmised that Tunny was based on a kind of Vennam cipher (that is, like a one-time pad employing a XOR operation, as in table 1). Initially, in order for the receiver to know how to set their machine, the sender had to transmit the machine settings within the message (which obviously compromised security). This was done by putting what Bletchley Park cryptanalysts called an indicator at the beginning of the transmission. From 1942, as Lorenz SZ40/42 came into full operation, the indicator system consisted of the transmission of the plaintext letters "QEP" followed by a two digit number. This could be looked up in a codebook issued to both operators and showed the settings of the twelve wheels.

Indeed, when two messages use the same key, if they are added together or compared in a XOR operation (as in table 1), then the effect of the key

¹⁹The character X means that the two symbols being compared are different; • means they are the same: X + X or • + • would both result in •, and X + • or • + X would result in X.

²⁰Adapted from the Bill Tutte Memorial Fund website (<https://billtuttememorial.org.uk/codebreaking/teleprinter-code>)

sto significa che il risultato dell'addizione di due testi non cifrati dovrebbe essere lo stesso di due testi cifrati. In teoria, la chiave potrebbe essere recuperata dalla coppia dei messaggi cifrati e non. Questo è un processo lungo e laborioso se fatto a mano su ogni insieme di caratteri corrispondenti che sono stati generati dalla stessa chiave. Si può calcolare la somma dei caratteri cifrati (cioè il codice della telescrivente che corrisponderebbe all'operazione XOR dei due caratteri cifrati), ma non c'è un algoritmo matematico per decomporlo e trovare quale carattere del testo originale possa essere aggiunto per produrlo (allo stesso modo in cui non possiamo determinare in maniera univoca quali siano i due numeri da sommare per fare, diciamo 34: $14 + 20$, $33 + 1$, $9 + 25$ etc.).

Tiltman, essendo un dotato paroliere di vecchia scuola, si affidava ad intuizioni linguistiche e perseveranza. Ad esempio, se la somma di due testi cifrati è RSEZLS, e si immagina che si riferisca ad una grande città della Gran Bretagna, allora, ricordando che RSEZLS è la somma di due testi non criptati si può fare una ragionevole ipotesi che una di loro sia, diciamo, London. Per verificarlo, si possono aggiungere i codici da telescriventi per RSEZLS e LONDON (come nella tabella 1) che produrrebbe OXFORD²². In questo modo, si può essere sicuri che si sono identificati con successo i due testi non criptati dato che hanno senso compiuto nel contesto, ed è estremamente improbabile che sia il prodotto di un effetto casuale. Questi metodi si rivelarono efficaci, ma troppo lenti e avrebbero funzionato solo se si fosse in qualche modo dedotto il tipo di contenuto di ciò di cui si parlava.

Nell'agosto del 1941, i Britannici intercettarono due messaggi quasi identici inviati da un operatore negligente che fu forzato a rimandare un lungo messaggio di 4 mila caratteri. Nella sua fretta, o pigrizia, lui / lei non cambiò il settaggio della macchina Lorenz SZ40/42. Il fatto che ci fossero delle piccole differenze tra le due trasmissioni fu cruciale poiché due messaggi erano sicuramente meglio di uno. Due messaggi quasi identici fornivano profondità informativa:

²²Questo esempio è dato da Frank Carter, a Bletchley Park Advisor, nell'eccellente documentario: *Timewatch, Code-Breakers: Bletchley Park's Lost Heroes* (BBC 2011).

is eliminated [17]. This means that the result of the addition of the two plaintexts should be the same as that of the two ciphertexts. In theory, the key can then be recovered from either ciphertext / plaintext pair. This is a long and laborious process if done by hand on every set of corresponding characters in two messages that have been generated by the same key. One is able to calculate the sum of the ciphertext characters (that is the teleprinter code that would correspond to the result of the XOR operation of the two ciphertext characters), but there is no mathematical way to decompose this and find out which plaintext characters could be added together to produce it (in the same way that what cannot determine mathematically which two numbers have been added together to make, say, 34: $14 + 20$, $33 + 1$, $9 + 25$ etc.).

Tiltman, being an exceptionally gifted old-school wordsmith, relied on inspired guesswork, linguistic intuition, and perseverance. For example, if the sum of two ciphertexts is RSEZLS, and one suspects that each refers to a major British city, then, remembering that RSEZLS is the sum of the two plaintexts, one may make an educated guess that one of them is, say, London. To check this, one may add together the teleprinter codes for RSEZLS and LONDON (as in table 1), which would give OXFORD²¹. In this way, one could be fairly sure that one had successfully identified the two plaintexts, as they both made sense in the context, and were extremely unlikely to be a product of random chance. Such methods proved effective, but were slow and would only work if one could somehow deduce the kind of content one was dealing with.

In August 1941, the British intercepted two nearly identical messages sent by a sloppy operator who had been forced to resend a long 4,000-character message. In his / her haste, or out of laziness, he / she omitted to change the settings on the Lorenz SZ40/42 machine. The fact that there were minor differences between the two transmissions was crucial as two identical texts would of course have been no better than one. It meant that the almost identical messages pro-

²¹This example is given by Frank Carter, a Bletchley Park Advisor, in the excellent documentary: *Timewatch, Code-Breakers: Bletchley Park's Lost Heroes* (BBC 2011).

costituivano due diversi criptaggi generati dallo stesso settaggio (piuttosto che due possibilità di 1.6 miliardi di miliardi). In sintesi, fornivano una rara sequenza di 8 mila caratteri sulla quale Tiltman poteva lavorare.

Tiltman ipotizzò che il messaggio iniziasse con la frase SPRUCHNUMMER "numero del messaggio". Quando fu richiesto di ripetere, l'operatore impaziente lo accorciò in S P R U C H N R "messaggio no.". Questa piccola differenza significò che tutto ciò che seguiva la N nel secondo testo era diverso da quello della prima trasmissione. Tiltman continuò ad applicare la stessa tecnica, come illustrato nella tabella 1, usando la stessa chiave già adoperata per gli altri testi che fornivano la profondità informativa. Fu capace di verificare le sue intuizioni per ognuno dei testi dato che sapeva che criptavano, fondamentalmente, lo stesso messaggio. Questo gli dette la certezza che stava procedendo bene invece che trovare delle coincidenze accidentali. Tiltman non era un matematico e non aveva idea di come la macchina Lorenz S40/42 potesse funzionare. Ci vollero 10 giorni per decifrare questi due messaggi (comunque una notevole impresa). Aveva perfezionato il metodo, ma non aveva la possibilità di ottenere una descrizione matematica della chiave che genera il processo. Se si fosse trovata, forse il processo di decifrazione si sarebbe velocizzato, o, addirittura, automatizzato.

3.2 Usare la matematica per ricostruire la macchina Lorenz S40/42

Tiltman diede questo compito al nuovo arrivato, il ventiquattrenne laureato in chimica William "Bill" Tutte, che lavorò nella divisione di cui lui era il capo: la Research Section²³. Come parte del suo addestramento come criptoanalista, a Tutte era stata insegnata la tecnica d'esame di Kasiski, inventata da un ufficiale prussiano nel

²³Umile di carattere, i talenti di Tutte non furono immediatamente evidenti. Fu esaminato da Alan Turing e bocciato come candidato per lavorare ad Enigma. Nella sua vita successiva, ottenne un PhD in Matematica da Cambridge e diventò professore all'Università di Waterloo nell'Ontario, dove contribuì a fondare il Dipartimento di Combinatoria e Ottimizzazione. Il suo ruolo nella decifrazione di Tunny non fu reso pubblico fino al 1990, molto dopo il suo pensionamento.

vided depth: they constituted two different encryptions generated by the same setting (rather than two randomly out of the 1.6 billion billion [sic] other possibilities). In short, they provided a rare 8,000 worth of characters with which Tiltman could work.

Tiltman guessed that the message began with the phrase SPRUCHNUMMER "message number". When asked to repeat, the impatient operator shortened this to S P R U C H N R "message no.". This tiny difference meant that everything from the first N in the second text was different from that in the first transmission. Tiltman was able to continue with the same additive technique, as illustrated in table 1, working through the ciphertext as he had with other depths that he had previously worked upon. This gave him the certainty that he was on the right track rather than just finding coincidental matches. Tiltman was no mathematician and had little idea of how the Lorenz S40/42 machine might work. It took him 10 days to decrypt just these two messages (nonetheless a remarkable feat). He had however, perfected a method, but had no way to derive a mathematical description of the key generating process. If one could be found then perhaps the decryption process could be speeded up, perhaps even automated.

3.2 Using mathematics to reconstruct the Lorenz S40/42 machine

Tiltman gave this task to a newly-arrived 24 year-old chemistry graduate, William "Bill" Tutte, who worked in the unit he was head of: the Research Section²². As part of his training as a cryptanalyst, Tutte had been taught the Kasiski examination technique which had been invented by a nineteenth century Prussian officer to crack

²²Humble in character, Tutte's talents were not immediately apparent. He had been interviewed by Alan Turing and rejected as a candidate to work on Enigma. In later life, he obtained a PhD in Mathematics from Cambridge and became a professor at Waterloo University, Ontario, where he helped found the Department of Combinatorics and Optimization. His role in cracking the Tunny was not public knowledge until the 1990s, long after his retirement.

XIX secolo per decifrare il codice di Vigenère (sez. 2.1). Questo metodo implica di cercare ripetizioni di stringhe di tre o più caratteri in un testo cifrato e misurare la distanza tra loro. È probabile che queste siano multipli della lunghezza della parola chiave, e, nel caso del cifrario Tunny permettono di invertire il processo che ha criptato il messaggio. Usando questo metodo laborioso, che era scomodo con lunghe sequenze che non potevano facilmente essere dispiegate su carta, Tutte, come Tiltman, ricorse a ragionevoli ipotesi. Dagli indicatori trasmessi con il messaggio (QEP più 2 cifre) fece la congettura che gli indicatori di Tunny fossero di 25 lettere (escludendo la J) per un settaggio e di 23 lettere per l'altro. Quindi, Tutte applicò la tecnica Kasiski sul primo impulso usando una ripetizione di 575 (cioè 25×23). Questo metodo non funzionò, eppure Tutte notò delle evidenti ripetizioni con un andamento non verticale, come previsto da questo metodo, ma leggermente diagonale. Quindi provò ancora con 574 con maggiore successo. I fattori primi di 574 sono 2, 7 e 41, quindi riprovò con una riga di 41 quadrati sul suo tavolo e riuscì ad osservare nelle colonne una moltitudine di sequenze ripetute indicando che la sua intuizione era probabilmente corretta.

Questo fu solo il primo passo di un lungo percorso. Era chiaro che la macchina era formata da ruote separate, come Enigma, e Tutte dovette trovare quante fossero, e quale ruolo avessero nel processo di crittaggio. Fu capace di calcolare che c'erano due gruppi di cinque ruote ed altre due individuali.

Tutte indicò la prima componente della chiave come χ (chi) e immaginò che ci fosse un'altra componente che le era legata attraverso un'operazione XOR. Questo tuttavia non cambiava con ogni carattere (dato che le due ruote aggiuntive creavano un "balbettio"). Il risultato di questa ruota era detto ψ (psi). Trovò che questo assetto si applicava ad ognuno degli impulsi prodotti dai due gruppi di cinque ruote. Quindi, per un carattere individuale, la chiave era il risultato dell'operazione XOR di χ e ψ . La sequenza di caratteri aggiunti dalle ruote ψ era detto ψ esteso. Tutte riuscì a derivare le componenti di ψ grazie ad un difetto della macchina Lorenz S40/42 (corretto in seguito dai costruttori) che implicava che punti e croci fossero lievemente più proba-

Vigenère ciphers (§ 2.1). This method involves looking for repetition of strings of three or more characters in the ciphertext and measuring the distances between them. These are likely to be multiples of the length of the keyword, and, in the case of Tunny cipher, allow one to start to reverse engineer the way that it encrypted messages. Using this laborious method, which was cumbersome with longer sequences that could not be set out so easily on paper, Tutte, like Tiltman, resorted to educated guesses. He surmised from the indicators transmitted with the message (QEP plus 2 digits) that the Tunny indicators used 25 letters (excluding J) for one setting and 23 letters for the other. Tutte therefore tried a Kasiski examination on the first impulse of the key characters using a repetition of 575 (i.e. 25×23). This did not work but he did notice some obvious repetitions not on the vertical, as this method would have envisaged, but on a slight diagonal. He therefore tried again with 574, with more success. The prime factors of this number 574 are 2, 7 and 41, so he tried again with a row of 41 squares on his table and was able to observe in the columns a multitude of repeated patterns indicating that his intuition had probably been correct.

This was only the first step of many. It was clear that the machine consisted of separate wheels, like Enigma, and Tutte had to set about finding out how many more there were, and what part they each played in the encryption process. He was able to calculate that there were two sets of five wheels and two more individual wheels.

Tutte called the first component of the key χ (chi) and speculated that there was another component which was linked to it via an XOR operation. This however did not always change with each character (because of the two additional wheels that created a "stutter"). He called the product of this wheel ψ (psi). This arrangement, he found, applied for each of the impulses produced by the sets of five wheels. So for a single character, the key was the product of the XOR operation of χ and ψ . The sequence of characters added by the ψ wheels was referred to as the extended ψ . Tutte was able to derive the ψ component because a flaw in the key setting of the Lorenz S40/42 machine (one later rectified by the manufacturers) meant that dots and

bilmente ripetuti (cioè seguiti da un altro punto o una croce). Avendo fatto questa scoperta, il resto della Research Section lo aiutò nello studio degli altri impulsi. Si trovò che le cinque ruote ψ si muovevano insieme sotto il controllo di due ruote - motore separate che chiamarono μ (mu).

In realtà, leggere un messaggio Tunny richiedeva non solo riuscire ad afferrare la struttura logica del sistema, come avevano fatto Tutte ed il resto della Research Section, ma implicava anche che si potesse derivare la sequenza cangiante delle camme (o dei *pin*) delle ruote²⁴. Infine bisognava stabilire la sequenza iniziale (o la chiave del messaggio) del mescolatore delle ruote. La derivazione delle sequenze delle camme ("*wheel breaking*") era lo scopo della procedura chiamata *Turingery* dal suo inventore Alan Turing, che la propose nel 1942. Inutile dire che, insieme allo stabilire la sequenza iniziale, anche questo fu un altro compito molto difficile. Tutti questi passaggi furono inizialmente eseguiti laboriosamente a mano, ma avrebbero potuto eventualmente essere automatizzati, se gli alleati fossero stati capaci di leggere i messaggi Tunny abbastanza velocemente da capitalizzare il prezioso spionaggio che fornivano.

La storia del progetto e della costruzione di Colosso, il primo calcolatore elettronico programmabile, è un'altra storia (questa volta nel campo dell'ingegneria) del trionfo di conoscenza, intelletto, ingenuità e perseveranza contro le avversità. Max Newmann, il matematico, si rese conto che il metodo di Tutte (una parte del processo di decriptazione di Tunny) poteva essere meccanizzato ma, avendo poca esperienza in ingegneria, la macchina che aveva costruito continuava a surriscaldarsi e a fermarsi. Tommy Flowers (un giovane ingegnere delle Poste che aveva una profonda conoscenza di valvole termoioniche e tubi a vuoto) fu coinvolto. Invece di aggiustare la macchina di Newton, progettò dall'inizio una macchina più efficiente e affidabile che funzionò

crosses were marginally more likely to be repeated (i.e. followed by another dot or cross). Having made this breakthrough, the rest of the Research Section could assist him in studying the other impulses. It was found that the five ψ wheels all moved together under the control of the two separate motor wheels, which they labelled μ (mu).

In fact, reading a Tunny message required not only that one had grasped the logical structure of the system, as Tutte and the rest of the Research Section had now done, it also demanded that one could derive the periodically changed pattern of active cams (or pins) on the wheels²³. Finally, one had to establish the starting positions (or message key) of the scrambler wheels. The derivation of the cam patterns ("*wheel breaking*") was the target of a procedure called *Turingery* after its inventor Alan Turing, who came up with it in July 1942. Needless to say, together with the establishment of the starting settings, this was yet another highly complex process. All these different stages initially had to be done laboriously by hand, but could eventually be automated, indeed urgently needed to be done so if the Allies were to be able to read Tunny messages quickly enough to capitalise on the precious intelligence that they provided.

The story of the design and construction of the Colossus, the world's first programmable electronic computer, is another tale (this time in the field of engineering) of the triumph of insight, intellect, ingenuity, and sheer perseverance against incredible odds. Max Newman, the mathematician, had realised that Tutte's method (just one part of the process for cracking Tunny) could be mechanised but, with little experience of practical engineering, the machine that he built kept overheating and breaking down. Tommy Flowers (a young General Post Office telecoms engineer who had a thorough practical knowledge of things like valves and vacuum tubes) was brought in. Instead of fixing Newman's machine, he designed a more efficient and reliable one from

²⁴Se incrementate producevano una X (1 in binario); decrementate generavano uno spazio (0 in binario) o • nella tabella 1.

²³If raised, they generated an X (1 in binary); lowered, they generated a space (0 in binary) or • on table 1.

perfettamente la prima volta²⁵.

Senza dubbio lo sforzo valeva la pena. In un solo anno, gli Alleati conoscevano le intenzioni e le direttive dell'alto comando tedesco, a volte anche prima che fossero note ai comandanti di basso livello. Grazie a Colossus, i britannici riuscirono a passare ai Sovietici l'intero ordine di battaglia dei tedeschi per il loro attacco a Kursk nel Luglio 1943, senza rivelare i progressi che avevano fatto nella criptoanalisi²⁶. Questa enorme battaglia, il più grande scontro di mezzi corazzati della storia, fu vinta dall'Unione Sovietica, e segnò l'inizio dell'inarrestabile avanzata dell'Armata Rossa su Berlino. La capacità di decifrare Tunny fu vitale anche nello sbarco del D-Day. Tra le altre cose, mostrò agli alleati che il massiccio piano per ingannare l'Asse sull'eventuale luogo dello sbarco ebbe successo. Lo spionaggio Tunny confermò che lo stesso Hitler era convinto, contro la più corretta opinione di alcuni suoi generali, che lo sbarco in Normandia del 6 Giugno fosse solo un diversivo.

4 Conclusion

La criptoanalisi è una delle aree dove la matematica è rilevante per la linguistica. È chiaramente un caso dove un approccio matematico può essere combinato con uno linguistico per pro-

²⁵L'esistenza di Colossus, ed il lavoro dei suoi creatori, Newman e Flowers fu mantenuto in stretto riserbo fino agli anni '70. Nel 1980, Flowers che, dopo la guerra, era ritornato a lavorare per le poste, ricevette un riconoscimento dovuto: fu il primo vincitore della medaglia Martlesham per i suoi risultati nell'informatica. Nel 1993 ricevette un semplice certificato in informatica come programmatore di *personal computer* dall'Hendon College (una scuola per adulti): questa è la sola qualifica in informatica che riuscì ad ottenere l'ingegnere responsabile del primo calcolatore elettronico programmabile.

²⁶Il generale Zhukov, il comandante Sovietico, era convinto che i britannici avessero una talpa nel quartier generale dell'esercito tedesco (alimentando il sospetto che gli alleati occidentali potessero raggiungere un accordo di pace separato con i Nazisti), così impressionato con le informazioni che aveva ricevuto. Parte dei motivi per la segretezza nel lavoro di decodifica Tunny e dell'esistenza di Colossus poteva essere stata che i britannici sospettavano che nel mondo post-bellico i Sovietici (di cui Churchill diffidava sempre più) potessero cominciare ad usare Lorenz, o qualche cosa di analogo.

scratch. Amazingly, it worked perfectly first time²⁴.

Undoubtedly, all the effort was worth it. Within only a year, the Allies were learning about the intentions and directives of the German High Command, sometimes before Axis commanders lower down in the field were. Because of Colossus, the British were able to pass on the entire German order of battle for their attack on Kursk in July 1943 to the Soviets, without revealing the advances they had made in cryptanalysis²⁵. This huge battle, the largest clash of armour in history, was won by the Soviet Union, and marked the beginning of the unstoppable Red Army advance on Berlin. The ability to crack Tunny ciphers was also vital in the D-Day landings. Among other things, it showed to the Allies that the massive plan to deceive the Axis about the location of the eventual landings, Operation Fortitude, had been successful. Tunny intelligence confirmed that Hitler himself was convinced, against some of his generals better judgement, that the landings in Normandy on 6th June were just a diversion.

4 Conclusion

Cryptanalysis is just one area where mathematics is relevant to linguistics. It clearly constitutes a case where a mathematical approach may be combined with a linguistic one to produce results which neither field by itself would produce.

²⁴The existence of Colossus, and the work of its creators, Newman and Flowers was kept a tight secret until the 1970s. In 1980, Flowers, who had quietly returned to work for the GPO after the war, received some long overdue recognition: he was the first winner of the Martlesham Medal for his achievements in computing. In 1993, he received a simple certificate in basic information processing on a personal computer from Hendon College (a vocational school for adults): as it happened, the only formal qualification in computing that the engineer responsible for the first programmable electronic computer was ever to obtain.

²⁵General Zhukov, the Russian commander, was convinced the British had a mole at the German army headquarters (further fuelling Soviet suspicions that the western Allies might reach a separate peace deal with the Nazis), so impressed was he with the information that he had received. Part of the reason for the secrecy over the work in deciphering Tunny and the existence of Colossus may have been that the British suspected that in a post war world, the Soviets (who Churchill increasingly distrusted) might start using Lorenz, or something based on it, themselves.

durre risultati che nessuno dei due campi isolati potrebbe produrre separatamente.

Oggi, con l'avvento della calcolo elettronico, annunciato da Colossus, e dell'Intelligenza Artificiale, è probabile che le applicazioni della matematica alla linguistica crescano esponenzialmente in molte direzioni differenti. La traduzione assistita da *Computer* è un'area dove sono avvenuti grandi progressi negli anni recenti; ci sono applicazioni gratuite *on-line*, come Google Translate, capaci di fornire passabili traduzioni da e verso l'Inglese per molte lingue²⁷. Questi programmi possono offrire la prospettiva di non dover imparare una lingua straniera, dato che sarà possibile usare i *computer* per parlare nella propria lingua ed essere istantaneamente capiti da chi parla qualsiasi altra lingua, analogamente a come Lorenz SZ40/42 permetteva di trasformare un testo in un codice cifrato e quindi ancora nel testo originale. Ostler [18] sostiene che questa tecnologia eliminerà la necessità di una lingua franca globale, un ruolo attualmente goduto dall'Inglese:

"È possibile guardare avanti nel mondo dei *media* elettronici interlinguistici, un mondo che si migliora e arricchisce dinamicamente. Come la rivoluzione della stampa, e altre varie rivoluzioni sociali associate all'urbanizzazione, hanno cambiato le regole di base della comunicazione tra gli europei del sedicesimo secolo, allo stesso modo, la tecnologia elettronica, se continua a seguire l'andamento attuale, è pronta per cambiare l'antica necessità di una lingua franca per tutti coloro che desiderano partecipare direttamente alle principali conversazioni internazionali. In breve, se l'elettronica può rimuovere la necessità di un intermediario umano per interpretare, o tradurre, le frustrazioni delle barriere linguistiche possono essere superate senza un mezzo universalmente condiviso al di là di un *software* compatibile. Discorsi registrati e

²⁷Per l'accoppiamento di altre lingue, queste app potrebbero essere meno affidabili dato il minor numero di utenti e di dati di *input* con cui lavorare ed imparare.

Now with the advent of computing, as heralded by Colossus, and of Artificial Intelligence, the applications of mathematics to linguistics are likely to rise exponentially and in many different directions. Computer-Assisted Translation is an area where great progress has been made in recent years; even free online apps such as Google Translate are able to provide passable translations to and from many languages into English²⁶. Such programs, may offer the prospect of no one having to learn foreign or second languages any more, as it will be possible to use computers to speak in one's own language and be instantly understood by speakers of any other language, analogously to the way that the Lorenz SZ40/42 allowed plaintext to be transformed into ciphertext and then back again. Ostler [18] argues that such technology will do away with the need for a global lingua franca, a role currently enjoyed by English:

"It is possible to look ahead into the dynamically improving, and enriching, world of interlingual electronic media. Just as the print revolution - and various other social revolutions associated with urbanization - changed the ground rules of communication among Europeans in the sixteenth century, so modern electronic technology, if it follows its current path, is set to change the ancient need for a single lingua-franca for all who wish to participate directly in the main international conversation. In brief, if electronics can remove the requirement for a human intermediary to interpret or translate, the frustrations of the language barrier may be overcome without any universal shared medium beyond compatible software. Recorded speeches and printed texts will become virtual media, accessible through whatever language the listener or speaker prefers.

²⁶With pairings of other languages, they may be less reliable due to the smaller numbers of users and the less input they have to work with and learn from.

testi scritti diventeranno *media* virtuali, accessibili in ogni linguaggio che l'ascoltatore o il relatore preferisce.

Alla fine, e forse abbastanza presto, diciamo a metà del XXI secolo, ognuno potrà esprimere un'opinione nella propria lingua, sia oralmente o scrivendo, e il mondo comprenderà."

Ci si può chiedere se, dato che il linguaggio non è solo un mezzo per esprimere idee, ma anche una parte fondamentale della nostra identità e personalità, non sarà, nelle parole di Christiansen ([19], p.150),

"dubbioso che gli esseri umani vorranno delegare la comunicazione con persone di altra lingua ad apparati digitali così come hanno entusiasticamente rinunciato all'aritmetica mentale per quella dei calcolatori elettronici. "

Tuttavia, il fatto che i traduttori automatici oggi siano così economici e facili da usare avrà certamente un effetto sul numero di persone che desiderano investire tempo, denaro e sforzo nell'imparare una seconda lingua straniera, che sarà una grande perdita per quanto riguarda i ben noti benefici psicologici e cognitivi di un individuo multilingua [20].

Un'altra area, con implicazioni indubbiamente molto a più ampio raggio, è il programma di facilitazione linguistica tra umani e *computer / robot*, cioè i campi in rapida crescita del processamento del linguaggio naturale su cui lavorano molti informatici, ingegneri e matematici²⁸. Il linguaggio parlato spontaneamente, con tutte le sue imperfezioni, false partenze e cambi di direzione, si è dimostrato difficile da processare per i *computer*, ma recentemente ci sono stati progressi enormi. Questi programmi funzionano perché le lingue possono essere analizzate e si possono scrivere algoritmi per permettere ai *computer* di

Ultimately, and perhaps before too long - say by the middle of the twenty-first century - everyone will be able to express an opinion in his or her own language, whether in speech or in writing, and the world will understand."

One may wonder whether, given that language is not only a means by which we express our ideas, but also a key part of our identity and our personality, it will not, in the words of Christiansen ([19], p.150), be

"doubtful that human beings will want to delegate communication with people of other languages to digital devices as enthusiastically as they gave up mental arithmetic for electronic calculators."

However, the fact that useable translations are now so cheap and easy to come by will no doubt have an effect on the numbers of people willing to invest the time, money and effort into learning foreign or second languages, which would be a great loss given the well-attested cognitive and psychological benefits to an individual of being plurilingua [20].

Another area, with doubtless more far-reaching implications, is programs facilitating the linguistic or spoken interaction between humans and computers / robots, i.e. the rapidly expanding fields of natural language processing that many computer scientists, engineers, and mathematicians are working on²⁷. Natural spontaneous spoken language, with all its imperfections and false starts and changes of direction, has proved difficult for computers to process, but recent years have seen enormous advances. These programs function because language can be analysed and algorithms written to allow computers to use, or

²⁸Questa è un'area ovviamente popolata dai linguisti matematici, un ramo che applica specifici metodi e concetti matematici a sistemi linguistici, una tradizione, osserva qualcuno, che risale ad Euclide (circa 325 - 265 a.c.) e allo studioso sanscrito Pāṇini, (circa 520-460 a.c.) [21].

²⁷This is an area covered most obviously by mathematical linguistics, a branch which specifically applies mathematical methods and concepts to linguistic systems, a tradition, some observe, going back to Euclid (c. 325-265 BCE) and Pāṇini, the Sanskrit scholar, (c. 520-460 BCE) [21].

usare il linguaggio, o di simularne l'uso²⁹. La prospettiva che gli essere umani possano usare il loro linguaggio naturale come interfaccia con *computer* ed intelligenze artificiali (o tra *computer* e intelligenze artificiali senza alcuna intermediazione umana) è qualcosa che la fantascienza ha ipotizzato già da molto tempo. È un'ipotesi stuzzicante con cose come Siri di Apple o Alexa di Amazon, sebbene sia difficile prevedere quando diventerà realtà, ed esseri umani e *computer* potranno avere lunghe, e significative, conversazioni come avviene, ad esempio, nei film come 2001: Odissea nello spazio di Kubrick, con HAL 9000 l'inizialmente geniale *computer* di controllo della nave spaziale Discovery One. Per ottenere questo livello, come nel caso della decifrazione del cifrario Lorenz (Tunny), linguisti, matematici, ingegneri, ed informatici dovranno lavorare insieme, poiché, separatamente nessuno di loro sarebbe capace di superare la miriade di sfide da affrontare.

Il fatto che matematici, linguisti, e studiosi di altri campi possano collaborare in maniera proficua non è una sorpresa. È semplicemente il risultato di una profonda connessione tra linguaggio, matematica e altre scienze; ognuna è una manifestazione di diversi ambiti cognitivi la cui funzione è quella di aiutare a comprendere aspetti differenti del mondo, e che forniscono diversi strumenti concettuali e affrontano lo stesso problema da prospettive diverse.

²⁹Così è come Alan Turing presentò il *test* chiamato con il suo nome: non come se una macchina potesse usare il linguaggio come un essere umano (forse impossibile da verificare visto che il lavoro della mente e la locazione del linguaggio sono ancora coperti dal mistero) ma piuttosto se un essere umano sarebbe capace di distinguere se sta parlando con un *computer* o con un altro essere umano.

at least appear to use language²⁸. The prospect of humans being able to use their natural language as an interface with computers and artificially intelligent devices (or indeed between computers and AI devices without any human agency) is something that science fiction has long assumed would one day be possible. Such a thing seems tantalisingly close with things like Apple's Siri or Amazon's Alexa, although it is difficult to predict when it will become reality, and humans and computers will be able to have long, meaningful conversations of the kind seen in films such as Kubrick's 2001: A Space Odyssey (1968) with HAL 9000 the initially genial control computer on the spacecraft Discovery One. To achieve this, as was the case with the cracking of the Lorenz cipher (Tunny), linguists, mathematicians, engineers, and computer scientists will have to work together, each looking at different aspects of the problem, because, separately, none of them will be able to overcome the myriad challenges.

The fact that mathematicians, linguists, and scholars from other fields can profitably work together should come as no surprise. It is merely a result of the underlying connection that exists between language, mathematics and the other sciences; they are each manifestations of different cognitive frameworks whose function is to help us understand different aspects of the world, and which give one different conceptual tools and thus approach the same problem from different perspectives.

²⁸This, in essence, is how Alan Turing presented the problem when he thought up the test named after him: not as such whether a machine could actually use language in a humanlike way (something perhaps impossible to test seeing that the inner workings of the mind and language's place within it are still shrouded in mystery) but rather whether a human would be able to tell whether they were conversing with another human or not.



- [1] N. Chomsky: *Cartesian linguistics: a chapter in the history of rationalist thought*, Harper & Row, New York (NY) (1966).
- [2] C. R. Darwin: *The descent of man, and selection in relation to sex*, John Murray, London (1871).
- [3] S. Pinker: *The Blank Slate: The Modern Denial of Human Nature*, Viking, New York (NY) (2002/2016).
- [4] N. Chomsky: *Lectures on Government and Binding*, Foris, Dordrecht (1981).
- [5] F. De Saussure: *Cours de Linguistique Générale*, Payot, Paris (1916).
- [6] N. Chomsky: *Aspects of the Theory of Syntax*, Massachusetts Institute of Technology Press, Cambridge (MA) (1965).
- [7] N. Chomsky: *Language and Mind*, Cambridge University Press, Cambridge (2006).

- [8] G. H. Hardy: *A Mathematician's Apology*, Cambridge University Press, Cambridge (1940).
- [9] I. A. Al-Kadi: *The origins of cryptology: The Arab contributions*, Cryptologia, 16 (1992) 97.
- [10] F. C. Barlett: *Remembering*, Cambridge University Press, Cambridge (1932).
- [11] S. Pinker: *The Language Instinct*, Penguin, Harmondsworth (1994).
- [12] M. Drosnin: *The Bible Code*, Simon & Schuster, New York (NY) (1997).
- [13] B. McKay, D. Bar-Natan, M. Bar-Hillel, G. Kalai: *Solving the Bible Code Puzzle*, Statistical Science, 14 (1999) 150.
- [14] J. Ferris: *Behind the Enigma: The Authorized History of GCHQ, Britain's Secret Cyber-Intelligence Agency*, Bloomsbury Publishing, London (2020).
- [15] T. Dunlop: *The Bletchley Girls*, Hodder & Stoughton, London (2015).
- [16] M. Smith: *The Debs of Bletchley Park*, Aurum, London (2015).
- [17] W. T. Tutte: *Appendix 4: My Work at Bletchley Park*, in B. J. Copeland (ed.) *Colossus: The Secrets of Bletchley Park's Codebreaking Computers* Oxford University Press, Oxford, (2006), pp. 352-369.
- [18] N. Ostler: *The Last Lingua Franca: English until the Return of Babel*, Penguin, London (2010).
- [19] T. Christiansen: *The rise of English as the global lingua franca. Is the world heading towards greater monolingualism or new forms of plurilingualism?*, Lingue e Linguaggi, 15 (2015) 129.
- [20] D. Mehmedbegovic, T. H. Bak: *Towards an interdisciplinary lifetime approach to multilingualism*, European Journal of Language Policy, 9 (2017) 150.
- [21] A. Kornai: *Mathematical Linguistics.*, Springer, Berlin (2008).



Thomas Christiansen: Thomas Christiansen è professore associato di Lingua e Traduzione, per la Lingua Inglese presso l'Università del Salento e, dal 2016, Direttore del Centro Linguistico di Ateneo. Ha insegnato in varie università della Puglia, del Regno Unito e della Polonia. Ha completato il dottorato di ricerca in linguistica testuale a Salford (UK). Ha svolto ricerche in varie aree della linguistica e ha pubblicato monografie e articoli su molti campi, tra cui linguistica sistemica e grammatica funzionale, varietà di inglese, inglese come lingua franca, didattica della lingua inglese, valutazione delle competenze linguistiche, e analisi di diversi corpora, incluso il discorso parlato. Ha anche lavorato per molti anni come consulente esperto per Cambridge Assessment English.

Thomas Christiansen: Thomas Christiansen is associate professor in English Language and Translation at the Università del Salento (Lecce, Italy) and since 2016, Director of the University Language Centre. He has taught in various positions at various universities in Apulia (Italy), the UK, and Poland. He completed his PhD in textual linguistics at Salford (UK). He has researched into various areas of linguistics and published books and articles on many fields including systemic linguistics and functional grammar, varieties of English, ELF, teaching English, language testing, and analysis of different corpora, including spoken discourse. He has also worked as an expert consultant for Cambridge Assessment English for many years.

La Fisica e la Matematica negli Studi di Chopin

Massimo Blasone

*Dipartimento di Fisica "E. Caianiello" - Università di Salerno
Istituto Nazionale di Fisica Nucleare, Gruppo Collegato di Salerno.*

Si propone l'utilizzo di concetti fisici per rappresentare l'evoluzione nel tempo delle strutture musicali in termini di processi dinamici. Applicando questo approccio a tre Studi di F.Chopin, l'Op.10 n.1 e n.3 e l'Op.25 n.1, se ne ottengono delle descrizioni originali basate sui concetti di rottura della simmetria, coerenza quantistica e frattali, potenzialmente rilevanti per l'interpretazione pianistica. L'analisi si avvale di spartiti musicali colorati, generati tramite l'implementazione del codice colori/suoni di Scriabin alla notazione musicale.¹

1. Introduzione

Il rapporto tra Musica e Matematica ha una storia antica [2, 3, 4], a partire da Pitagora che per primo associò numeri razionali agli intervalli musicali, fino ai nostri giorni, con l'applicazione di modelli matematici e algoritmi nella produzione di musica, come per esempio nella musica stocastica [5] e frattale [6].

Tuttavia, se si guarda alla pratica musicale, questi sviluppi sembrano avere poco impatto: in particolare per la Musica Classica, esiste una consolidata tradizione nella esecuzione musicale,

¹Il presente lavoro è basato in gran parte su Ref.[1]. Il materiale di Sez.5 è stato presentato nel corso di un seminario/concerto a Schwarzenberg (Austria) ad Agosto 2019, nell'ambito della conferenza Symmetries in Science.

che di fatto non si avvantaggia delle conoscenze sulla struttura matematica della musica. In questo contesto, la posizione del musicista è in qualche modo simile a quella di un alchimista, che ben conosce la procedura per combinare certi elementi allo scopo di ottenere una determinata sostanza, e non si interessa della complessa dinamica di atomi e molecole che è alla base delle reazioni che utilizza.

Naturalmente, c'è anche un'altra ragione, più profonda, per questo atteggiamento: facciamo musica per il nostro piacere, essa rappresenta un veicolo efficiente e universale per comunicare le nostre emozioni. Quindi sicuramente non basta studiare la struttura matematica nascosta nel linguaggio musicale, senza considerare il livello della percezione, che è quello in cui queste strutture vengono elaborate e apprezzate. In altre parole, non necessariamente a una struttura matematica bella in senso formale è associata una musica bella (nel senso artistico), e vice-versa.

Così si potrebbe essere tentati di liquidare questi studi come interessanti ma in realtà non capaci di toccare il livello della esperienza estetica, che è quello rilevante per l'ascolto e l'esecuzione della musica. E però, è sicuramente un'esperienza comune a chiunque suoni una pagina di Bach o Mozart, il profondo senso di meraviglia al confronto di certi capolavori e la sensazione che debba esserci qualcosa dietro a quello che percepiamo come Bellezza. Questa sensazione è forse ancora più forte, se il musicista è anche uno scienziato,

perchè inevitabilmente cercherà di investigare e comprendere i segreti nascosti nell'opera d'arte. È questa la motivazione principale per le considerazioni contenute in questo lavoro: esse non sono il risultato dell'applicazione di un qualche programma teorico a un determinato pezzo musicale; piuttosto nascono quando, seduti al pianoforte, si cerca di capire la logica profonda di una composizione, allo scopo di eseguirla correttamente, e cioè trasferire efficientemente il suo contenuto informativo a se stessi e agli altri.

Un modo tipico di descrivere la musica è attraverso l'uso di metafore, nella forma di immagini o situazioni emotive, che possano aiutare a dare un senso al linguaggio musicale, per sua natura sostanzialmente privo di un livello semantico [8]. Questo è particolarmente evidente nel caso della musica a programma, che per definizione si riferisce a un soggetto specifico. In ogni caso, queste associazioni sono necessariamente vaghe dal momento che sono legate alla musica attraverso gli effetti che entrambe (musica e immagini) producono nei nostri cervelli, effetti che ovviamente sono altamente soggettivi. L'idea è quindi quella di usare come metafore fisiche² di una data musica, specifici processi dinamici, che a loro volta possano essere associati a una molteplicità di processi fisici (immagini). Il vantaggio di questo approccio sta nel fatto che il processo dinamico può essere direttamente collegato alla struttura matematica/tonale della musica: si possono così applicare alla musica concetti come quelli di simmetria, forze, ordine e disordine, transizioni di fase, coerenza ecc., che hanno un significato ben preciso in ambito fisico.

Questa operazione va condotta con molta cura e con un certo grado di approssimazione, mantenendo un approccio qualitativo, allo scopo di evitare di cadere nel riduzionismo, di pretendere cioè che la musica possa essere descritta completamente in termini matematici. Un altro rischio, questo forse inevitabile, è quello di manipolare la musica, forzandone una interpretazione non completamente appropriata. D'altra parte, è anche plausibile pensare che in un'opera d'arte possa esserci più di quello che l'autore stesso abbia potuto immaginare. Una situazione

²L'importanza della metafora come elemento strutturale del pensiero e strumento fondamentale di comprensione è discussa in Ref.[9].

simile non è infrequente in Fisica, basti pensare al caso di Dirac, che definiva la sua equazione "più intelligente dell'autore"! [10]

L'approccio sopra descritto può essere applicato, in linea di principio, a qualsiasi tipo di musica. Il fatto che qui consideriamo in particolare gli Studi di Chopin, ha varie ragioni: una è di carattere personale e ha a che fare con la preferenza per questo Autore e in particolare per i suoi Studi e i Preludi. Ci sono poi motivazioni più oggettive: la forma relativamente semplice di queste composizioni (miniature, nel caso di alcuni Preludi), poco strutturate rispetto ad altre, come ad esempio la Sonata, rende più semplice la loro analisi; inoltre, il carattere di questi pezzi è particolarmente evocativo di immagini (specialmente naturali), come testimoniato da vari autori/editori (vedi Sezione 2). Un altro aspetto notevole della musica di Chopin è costituito dalla sua universalità, essendo apprezzata da un grandissimo numero di persone, indipendentemente dal loro *background* culturale. Infine, un certo carattere enigmatico di questo Autore, come già riconosciuto dai suoi contemporanei (Le Pianiste, 1833), insieme con il fatto che rifiutò sempre ogni interpretazione descrittiva per la sua musica, ne rende particolarmente interessante l'analisi.

L'articolo è organizzato nel modo seguente: in Sezione 2, diamo alcune note storiche sugli Studi di Chopin; le Sezioni 3, 4 e 5 sono dedicate all'analisi degli studi Op.10 n.3, Op.25 n.1 e Op.10 n.1 rispettivamente; la Sezione 6 contiene le conclusioni e le prospettive di questo lavoro. In Appendice (Sez.7), è illustrata l'implementazione grafica del codice tonale di colori di Scriabin, che è utilizzata per l'analisi degli Studi.

2. Una nota storica sugli Studi di Chopin

Il termine *Studio* denota una composizione strumentale con fini didattici, intesa a esercitare una particolare combinazione di note o un movimento del corpo in relazione allo strumento musicale in questione. Queste composizioni hanno tipicamente una struttura piuttosto semplice rispetto a quella di altre forme musicali.

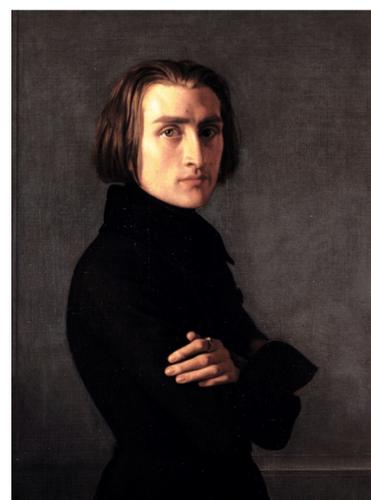
Tra i primi esempi di studi per strumenti a tastiera, ci sono gli "Esercizi per gravicembalo" (1738) di Domenico Scarlatti, e poi le raccolte



N. Paganini (1782-1840)



F. Chopin (1810-1849)



F. Liszt (1811-1886)

Figura 1: Chopin tra Paganini e Liszt

di Studi di J.B.Cramer, M.Clementi, C.Czerny e I.Moscheles, per citare solo i principali.

Fryderyk Chopin compose due raccolte di Studi: 12 Studi Op.10, pubblicati nel 1833 e 12 Studi Op.25, pubblicati nel 1837; successivamente pubblicò tre Studi per il metodo Moscheles. Inoltre, alcuni dei 24 Preludi Op.28 (pubblicati nel 1839) hanno caratteristiche comuni con le Op.10 e 25. Gli Studi di Chopin sono dedicati allo sviluppo di aspetti tecnici specifici (Op.10 n.2 per le scale cromatiche, Op.25 n.10 per le ottave, Op.25 n.6 per le doppie terze, ecc.) spinti ai massimi livelli di difficoltà (almeno per l'epoca).

L'intento virtuosistico è confermato dalla dedica dell'Op.10 a F. Liszt (l'Op.25 è dedicata a M. d'Agoult, compagna di Liszt). È anche noto [11] che Chopin scrisse alcuni degli Studi dell'Op.10 come esercizi per se stesso durante lo studio del suo Concerto n.1 per pianoforte e orchestra. Un altro *input* fu rappresentato dai dieci concerti dati da N. Paganini a Varsavia nel 1829, a cui il giovane Chopin ebbe occasione di assistere. Come molti suoi contemporanei (compreso Liszt), fu profondamente impressionato dal virtuosismo e dalla personalità del violinista italiano: i 24 "Capricci per violino" furono un modello per Chopin per i suoi Studi e Preludi (insieme al "Clavicembalo Ben Temperato" di J.S. Bach).

Per Chopin, comunque, l'aspetto tecnico non è mai limitato alla dimensione didattica, piuttosto diventa strumentale per l'espressione di una grande varietà di stati d'animo: egli fu in grado

di elevare gli studi dallo stato di esercizi a quello di opere d'arte, creando un nuovo genere – lo studio da concerto – che fu successivamente sviluppato da Liszt, Debussy, Rachmaninov e altri. Questo risultato è ben sintetizzato da L. Kentner, per il quale gli Studi di Chopin rappresentano una "fusione perfetta di atletica ed estetica" [12]: in essi la tecnica è intesa nel senso etimologico di *τέχνη*, ossia di Arte nel suo significato più ampio. Queste composizioni, che ancora oggi rappresentano una summa del virtuosismo pianistico, presentarono tali novità da risultare rivoluzionarie a loro tempo: posizioni molto larghe per la mano, nuovi modi di attacco della tastiera, insieme con nuove combinazioni ed estese progressioni armoniche, figure strumentali usate come temi (carattere cantabile della figurazione). Per alcuni aspetti, gli originali trattamenti armonici contenuti negli Studi di Chopin anticipano la dissoluzione della tonalità [13].

Per la loro originalità, queste caratteristiche non furono sempre comprese dai contemporanei e costarono alcune critiche all'autore, come quella dell'editore e compositore L.Rellstab che definì gli Studi Op.10 come "ottimi esercizi per chi ha le dita storte e vuole raddrizzarle, consigliando coloro che hanno dita ben diritte di non suonarli se non hanno sottomano un chirurgo" [12]. Anche F.Mirecki criticò la monotonia (!) dei temi negli Studi, dicendo che "se le note fossero colorate, anziché nere, gli spartiti potrebbero essere usati come carta da parati" [12].

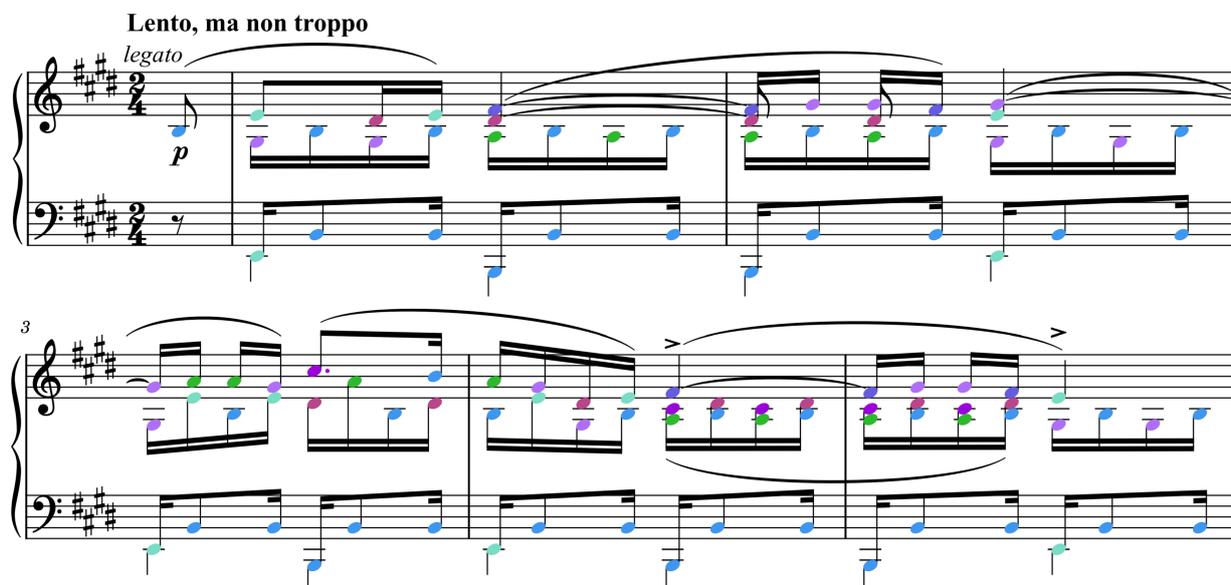


Figura 2: Op.10 n.3 – Tema iniziale.

D'altra parte, altri riconobbero immediatamente il grande valore della musica di Chopin: gli Studi Op.10 furono recensiti molto positivamente sulla rivista "Le Pianiste", subito dopo la loro pubblicazione. Anche Robert Schumann, in veste di critico musicale, apprezzò molto il giovane compositore ("Giù il cappello, signori, un genio!" è il famoso incipit della sua recensione delle Variazioni Op.2 sul "Là ci darem la mano" di Mozart) e lo incluse persino come uno dei personaggi nel *Carnaval* (insieme a Paganini). Più di recente [14], J. Huneker ha definito gli Studi di Chopin "esperimenti titanici" e, a proposito dell'Op.25 n.11, ha espresso la seguente considerazione: "Uomini dall'anima piccola, per quanto agili siano le loro dita, non dovrebbero cimentarvisi".

È importante notare come, sebbene Chopin abbia sempre usato titoli generici per le sue composizioni (Studi, Preludi, Notturmi, Ballate, ecc.) rifiutando ogni intento programmatico per la sua musica, tuttavia molte sue composizioni abbiano ripetutamente attratto dei titoli, essendo evidentemente molto evocative di immagini ed emozioni. Alcuni esempi per l'Op.10: n.1 (Cascata), n.3 (Tristezza), n.8 (Raggio di sole), n.12 (La caduta di Varsavia) e per l'Op.25: n.1 (Arpa eolia), n.2 (Le api), n.8 (Farfalla), n.11 (Vento d'inverno), n.12 (Oceano).

È chiaro che queste immagini hanno un carattere estremamente soggettivo e in fondo ab-

bastanza arbitrario, ma è significativo come gli Studi di Chopin appaiano capaci di comunicare in modo molto distinto delle specifiche emozioni. In un certo senso, possono essere visti come un distillato di diverse situazioni emotive, associate a problemi di tecnica pianistica che ne costituiscono la motivazione esteriore.

3. Op.10 n.3 – Cristallo blu con macchie rosse

Lo Studio Op.10 n.3, anche noto come "Tristezza" (vedi sopra) è sicuramente una tra le composizioni più note di Chopin, per il suo carattere lirico. È riportato [12] come l'autore stesso abbia detto a suo proposito: "In tutta la mia vita non sono stato più capace di trovare una melodia così bella". Questa melodia è il tema che appare nella prima e terza sezione dello studio – che ha la struttura A-B-A' – ed è infatti così bello da oscurare la sezione centrale dal carattere virtuosistico. A questo proposito, è interessante leggere la recensione di C. Chaliou, nella rivista "Le Pianiste" (n.1, 1833), su questo studio: "Il terzo è tra i più difficili. La melodia è meravigliosa, ma difficile da rendere. Raccomando la scala cromatica in settime diminuite, che è originale e di buon gusto; non posso dire lo stesso per il passaggio che segue [quello marcato con bravura], che non ha un effetto piacevole...".

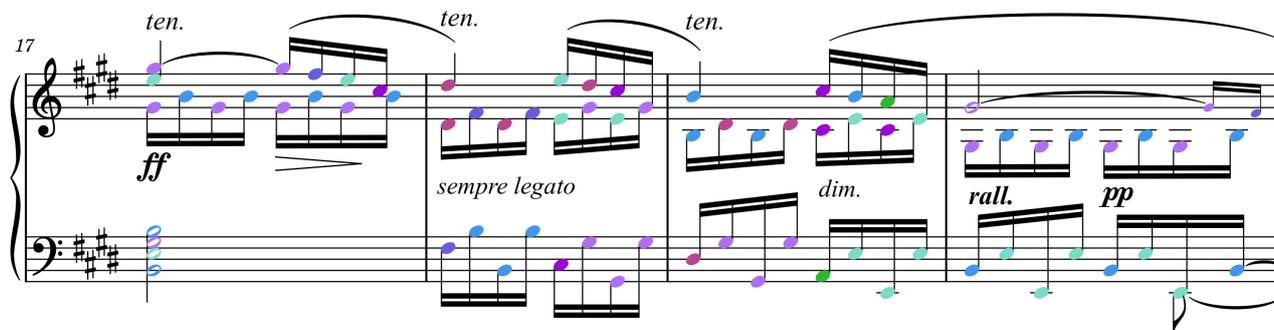


Figura 3: Op.10 n.3, prima sezione – Fase ordinata (cristallo blu).

Si è allora portati a chiedersi la ragione di questa struttura così contrastante, che può apparire sorprendente e quasi sconcertante a prima vista. Di seguito, sulla base di argomenti matematici molto generali (simmetrie), si propone una possibile chiave di lettura per questo Studio, che crediamo possa essere di aiuto anche per la sua interpretazione pianistica.

Partiamo da alcune considerazioni generali circa il sistema tonale usato nella musica occidentale [15]: esso è basato su dodici note ugualmente spaziate (temperamento equabile), ossia una discretizzazione simmetrica dell'intervallo naturale (ottava) corrispondente al raddoppio di una data frequenza – laddove i toni con frequenza raddoppiata sono percepiti come la stessa nota ad una diversa altezza. Di conseguenza, la musica tonale consiste di operazioni su 12 elementi (note) modulo 12 (gruppo ciclico \mathbb{Z}_{12}) [16, 17, 18].

Dal punto di vista della percezione, l'affermazione di una data tonalità corrisponde quindi a un processo di rottura della simmetria verso una tra le *a priori* inizialmente equivalenti 12 tonalità (non distinguiamo qui tra modi maggiori e minori). L'idea di rottura della simmetria è una delle più feconde nella Fisica moderna, essendo associata alla emersione di strutture ordinate dal disordine (simmetria) [19] (vedi Box n.1).

Nel caso specifico dell'Op.10 n.3, la tonalità di Mi maggiore si manifesta in modo molto netto nel tema principale: questo processo può essere visto come uno stato ordinato³ (p.es. un cristallo) che cresce nel tempo di fronte a noi (vedi Fig.2). Usando la corrispondenza tra suoni e co-

³È importante sottolineare come questo stato ordinato consista nella correlazione tra percezioni che avvengono a tempi successivi, per cui è intrinsecamente legato ad un processo di memorizzazione ed elaborazione della musica da parte del cervello.

lori di Scriabin (vedi Appendice), si ottiene un cristallo blu/verde, una immagine di chiarezza e perfezione (Fig.3).

La seconda sezione, marcata "più mosso" ha un carattere molto diverso: ci sono scale cromatiche e sequenze dissonanti di doppie note seguite da un lungo passaggio in doppie seste, marcato "con bravura" per evidenziarne il carattere di cadenza virtuosistica, e basato su accordi di settima diminuita. Questi accordi hanno la caratteristica di essere simmetrici, essendo costituiti ognuno di quattro note, ugualmente spaziate (vedi Fig.6). Nel nostro sistema tonale esistono tre distinti accordi di questo tipo, che quindi coprono tutte e dodici le note.

Il passaggio alle battute 46-53 (vedi Fig.7) è una progressione basata su questi tre accordi, sviluppati su intervalli fissi (seste) su tutte le dodici note: può essere quindi visto come una fase simmetrica (disordinata) – di fatto atonale, anche se formalmente basata sulla dominante Si. Qui lo stato ordinato iniziale, rappresentato dalla tonalità di Mi+, è distrutto: il cristallo blu si è fuso nel fuoco e tutti i colori sono presenti (ripristino della simmetria).

Il ritorno al tema iniziale basato sulla tonalità di Mi+ richiede una rottura della simmetria tonale. Questa è realizzata alla battuta 53 (vedi Fig.7) abbassando il Do# a Do, trasformando l'intervallo di sesta maggiore Mi/Do# in uno di sesta minore Mi-Do, per giungere poi alla dominante Si. A questo punto segue una breve sezione (battute 54-61): qui si verifica una dissipazione della grande energia accumulata precedentemente. Questo effetto è ottenuto tramite gruppi irregolari (terzine) nel basso, che producono una impressione di rallentamento (attrito) (Fig.8).

Finalmente, la tonalità originale di Mi maggio-

Box 1. Rottura spontanea della simmetria e difetti topologici

Il concetto di simmetria è di grande importanza per la fisica moderna e in particolare rappresenta un ingrediente fondamentale del Modello Standard delle interazioni fondamentali.

In Fig.4 è schematizzato il processo di rottura spontanea della simmetria: il sistema è in uno stato disordinato (simmetrico) al di sopra della temperatura critica T_c . Quando la temperatura scende sotto T_c , il sistema si ordina spontaneamente lungo una certa direzione (casuale). Questo tipo di dinamica è alla base della magnetizzazione di un pezzo di ferro che, inizialmente sopra la temperatura di Curie, venga raffreddato in un campo magnetico esterno.

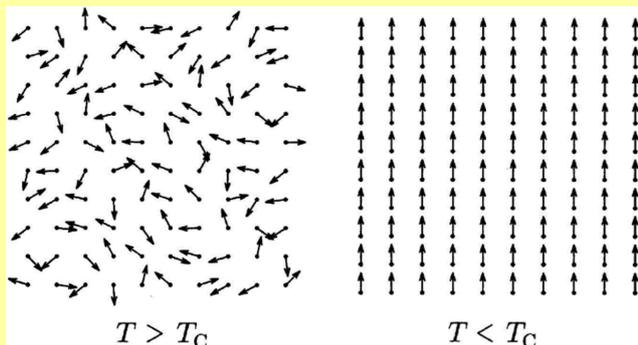


Figura 4: Rottura spontanea della simmetria

Quando il passaggio dalla fase ordinata a quella disordinata avviene molto rapidamente (rispetto ai parametri caratteristici del sistema), può succedere che regioni contenenti la fase disordinata rimangano intrappolate nella nuova fase ordinata. Si hanno così dei difetti topologici (meccanismo di Kibble–Zurek). Questo fenomeno è comune a molti sistemi fisici su scale molto diverse, dalla materia condensata alla cosmologia.

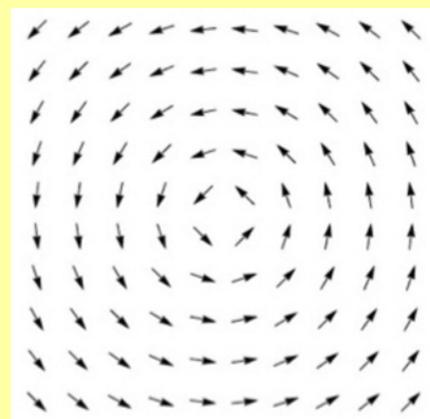


Figura 5: Fase ordinata con vortice

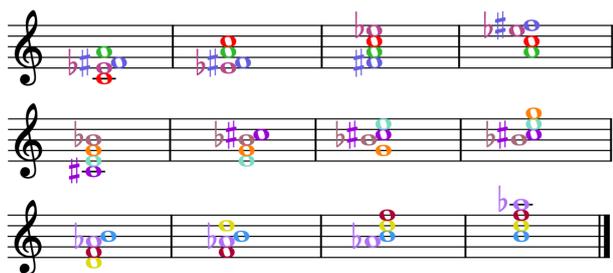


Figura 6: Accordi di settima diminuita con i rispettivi rivolti. Si noti come i tre accordi esauriscano le 12 note del sistema tonale.

re è raggiunta e il tema riappare: il sistema si è raffreddato e il cristallo si è riformato nel crogiolo. Tuttavia, questa sezione non è una copia esatta della prima: alcune battute mancano, le indicazioni dinamiche sono differenti (f invece di ff alla battuta 70), ma in particolare, ora appa-

re un Do \sharp nel tema principale alle battute 73-74. Questo fatto è notevole in quanto questa nota non è nella armonia di Mi \sharp e produce l'effetto di aggiungere una nota nostalgica alla melodia (giustificando forse il titolo Tristezza). Da notare come questa nota sia la stessa che ha prodotto la rottura della simmetria alla battuta 53.

Ritornando all'immagine del cristallo che si fonde e poi ricristallizza, vediamo che questa non è più la struttura perfetta apparsa originariamente. Ora alcuni difetti – macchie rosse – appaiono nella gemma blu (Fig.9). Questa visione è liberamente ispirata al meccanismo di Kibble–Zurek per la generazione di difetti topologici nel corso di transizioni di fase associate a una rottura della simmetria [20] (vedi Box 1): questi difetti appaiono durante il (rapido) processo di ordinamento

Figura 7: Op.10 n.3, seconda sezione – La fase simmetrica (tutti i colori sono presenti). Rottura della simmetria tramite ripetizione del Do \sharp (note rosse alla battuta 53).

del sistema e rappresentano una memoria della fase simmetrica. Nel caso specifico, abbiamo il ripristino e la rottura della simmetria tonale con la comparsa di difetti (armonici).

A questo punto, bisogna ammetterlo, il Lettore potrebbe ben obiettare sul fatto che, dopo aver criticato (vedi Sez.2) l'arbitrarietà dei titoli associati alle composizioni di Chopin, siamo arrivati, dopo una lunga discussione, a dare un nuovo titolo a questo studio (Il cristallo blu con macchie rosse)!

In realtà, quello che stiamo proponendo è di associare alla musica un processo dinamico (metafora fisica), piuttosto che una immagine specifica. Infatti, allo stesso modo di come il meccanismo di Kibble-Zurek descrive la formazione di difetti in un gran numero di sistemi fisici molto differenti tra loro (dalla cosmologia ai cristalli nematici o ai superconduttori), qui l'aspetto fondamentale è rappresentato dalla dinamica collegata alla

evoluzione dei *pattern* di simmetria, mentre il cristallo che si scioglie e si riforma con i difetti è solo un utile (e certamente non unico) strumento per visualizzare questo processo. È importante sottolineare come il reale processo dinamico associato alla musica abbia luogo a livello della percezione (cioè nel nostro cervello): l'Artista, grazie a una sensibilità fuori dal comune, è capace di dare una forma musicale alle sue emozioni, in un modo che può esser comunicato ad altri soggetti, attraverso una risonanza di queste forme nei loro cervelli.

Ritornando al processo fisico sopra descritto, possiamo ora provare a proporre una interpretazione per lo Studio Op.10 n.3: una metafora di un viaggio interiore da una età adolescenziale di sogni illimitati e perfetti – passando per alcuni eventi drammatici – fino a uno stadio successivo, in cui le stesse emozioni e sogni si manifestano necessariamente in modo differente, esibendo le cicatrici di quegli eventi traumatici. In questo

Figura 8: Op.10 n.3, seconda sezione – Episodio di transizione (dissipazione).

Figura 9: Op.10 n.3 – terza sezione – Fase ordinata con difetti (cristallo blu con macchie rosse).

schema, la sezione centrale dello studio, che sembrò così disturbante ai contemporanei di Chopin, gioca un ruolo cruciale, celato sotto un'apparente motivazione tecnica.

Come passo finale nel nostro ragionamento, possiamo chiederci in che modo la struttura matematica associata a specifici processi fisici possa avere a che fare con le nostre emozioni (intese come stati della dinamica cerebrale). Una possibile risposta è che le forme musicali sono capaci di eccitare strutture ordinate già presenti nel nostro cervello (esperienza estetica), che hanno una loro dinamica caratteristica. L'ipotesi più semplice è quella che i processi fisici associati alla musica ab-

biano una corrispondenza diretta con quelli che avvengono nel cervello: in questo senso la musica sarebbe niente altro che una rappresentazione della fisica delle emozioni.

Allegro sostenuto

Figura 10: Op.25 n.1, battute iniziali. Stati coerenti – dinamica imperturbata.

Figura 11: Op.25 n.1, battute 32-35. Stati coerenti - dinamica perturbata (squeezing).

4. Op.25 n.1 – Stati coerenti

Questo studio è noto come “Arpa Eolia”, dal racconto lasciatoci da Robert Schumann, che lo ascoltò eseguito dall’autore:

"S'immagini un'arpa eolia che abbia tutte le gamme sonore e che la mano d'un artista le mescoli in ogni sorta d'arabeschi fantastici, in modo però da udire sempre un suono grave fondamentale e una morbida nota alta; s'avrà così press'apoco un'immagine del modo di sonare di Chopin.

...

Sbaglierebbe chi pensasse ch'egli facesse udire chiaramente ognuna delle piccole note; si sentiva piuttosto un'ondulazione dell'accordo in la bemolle maggiore, rinnovato di tempo in tempo dal pedale, ma attraverso le armonie si distinguevano melodie dai suoni ampi,

meravigliosi; una volta sola, a metà del pezzo, si sentiva chiara fra gli accordi una voce di tenore, insieme al canto principale. Finito lo studio, si prova l'impressione di chi si vede sfuggire una beata immagine apparsa in sogno e che, già mezzo sveglio, vorrebbe ancora trattenere."

L'aspetto tecnico considerato in questo Studio è costituito dagli arpeggi, che però sono trattati in modo differente rispetto ad altri Studi sullo stesso soggetto (p.es Op.10 n.1 and n.11, Op.25 n.12). Qui gli arpeggi sono sviluppati in modo compatto attorno ad una nota (superiore o inferiore) e le note interne sono scritte in caratteri piccoli con l'intenzione, come testimoniato da Schumann, di dare a queste (piccole) note solo la funzione di colore – un'anticipazione delle trame impressionistiche come per esempio in Debussy. Questo è ben visibile in Fig.10 dove sono mostrate le prime battute del pezzo.

Box 2. Stati coerenti

In Meccanica Quantistica vale la relazione di incertezza di Heisenberg per due osservabili coniugate A e B , non commutanti tra loro: $\Delta A \Delta B \geq \frac{\hbar}{2}$. Esempi tipici sono le coppie posizione–momento e tempo–energia. Gli stati coerenti sono particolari stati quantistici che minimizzano la relazione di Heisenberg. Nel caso di posizione e momento si ha

$$\Delta x \Delta p = \frac{\hbar}{2} \quad (1)$$

Gli stati **coerenti** sono gli stati quantistici che approssimano meglio i corrispondenti stati classici.

Gli stati **squeezed** (coerenti) sono stati per cui vale ancora la relazione (1) ma con $\Delta x \neq \Delta p$, in pratica l'incertezza in una variabile viene spremuta, con corrispondente aumento della incertezza nella variabile coniugata, in modo che l'equazione (1) continui a valere.

Esempi di stati **coerenti** e **squeezed**: In Fig.12 è riportata la funzione di Wigner per uno stato coerente e in Fig.13 quella per uno stato squeezed coerente.

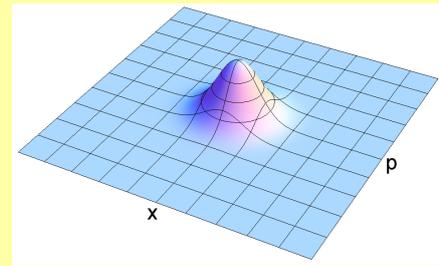


Figura 12: Stato coerente

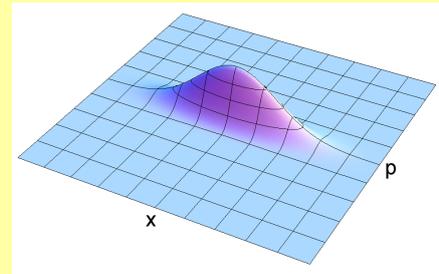


Figura 13: Stato squeezed

Il peculiare effetto sonoro di questa scrittura è quello di gocce con una struttura interna indefinita. Per un fisico è naturale pensare a questi oggetti sonori in termini di stati coerenti (vedi Box 2), con una forma (macroscopica) ben definita e un contenuto (microscopico) di particelle non definito e fluttuante, come dettato dalla relazione di incertezza fase–numero:

$$\Delta n \Delta \phi \geq \frac{1}{2} \quad (2)$$

È importante nuovamente sottolineare che una tale immagine suggestiva debba essere applicata a livello della percezione piuttosto che a quello relativo allo spartito musicale. Il significato della Eq.(2) è pertanto il seguente: la percezione dettagliata della struttura interna (note piccole) dell'arpeggio non è compatibile con quella dell'insieme.

Lo studio inizialmente procede in una maniera piuttosto uniforme, in una atmosfera estatica, come riportato nel racconto di Schumann: abbiamo quindi stati coerenti che evolvono secondo una dinamica libera e non perturbata, nella prima parte del pezzo (Fig.10).

Successivamente, però, una serie di modula-

zioni produce una dinamica perturbata, generata dai gradienti nell'armonia (potenziale) con conseguente deformazione degli stati coerenti (squeezing): questo è visibile in Fig.11 nella alterata struttura ritmica degli arpeggi alla mano sinistra (pentagramma inferiore).

L'idea della linea di basso che agisce come una specie di potenziale è ulteriormente supportata dall'analisi del passaggio alle battute 22-25 (Fig.14): qui la voce superiore si muove per due volte su e giù partendo dal Mi, e solo la terza volta è capace di raggiungere il Fa \sharp da dove poi procede. Questo movimento è accompagnato da un cambiamento progressivo nell'armonia, che solo alla battuta 24 fornisce sufficiente energia per allontanarsi dall'attrattore Mi. Questa dinamica è simile al comportamento di una particella in una buca di potenziale: se mossa dalla posizione di equilibrio, inevitabilmente tornerà indietro a meno che non venga fornita sufficiente energia a scavalcare la barriera. In qualche modo, possiamo quindi considerare questo passaggio come una sorta di trappola armonica.

Verso la fine, la dinamica si rilassa nuovamente e, come in un sogno, il quadro tende a svanire.

Figura 14: Op.25 n.1, battute 22-24. “Trappola armonica”: la voce superiore non può raggiungere il Fa# finchè una energia sufficiente non è generata dai gradienti armonici.

Figura 15: Op.25 n.1, battute finali. Stati coerenti – Decoerenza ed evaporazione.

È molto interessante osservare quello che accade nelle battute finali 44-49 (vedi Fig.15): il Lab basso alla battuta 44 “apre” gli arpeggi finora chiusi, in un arpeggio che si estende a tutta la tastiera e infine dissipa l’energia residua nel trillo finale. Questo processo può essere associato alla decoerenza ed evaporazione degli stati coerenti.

5. Op.10 n.1 – Spirali frattali

È uno tra i più celebri e più ostici di tutti gli Studi di Chopin. La principale difficoltà nasce dal continuo movimento di estensione e contrazione della mano destra nell’eseguire i rapidi arpeggi

che coprono tutta la tastiera. Lo stesso Chopin diceva ad una sua allieva: “Questo Studio vi farà del bene. Se lo studiate come io l’intendo, vi allargherà la mano e vi darà una gamma di accordi, come dei colpi d’arco. Ma spesso, ahimè, invece d’insegnare tutto ciò, lo fa disimparare”.

È stata notata una profonda analogia strutturale con il Preludio n.1 del “Clavicembalo ben Temperato” di J.S. Bach a cui anche idealmente si rifà come primo pezzo della serie: l’Op.10 n.1 può essere visto come una sorta di manifesto del pianismo romantico.

L’effetto dirompente della rapida sequenza di note degli arpeggi richiama alla mente il movimento di un fluido, da cui il nome “la cascata” (vedi Sez.2). Huneker [14] associa gli arpeggi di

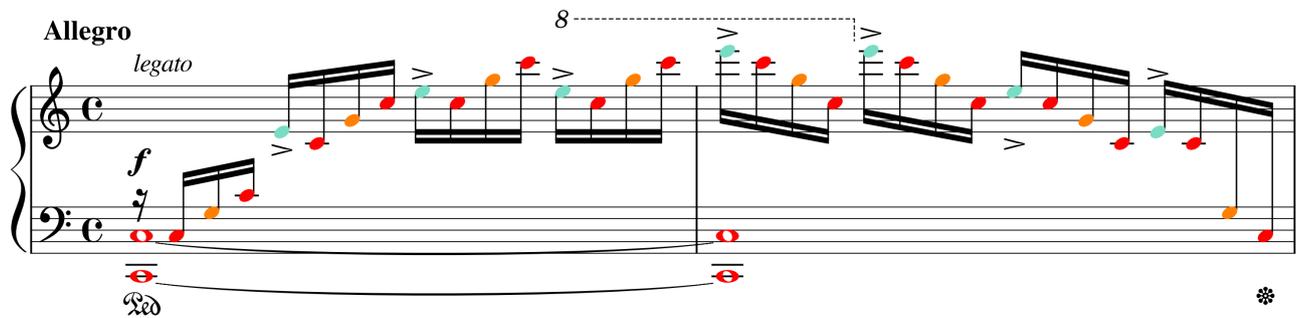


Figura 16: Op.10 n.1 – battute iniziali.

Box 3. I frattali

I **frattali** sono particolari strutture geometriche ottenute tramite (infinite) iterazioni di operazioni elementari a partire da un elemento iniziale. Il nome deriva dal latino *fractus* ed è riferito alla dimensionalità non intera di queste figure.

La caratteristica più evidente dei frattali è costituita dalla auto-similarità: si ritrova la stessa struttura a qualunque scala l'oggetto venga osservato.

Strutture di questo tipo sono ubiquie in natura: alberi, foglie, nuvole, montagne mostrano tutti caratteristiche frattali. In fisica, gli attrattori di alcuni sistemi dinamici esibiscono una struttura frattale e sono perciò detti attrattori strani.

In Fig.17 sono riportate le prime quattro iterazioni del frattale noto come Koch Snowflake. In verde il segmento che, ad ogni passo, viene rimosso e rimpiazzato da due segmenti contigui.

Il frattale che, con un pò di immaginazione, può essere associato allo Studio op.10 n.1 è una spirale frattale (Fig.18) dove ogni spirale piccola rappresenta l'arpeggio elementare e quella grande l'arpeggio completo (vedi testo).

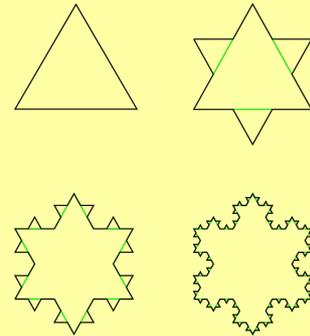


Figura 17: Fiocco di neve

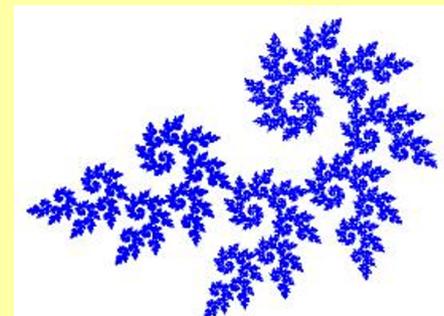


Figura 18: Spirale frattale

questo pezzo alle fantasmagoriche scale dipinte da Giambattista Piranesi, sottolineando l'effetto ipnotico che esercitano sull'ascoltatore.

Osservando la struttura degli arpeggi (Fig.16), si può notare come i vari intervalli costituenti l'arpeggio (Do-Sol, Do-Mi, Sol-Do, etc.) appaiano più volte a diverse scale lungo le 16 note che costituiscono l'arpeggio ascendente⁴, e analogamente per l'arpeggio discendente. È quindi presente una struttura self-similare, analoga a

⁴Per esempio, l'intervallo Do-Mi appare 20 volte nell'arpeggio ascendente.

quella che si ritrova nei frattali (vedi Box 3)⁵. È interessante notare come il rumore di una cascata (composta da gocce di acqua di tutte le dimensioni) sia un classico esempio di rumore frattale.

Se l'effetto della parte ascendente dell'arpeggio è quello di una caduta, il ritorno viene avver-

⁵Anche se la struttura ha dimensione limitata (16 note), l'effetto di un continuo è generato a livello della percezione, in modo analogo a come "vediamo" infinitamente dettagliate le immagini dei frattali in Figg.17 e 18.

Figura 19: Op.10 n.1, battute 69-72 – Struttura multifrattale

tito come una risalita. L'analogia che può essere usata in questo caso è quella di una palla che, lasciata cadere da una certa altezza, accelera fino a rimbalzare sul pavimento, per poi tornare al punto di partenza decelerando.

Seguendo questa analogia, si può proporre una esecuzione che preveda una leggera accelerazione nella parte ascendente dell'arpeggio di 16 note, per poi decelerare nella parte discendente, in modo che il tempo complessivo dell'intero arpeggio sia uguale a quello della esecuzione a velocità costante. Questo effetto ben si addice al fraseggio della melodia, che in questo studio è costituita dal basso in ottave.

Nella parte finale dello studio, gli arpeggi acquistano ulteriore struttura differenziandosi al loro interno (Fig.19). Frattali con più scale sono detti multifrattali.

6. Conclusioni

In questo lavoro abbiamo esplorato la possibilità di associare specifici processi dinamici alla musica allo scopo di ottenere immagini (metafore fisiche) utili per la sua comprensione ed interpretazione. Abbiamo applicato questa idea ad alcuni Studi di F. Chopin: nell'Op.10 n.3, abbiamo trovato un processo di rottura e ripristino della simmetria tonale con associata generazione di difetti armonici, nello spirito del meccanismo di

Kibble-Zurek. Nell'Op.25 n.1, la dinamica emergente è quella di stati coerenti, che evolvono nel tempo e infine sono soggetti a decoerenza. Nel caso dell'Op.10 n.1, infine, abbiamo evidenziato una struttura frattale nella distribuzione degli intervalli all'interno degli arpeggi.

Nello schema descritto, la dinamica è generata dalle relazioni armoniche (forze) tra diversi accordi nel sistema tonale (un'analisi di queste relazioni è data in Ref.[21]). Ci siamo avvalsi di spartiti musicali colorati, ottenuti attraverso l'implementazione del codice di colori di Scriabin per i suoni, come ausilio per la visualizzazione di tali relazioni. La corrispondenza tra dinamica musicale e forze dovute alle variazioni armoniche, visti come gradienti di un certo potenziale, è uno degli aspetti più interessanti di questo approccio. Su questa linea, si può considerare anche il tempo musicale, le cui fluttuazioni possono essere pensate come funzione della evoluzione dinamica delle strutture armoniche. Si ha così un tempo proprio associato all'esecutore/ascoltatore, in contrasto con un tempo esterno (metronomico). Altri concetti fisici che risultano rilevanti nella nostra analisi sono i frattali, gli orizzonti degli eventi (armonici) e l'*entanglement*.

Infine, abbiamo rimarcato come tutti i processi fisici che associamo alla Musica debbano essere in realtà pensati aver luogo a livello della percezione, cioè nel cervello. Da questo punto di

vista, l'utilizzo nella nostra analisi di concetti propri della fisica quantistica trova riscontro nella descrizione quantistica di alcuni aspetti della dinamica cerebrale (coscienza) come proposto in Reff.[22, 23].

A conclusione, come ulteriore elemento sulla peculiarità della musica di Chopin, riportiamo un brano dal recente libro "Anelli nell'io" di Douglas Hofstadter [24]. L'autore ricorda una conversazione con la madre avvenuta qualche tempo dopo la morte del padre Richard Hofstadter (premio Nobel per la Fisica nel 1961):

"In soggiorno abbiamo un volume con gli Studi per pianoforte di Chopin. Tutte quelle pagine sono soltanto pezzi di carta con delle tracce scure qui e là, bidimensionali, piatti e ripiegabili proprio come la fotografia di papà - eppure pensa al potente effetto che hanno avuto su tante persone in tutto il mondo negli ultimi 150 anni. Grazie a quei segni neri su quei piatti fogli di carta, migliaia e migliaia di persone hanno trascorso complessivamente milioni di ore muovendo le dita sui tasti del pianoforte in complicate configurazioni, producendo suoni in grado di dar loro un piacere indescrivibile e il senso di un messaggio profondo. Quei pianisti, a loro volta, hanno trasmesso a molti milioni di ascoltatori, tra cui te e me, le intense emozioni che si dibattevano nel cuore di Fryderyk Chopin, consentendo così a tutti noi un qualche parziale accesso all'interiorità di Chopin - all'esperienza di vivere nella mente, o meglio nell'anima di Fryderyk Chopin.

I segni su quei fogli di carta non sono niente di meno che schegge d'anima - resti sparsi dell'anima dispersa di Fryderyk Chopin. Ognuna di quelle strane geometrie di note ha il potere straordinario di riportare in vita, dentro i nostri cervelli, qualche minuscolo frammento delle esperienze interiori di un altro essere umano - le sue sofferenze, le sue gioie, le sue più profonde passioni e tensioni - e così sappiamo, almeno in parte, che cosa si provava a essere quell'essere umano, e molti nutrono per lui un

amore intenso.

In modo altrettanto potente, guardare quella fotografia di papà riporta, a noi che lo conoscevamo intimamente, la nitida memoria del suo sorriso e della sua gentilezza, attiva dentro i nostri cervelli viventi alcune delle rappresentazioni più centrali che di lui sopravvivono in noi, fa sì che piccoli frammenti della sua anima danzino ancora, sia pure in cervelli diversi dal suo. Come le note sullo spartito di uno studio di Chopin, quella fotografia è una scheggia d'anima di qualcuno che se ne è andato, ed è qualcosa di cui dovremmo fare tesoro finché viviamo."

7. Il codice dei colori di Scriabin nella notazione musicale

In questa appendice presentiamo una implementazione del codice colori/suoni di Scriabin negli spartiti musicali, utile per una migliore visualizzazione delle strutture armoniche negli studi sopra descritti. Inoltre, questi spartiti colorati possono essere usati per scopi didattici o semplicemente per una più agevole lettura della musica.

L'idea di associare colori e suoni è molto antica, risalente all'antica Grecia o anche prima. Isaac Newton propose una corrispondenza quantitativa tra i colori dello spettro della luce solare e le sette note della scala diatonica, come descritto nel trattato "Optica" (vedi Fig.20):

"Ho trovato che queste osservazioni concordano abbastanza bene con un'altra, e che le rette parallele MG ed FA sono divise dalle suddette linee verticali allo stesso modo delle note musicali. Consideriamo la retta GM in relazione a X , e poniamo che MX sia uguale a GM , quindi consideriamo che le rette GX , λX , ιX , ηX , ϵX , γX , αX , MX siano tra loro in proporzione come i numeri, $1, 8/9, 5/6, 3/4, 2/3, 3/5, 9/16, 1/2$. In questo modo verranno rappresentati

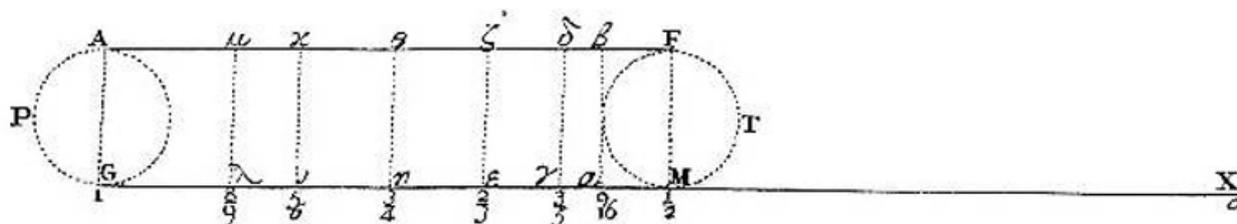


Figura 20: La corrispondenza suoni-colori di I. Newton.

l'intervallo di unisono, il tono, la terza minore, la quarta, la quinta, la sesta maggiore, la settima e l'ottava superiore: allo stesso modo gli intervalli $M\alpha$, $\alpha\gamma$, $\gamma\epsilon$, $\epsilon\eta$, $\eta\iota$, $\iota\lambda$, e λG , indicheranno gli spazi occupati dai rispettivi colori (rosso, arancio, giallo, verde, blu, indaco, violetto)".

Più tardi, uno schema di colori basato sul circolo delle quinte fu proposto da Alexander Scriabin [26, 27], che sosteneva di avere esperienze sinestetiche, ossia di visualizzare colori durante l'ascolto di suoni. Lo schema è riportato in Fig.21: vediamo che colori simili sono associati, per una data tonalità, al quarto e quinto grado della scala, mentre toni distanti hanno colori differenti.

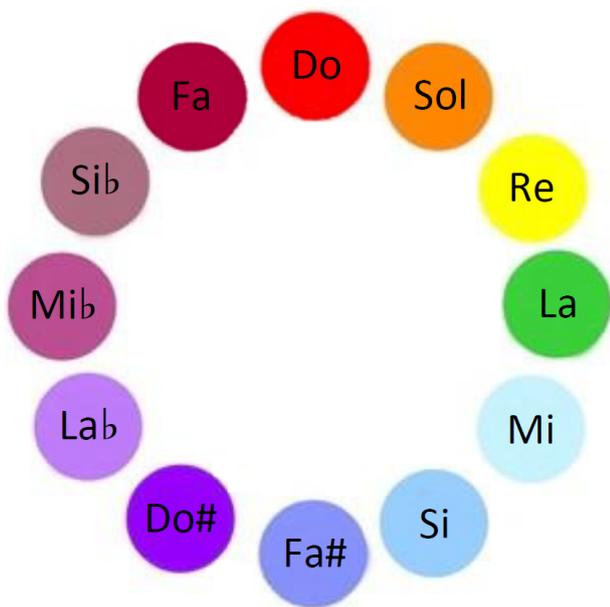


Figura 21: Colori associati da Scriabin ai suoni secondo il circolo delle quinte. Le identità enarmoniche sono sottintese ($Fa\sharp = Sol\flat$, $Do\sharp = Re\flat$, etc.).

È chiaro come questo schema fornisca solo i valori relativi per i colori, laddove quelli assoluti sembrano essere arbitrari (per esempio il rosso per il Do). Per Scriabin questa corrispondenza

aveva un significato preciso ed elaborò una complessa simbologia a riguardo, per esempio: Do (rosso) – Volontà umana; Re (giallo) – Felicità; Mi (blu chiaro/verde) – Sogno; Fa (rosso scuro) – Creatività; La (verde) – Materia; Si (blu) – Contemplazione.

Se si accetta questa corrispondenza, allora è interessante cercare di visualizzarla direttamente sugli spartiti. Questo si può fare in maniera relativamente agevole per mezzo del *software open source* MuseScore (ma anche in altri modi, per esempio tramite $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ - Lilypond). Il risultato è mostrato in Figg. 22 e 23:

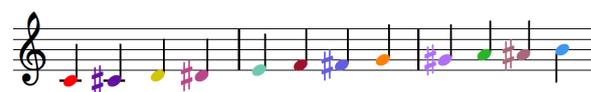


Figura 22: Realizzazione del codice di colori di Scriabin: scala cromatica.



Figura 23: Realizzazione del codice di colori di Scriabin: triadi maggiori.

Si noti come la aggiunta dei colori conferisca alla notazione un carattere assoluto e locale, a differenza della usuale notazione musicale, che richiede una interpretazione di una nota in una data posizione in termini delle alterazioni fisse e transitorie. Inoltre, la semplificazione è evidente nel caso delle enarmonie e di note molto alte o molto basse rispetto al pentagramma (con molti tagli aggiuntivi) che sono di difficile lettura. Queste caratteristiche rendono gli spartiti colorati un utile strumento didattico. Nel contesto del presente lavoro, l'aspetto interessante è costituito dalla possibilità di apprezzare le transizioni armoniche su una scala più larga di quella relati-

va alle singole note, e cioè per gruppi di battute o intere sezioni.



- [1] M.Blasone, *A Physicist's view on Chopin's Etudes*, Eur. Phys. J. Special Topics 226 (2017) 2715.
- [2] J.Fauvel, R.Flood and R.Wilson (Eds.), *Music and Mathematics: From Pythagoras to Fractals*, Oxford Univ. Press (2006).
- [3] D.R.Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, (1999).
- [4] A.Frova, *Fisica nella musica*, Zanichelli (1999).
- [5] I. Xenakis, *Musica Architettura*, Spirali (2003).
- [6] H.J.Brothers, *The Nature of Fractal Music*, in M.Frame and N.Cohen (Eds.) *Benoit Mandelbrot: A Life in Many Dimensions*, World Scientific (2015).
- [7] H.Neuhaus, *The Art of Piano Playing*, Kahn & Averill (2015).
- [8] L.B.Meyer, *Emotion and Meaning in Music*, Univ. of Chicago Press, (1956).
- [9] G.Lakoff and M.Johnson, *Metaphors we live by*, University of Chicago Press, (1980).
- [10] V.Weisskopf in *The Birth of Particle Physics*, Ed. by L.M.Brown and L.Hoddeson, Cambridge University Press, (1986).
- [11] G.Belotti, *Chopin*, EDT, (1984).
- [12] L.Kentner, *Piano*, Collection Yehudi Menuhin, Hatier, (1978).
- [13] M.Majchrzak, *Diversity of the Tonal Structure of Chopin's Etudes*, British Postgraduate Musicology 10 (2009) M1.
- [14] J.Huneker, *Chopin: the man and his music*, William Reeves, London (1901).
- [15] A.Schoenberg, *Theory of Harmony*, Univ. of California Press: 100th anniversary edition, (2010).
- [16] F.J.Budden, *The Fascination of Groups*, Cambridge Univ. Press (1972).
- [17] G.J.Balzano, *The Group-Theoretic Description of 12-Fold and Microtonal Pitch Systems*, Computer Music Journal 4 (1980) 66.
- [18] M.G.Bergomi, R.D.Jadanza, A.Portaluri, *Una geometrizzazione dello spazio degli accordi*, Ithaca: Viaggio nella Scienza (2014) 33.
- [19] M.Blasone, P.Jizba and G.Vitiello, *Quantum Field Theory and Its Macroscopic Manifestations*, Imperial College Press (2011).
- [20] T.W.B.Kibble, *Some implications of a cosmological phase transition*, Phys.Rep.67 (1980) 183.
W.H.Zurek, *Cosmological experiments in superfluid helium?* Nature 317 (1985) 505.
- [21] B.J.Gruber, *Mathematical-Physical Properties of Musical Tone Systems*, Sitz. Öst. Akad. Wiss. Wien, math.-nat. Kl., Abt. II 214 (2005) 43; 215 (2006) 45; 217 (2008) 13.

- [22] G.Vitiello, *My Double Unveiled: The Dissipative Quantum Model of Brain*, John Benjamins Publ. (2001).
- [23] R.Penrose, *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford Univ. Press (2016).
- [24] D.Hofstadter, *Anelli nell'io. Che cosa c'è al cuore della coscienza?*, Mondadori (2010).
- [25] I.Newton, *Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light* (1704).
- [26] L.Verdi, *Kandinskij e Skrjabin. Realtà e Utopia nella Russia pre-rivoluzionaria*, Akademos & Lim (1996).
- [27] F.Balducci, *Musica dell'apocalisse: la rivoluzione di Scriabin*, Simplicissimus Book Farm (2013).



Massimo Blasone: è Professore Associato di Fisica Teorica presso l'Università di Salerno. Si occupa di aspetti fondamentali delle teorie quantistiche. È diplomato in pianoforte.

Galileo, la matematica, la natura

La filosofia naturale è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi, io dico l'universo, ma non si può intendere se prima non s'impara a intendere la lingua e conoscer i caratteri nei quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi ed altre figure geometriche, senza i quali mezzi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro labirinto.

G. Galilei [1]

Abbiamo chiesto di commentare la frase di Galileo citata nel titolo. Presentiamo qui i contributi dei colleghi che hanno partecipato alla nostra iniziativa.

Gino Biondini e Barbara Prinari

Fisici Matematici

La frase di Galileo Galilei (1564 - 1642) ne "Il saggiaiore" si può sintetizzare nella tanto semplice quanto profonda affermazione "Il libro della natura è scritto in lingua matematica". Solo qualche anno prima, nello stesso spirito, Leonardo da Vinci (1452-1519) scriveva: "Nessuna humana investigatione si puo dimandare vera scienza s'essa non passa per la matematica dimonstrazione". Ancora forse più sorprendente è che Leonardo questo scrivesse nel suo "Trattato della pittura", suggerendo un legame senza precedenti tra arte e scienza.

Pur senza trascurare Leonardo, si deve però senza dubbio a Galileo, e in particolare al suo studio quantitativo del moto della caduta dei corpi, l'aver dimostrato il ruolo centrale della matematica nel tentativo di comprendere l'universo che

ci circonda. Da Galileo in poi la matematica è diventata uno strumento prezioso, e in molti casi indispensabile, in quasi tutte le discipline scientifiche. Al punto che Eugene Wigner parlava della "irragionevole efficacia della matematica nelle scienze naturali" [2]. E al punto che al giorno d'oggi siano in molti a ritenere che una disciplina non si possa considerare veramente scientifica a meno che non faccia ricorso in qualche misura a metodi quantitativi.

È inoltre suggestivo come non sia inusuale, nell'ambiente matematico, usare il concetto di bellezza nel descrivere la matematica [3]. Il che ingenera la domanda se si possa davvero pensare ad un'equazione, una teoria, o, più in generale, alla Matematica, come a qualcosa di intrinsecamente bello. Ovviamente l'affermazione non ha valore assoluto. Ma lo stesso si potrebbe dire per l'affermazione che un brano musicale, o qualsiasi altra opera d'arte siano belli. In entrambi i casi, vale il modo di dire che la bellezza risieda nell'occhio di chi guarda. La matematica può essere considerata alla stregua di un pezzo di musica contemporanea, o di un'opera d'arte moderna, che, pur sembrando astrusi e impenetrabili

li ad un profano, risultano chiari, intellegibili ed emozionanti all'occhio o all'orecchio allenati. Si pensi, ad esempio, alla "Grande Fuga" Op. 133 di Ludwig van Beethoven (1770 - 1827), definita dai contemporanei del compositore una mostruosità, il prodotto della follia senile del Maestro [4], ma oggi universalmente riconosciuta come uno dei capolavori della musica classica.

Per tornare a Galileo, la matematica non è che un linguaggio, e come i linguaggi naturali si può adoperare per esprimere una infinita varietà di concetti. Come per un'opera letteraria, sono sia il contenuto che le modalità con cui tale contenuto è espresso cui i matematici si riferiscono quando parlano di bellezza. Per illustrare questa idea, concludiamo così come abbiamo iniziato, con le parole di un gigante, in questo caso del pensiero matematico e filosofico del ventesimo secolo: "*Mathematics, rightly viewed, possesses not only truth, but supreme beauty - a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show.*" [5].

Ferdinando Boero

Zoologo

La filosofia naturale di Galileo è rivolta all'universo, dove l'approccio matematico e geometrico è l'unico possibile. Con il telescopio Galileo ha scrutato l'immensità dove l'occhio umano non era ancora arrivato e, assieme a Copernico, ha cambiato la nostra visione della posizione del mondo nell'universo. La filosofia di Lamarck e di Darwin (i fondatori della teoria dell'evoluzione e dell'ecologia) ha cambiato la nostra visione del mondo vivente e del nostro posto all'interno di esso. La teoria della biologia non si basa sulla matematica, e usa una lingua diversa da quella fatta di cerchi, triangoli e altre figure geometriche elaborata da Galileo. Ci sono più cose, in cielo e (soprattutto) in terra, di quante ne sogni una sola filosofia. La proposta di Galileo è veramente universale, potrebbe essere paragonata all'alfabeto della scienza. Le sue forme geometriche sono le lettere senza le quali non è possibile parlare. La chimica è il dizionario, con la tavola periodica di Mendeleev che mette assieme le sin-

gole parole di cui è composta la materia. E poi ci sono la geologia, e la biochimica, e poi lo stato della materia che non ha paragoni rispetto a tutti gli altri: la materia vivente. Organizzata in milioni di forme meravigliose che si assemblano in entità che cambiano continuamente, ogni giorno diverse dal giorno precedente, in un brandello imprevedibile di realtà. Non esiste nulla di più complesso nell'universo conosciuto. Non si descrive la vita con cerchi e triangoli, e non la si capisce attraverso formule. Bisogna usare l'approccio di Lamarck e di Darwin e di tutti quelli che ancora si affannano a comprendere i misteri della struttura e della funzione della materia vivente. Le lettere e le parole diventano letteratura e poesia, obbediscono alle regole sintattiche e grammaticali scoperte da Galileo e dai suoi seguaci, poi vanno oltre. Ci sono altre lingue e altre realtà. Non esiste unità in questo mondo, ma diversità. La musica non si può ridurre alle sette note e anche se può essere suonata da un computer, non può essere immaginata da un computer. I ricercatori di tutto il mondo devono essere grati a Galileo e al suo metodo. Relativamente a certi livelli di organizzazione è insostituibile, e chi studia cose più complesse ha il dovere di conoscerlo e rispettarlo, pretendendo altrettanto rispetto. Purtroppo c'è chi pensa che la scienza sia solo galileiana e matematizzata, e questo, con tutto il rispetto per Galileo, non è scientificamente corretto. Lo spiega Darwin: "Getta in aria una manciata di piume", dice ne l'Origine delle Specie, "e tutte cadranno al suolo secondo leggi ben definite". Si riferisce alla caduta dei gravi, ai cerchi e ai triangoli di Galileo. Poi continua con: "ma come è semplice questo problema se confrontato con le innumerevoli azioni e reazioni tra piante e animali che, nel corso dei secoli, hanno determinato l'insieme di alberi che formano la foresta davanti ai miei occhi".

La scienza parla molte lingue.

Fabio Bossi

Fisico Sperimentale

Possiamo capire il mondo? Riusciamo a tenere sotto controllo i mille eventi che accadono intorno a noi, indipendentemente da noi, a dispetto di noi, a volte, sembrerebbe, contro di noi? La risposta è sì, possiamo farlo, perché abbiamo svelato

il segreto che ci consente di riuscire ad interrogare il mondo e a riceverne risposte: abbiamo scoperto la matematica. La forza ordinatrice della matematica è talmente potente ed implacabile che ci consente di trovare un ordine nascosto anche nelle cose disordinate, nel caos.

Ci conforta questa consapevolezza? Sì, se ci aiuta a capire che il mondo intorno a noi non è necessariamente ostile, che possiamo comunicare con lui e possiamo tenerlo sotto controllo, e, a volte, addirittura, usarlo a nostro vantaggio. No, se ci fa riflettere sul fatto che la nostra esistenza è del tutto ininfluyente rispetto all'esistenza del mondo ed al suo funzionamento, che le nostre dimensioni sono irrисorie, le nostre aspirazioni irrilevanti.

La matematica è una moneta, che ci dà conforto o scoramento a seconda della faccia che guardiamo. Ma è l'unica moneta autentica che abbiamo a nostra disposizione.

Mario Castellana

Filosofo

Succede a volte anche nella storia delle idee, comprese quelle scientifiche, che la base di alcuni eventi, rivelatisi in seguito significativi come nel caso di Galileo, si trovi posta all'inizio quasi per gioco, come risulta dagli studi sulla nascita della prima Accademia dei Lincei, sorta a Roma nel 17 agosto del 1603 quando dei giovanissimi studiosi, su iniziativa del diciottenne Federico Cesi, si riunirono in casa sua

"in un'atmosfera quasi di congiura, di giuoco segreto e dopo polemico però provvisti *in corpore* dell'occhio della lince per guardare oltre e avventurarsi in quello che veniva chiamato "il gran theatro della natura" ".

Pur dispersi, come era prevedibile, dall'intervento successivo nel 1604 del Sant'Uffizio, il gruppo continuò ad incontrarsi col dare vita a quell'ideale di "lincealità" caratterizzato dall'osservazione dei fenomeni naturali con la matematica per spazzare via le false conoscenze sino all'entrata dello stesso Galileo nel 1611, che diede più concretezza a tale progetto. Non a caso tutti i componenti erano d'accordo nel dare vita a quella che chiameranno "nostra filosofica militia", idea dello

stesso Cesi in "Il natural desiderio di conoscere", grazie all'alleanza strategica con la matematica concepita come strumento *sine qua non*; grazie a questo Gianlo bifronte, il loro iniziale gioco si stava tramutando in qualcosa di molto più serio sino a porre le basi della scienza moderna.

Se in seguito la matematica anche per il ricco ventaglio di articolazioni al suo interno è stata interpretata come un puro linguaggio universale ed utile per ogni contesto di ricerca proprio per la sua presunta neutralità, rimane sempre aperto il problema relativo al fatto che i suoi prodotti sempre più astratti, pur essendo frutto dell'immaginazione umana e quasi "poesia matematica" a dirla con Federigo Enriques, prima o poi trovano corrispondenza quasi perfetta nei fatti reali sino a farli "cantare", per usare un'espressione di Gaston Bachelard, come il calcolo tensoriale prima per il reale relativistico, gli spazi astratti e l'algebra non-commutativa di Dirac, gli spazi hilbertiani per arrivare ai diagrammi di Feynman e così via; ma è da tenere presente nell'ambito di alcuni risultati provenienti dalla neuro-geometria la presa d'atto della forma geometrica degli stessi sillogismi aristotelici e degli schemi mentali di ordine cognitivo, quasi come iscritti nella profondità biologica. Tutto questo non può più essere inserito nel cosiddetto e controverso capitolo della "*The Unreasonable Effectiveness of Mathematics*" [2] che porta ad evidenziarne il carattere quasi miracolistico insieme ai conseguenziali aspetti convenzionali e linguistici, ma deve indurre a pensare le matematiche nel senso genuinamente galileiano come pensiero *tout court*, cioè in grado di conoscere sempre più in profondità le strutture del reale nelle diverse articolazioni.

Occorre, pertanto, lavorare a tracciare dei binari di un discorso fondato sulla "*Reasonable Effectiveness of Mathematics*" [2], di una nuova "filosofica militia", come voleva il giovane Cesi e sulla sua scia Galileo, più in grado di fare i conti coll'implicita razionalità, convinti che interrogandone la storia, stiamo interrogando in realtà la nostra razionalità da più parti messa in discussione. In tal modo, tra l'altro, riacquista un maggior senso l'invito fattoci da Einstein nel suo "Credo epistemologico" prima di morire, dove si lasciava in eredità tale questione dove oltre a credere nel valore esplicativo delle matematiche, era ritenuto un bisogno cruciale sforzarsi di darne una com-

preensione più adeguata, cosa che Galileo all'alba della scienza moderna aveva iniziato a fare nel tentativo di convincere gli aristotelici che anche "le cose cangianti di questo mondo sublunare" erano degni di "dimostrazioni matematiche".

Giampaolo Co'

Fisico teorico

La frase di Galileo va contestualizzata. A me, che opero nel XXI secolo, l'idea di una natura che segue leggi matematicamente definite, da scoprire, appare ingenua, ma devo considerare il contesto in cui Galileo lavorava. Doveva convincere che quanto si osservava nel cannocchiale non era un artificio dello strumento, ma una osservazione oggettiva; tutti coloro che guardavano con quello strumento, o con uno analogo, vedevano lo stesso fenomeno. Oggi, utilizziamo tele-, e micro-, scopi, di varia natura e funzionamento, che ci permettono di osservare fenomeni che appaiono a distanze 10^{-15} volte inferiori e 10^{26} volte superiori a quelle della nostra vita quotidiana, ma nessuno dubita dell'oggettività delle osservazioni fatte, anche perché, dopo Galileo, la ripetibilità dell'osservazione è diventata un paradigma dell'analisi scientifica di un fenomeno.

Abbiamo dovuto riscrivere molto delle leggi che regolano i fenomeni della vita quotidiana per poter descrivere quelli che avvengono su scale micro- e macro- scopiche. Nonostante l'osservazione di fenomeni assolutamente imprevedibili estrapolando la fenomenologia della quotidianità, gli strumenti teorici che adottiamo sono sempre gli stessi. Sono i triangoli, i cerchi e altre figure geometriche, l'oscillatore armonico, le equazioni differenziali. Ci siamo resi conto che i fenomeni che osserviamo sono più complicati e complessi di quanto noi riusciamo ad immaginare, eppure utilizziamo sempre gli stessi strumenti teorici di base.

Sorge, quindi, il pensiero che la matematica non sia nelle cose, come diceva Galileo, ma sia in noi. Certamente, la storia della matematica ci indica chi e quando ha formulato dei concetti ben precisi che oggi utilizziamo, ma rimane il dubbio che la matematica non sia solo un costrutto culturale, ma, nella sua parte più profonda, la logica non contraddittoria a due valori (vero o falso), abbia qualche connessione con qualche

cosa di profondo legato al nostro essere animali pensanti. In linguistica, Noam Chomsky, e chi ne segue le idee, ha formulato l'ipotesi che il linguaggio sia strettamente legato alla biologia umana, distinguendolo dalle lingue che ne sono la realizzazione pragmatica e formalizzata, e quindi prodotto culturale, di questo principio istintivo di base che è insito nella nostra essenza umana.

E se fosse così anche per le basi logiche della matematica? Se non solo esistesse una lingua universale, ma anche una logica universale, non più prodotto culturale, ma necessità biologica. Non è la natura che ci parla in lingua matematica, ma siamo noi che ci siamo evoluti per percepirla in termini matematici.

Stefano De Rubertis

Geografo economico-politico

Il passo che qui proviamo a commentare non esprime soltanto un punto di vista, esso contiene un vero e proprio programma, le basi per una rivoluzione culturale: Galileo indicava una strada per la conoscenza universale, in netta contrapposizione con quella ufficiale. A ben guardare, l'idea di universalità pervade tanto la scienza di Galileo quanto le credenze dell'Inquisizione, suo feroce nemico. Due generi di integralismo che, sia pure in forme molto diverse, si confrontano da secoli, nei contesti più vari e con alterne vicende.

Come osserva con grande lucidità il geografo francese Augustin Berque riflettendo sul rapporto tra scienza e potere, l'idea di uni-verso, nel suo contrapporsi a quella di di-verso, insinua la certezza che esista una sola verità, negando evidentemente qualunque alternativa. Questa assolutizzazione o universione, come la definisce l'autore, "è possibile soltanto in due modi: eliminando l'incommensurabile a vantaggio del commensurabile (che è la via del positivismo moderno) oppure abolendo tutte le misure (che è la via del misticismo)"[6]. Insomma, sembra emergere una consapevolezza che definirei postmoderna, foriera di non pochi problemi, non ultimo quello di veder mettere in discussione la stessa legittimazione della scienza. O, al contrario, di assumerla come verità assoluta: rendere la realtà trasparente, regolare, geometrica, cancellando

l'utilità e il senso di qualunque confronto, di fatto riducendo i gradi di libertà individuale.

Evidentemente, una solida fiducia nella scienza, alla cui costruzione Galileo ha tanto contribuito, non è un risultato ottenuto una volta per tutte. Razionale e irrazionale si fronteggiano, essendo l'uno complemento dell'altro. Il commensurabile ha senso grazie al suo opposto. La lettura e le interpretazioni del mondo quanto più diventano profonde e incisive tanto più restituiscono immagini sfocate della realtà. Le probabilità si sostituiscono alle certezze. Le relazioni sono più importanti degli elementi. La soggettività e le qualità dell'osservatore ridefiniscono il significato di oggettività delle osservazioni. La ricerca di nuovi equilibri ha pervaso, sia pure in misura e modalità differenti, ogni ambito disciplinare, con una nuova e più sofisticata consapevolezza delle proprie possibilità e delle proprie debolezze.

Così, anche la geografia moltiplica le sue anime, utilizza qualitativo e quantitativo, gradualmente (molto gradualmente) abbandona presunzioni di oggettività assoluta per giungere a verità parziali e temporanee, maturate attraverso processi di condivisione. In ogni epoca il prestigio della disciplina è derivato dalla sua utilità, che in futuro si misurerà sempre più con la sua capacità di rendere conciliabili rappresentazioni diverse, sottraendo a ognuna inaccettabili pretese di universalità, preferendo il confronto al conflitto. Ciò non toglie che la geografia sia e sarà ancora una scienza, visto che ne ha mutuato il metodo e che, anche quando non ricorre al linguaggio matematico, ne utilizza la logica rigorosa, sempre pronta a mettersi in discussione. Oggi la geografia della post-modernità è più matura e consapevole dei propri limiti. Questo non ne restringe la validità, anzi, è utile oggi almeno quanto lo è stata nel passato.

Sandra Lucente

Matematica

Caro Galileo, circa 400 anni fa inviasti *Il Saggiatore*, elegante trattato in forma di lettera, al poeta Virginio Cesarini. Allora ti rispondo anche io in forma di lettera. Hai notato che la parola lettera è un bisenso? Indica sia il singolo elemento dell'alfabeto sia la composizione di tanti segni imbustati.

Non è un caso che la tua lettera fosse indirizzata ad un poeta, tante lettere ad un letterato! Immagino la prima lettura di quella frase da parte di chi doveva solo controllare che tutti i caratteri fossero al posto giusto. Magari il tipografo si mise a pensare all'universo come libro. Quanti caratteri occorre per descriverlo? Quattrocento anni dopo i caratteri si infilano ordinati su un monitor e solo a volte sulla carta. Ma non sai cos'è un monitor e prima dovrei raccontarti del meraviglioso passaggio intermedio alla macchina da scrivere che ha messo un *set* di caratteri in tante case. Ogni tanto passavamo il dito sulla C per stampare la lettera sul polpastrello. C di cometa dirai, non ne sono esperta, ti voglio scrivere di C di carattere. Un altro bisenso, un segno tracciato ma anche le doti dell'individuo. Torniamo alla tua idea di una gigantesca stamperia o macchina da scrivere che ricostruisca il libro dell'universo e che contenga tutti i caratteri della matematica. Vedi Galileo, innanzitutto deve essere una macchina da scrivere in espansione, e non tanto perché si espande l'universo, cosa che tu puoi discutere meglio di me, quanto perché si espande la matematica e acquisisce nuovi caratteri. Così si riesce a girare altre pagine del libro dell'universo. Ti faccio un esempio. Ti ricordi delle dispute algebriche che il secolo precedente al tuo portarono ad introdurre i numeri immaginari? Ecco, oggi li usiamo per capire i segreti dell'infinitamente piccolo. Si chiama equazione di Schrödinger e dovrebbe interessarti tanto. Se tu hai studiato le grandi macchie solari, noi col tuo metodo scientifico ci siamo interessati alle particelle più piccole della materia. Se ripensi alla fatica di misurare distanze planetarie con il tuo cannocchiale, ti sembrerà grande fortuna il nostro essere dotati di satelliti artificiali che ci guardano da lontano. Ma anche per costruire questi abbiamo dovuto allargare il *set* di caratteri matematici, e per interpretare le notizie che ci mandano abbiamo persino deciso di usare caratteri appartenenti a diversi *font*. La macchina da scrivere immaginaria della matematica ha oggi infatti diverse geometrie che ci regalano i percorsi più brevi sulle superfici più disparate. Quando scriviamo di raggi di luce in spazi amplissimi e non teniamo fermo l'osservatore non ci importa più che le parallele non si incontrino. Ma nel semplice lavoro di chi mette in fila arbusti vo-

gliamo invece lasciare costante le distanze nel campo per favorire la crescita delle piante e i saggi assiomi di Euclide ci occorrono ancora. Lo so che non hai nemmeno letto l'ultima frase, stai ancora pensando all'osservatore che si muove e a quanto quell'Einstein che non hai mai sentito nominare sia stato influenzato dal tuo discorso sui sistemi inerziali. Ma ti prego seguimi ora sulla faccenda degli alberi. Hai sicuramente notato che le radici e i rami seguono la stessa regola della diramazione che si legge sulle foglie. Come la esse maiuscola dei libri antichi somigliava alla effe minuscola e per certi libri bisognava avere un monoclo per distinguere i caratteri, così noi abbiamo guardato da vicino e da lontano i fenomeni. In alcuni oggetti questa doppia visione non dà differenze. Mandelbrot nel 1970 ha iniziato a chiamare questi oggetti frattali, ha persino proseguito la tua frase dicendo "Le nuvole non sono delle sfere, le montagne non sono dei cono, le linee costiere non sono dei cerchi, il sughero non è liscio ed i fulmini non si muovono lungo linee diritte". Non era in polemica con te, voleva solo chiedere di allargare la stamperia, di creare nuovi caratteri per quello che cresce con poca regolarità. Sono nate nuove idee di perimetro, di area, di volume. Avrai subito pensato a quel che scrivevi nel Sidereus Nuncius sulla Luna che non è una sfera perfetta ma ha monti e valli, sulla meravigliosa Via Lattea che sembra un ammasso. I tuoi seguaci cosmologi stanno discutendo ancora su cosa questa teoria dice sul susseguirsi di vuoti e pieni di materia, su scale diverse. Di certo, per affermare o smentire una idea usano sempre la matematica i cui caratteri, come quelli delle parole, se scambiati cambiano senso. Causalità e casualità ad esempio. La teoria dei frattali è vantaggiosa perché mette insieme queste due parole e perché guarda i mondi in movimento. Sai, oggi abbiamo il carattere della quarta dimensione, il tempo. Anzi stiamo già leggendo le pagine del libro sulla dimensione infinita che ci serve per i movimenti dei fluidi, e ti prometto di spiegartela la prossima volta, così ti togli tutti i dubbi che avevi sulle maree. Noi ne abbiamo molti ancora sui vortici e sul moto dei gas. La vera rivoluzione (oltre quella dovuta a voi giganti del Seicento si intende) è arrivata quando abbiamo introdotto in stamperia i caratteri della probabilità e della incertezza, così le passeggiate di parti piccolis-

sime della materia le discutiamo non punto per punto ma con funzioni d'onda. Ovviamente i vecchi caratteri non sono desueti e le passeggiate di informazioni le discutiamo con la teoria dei grafi, un set di caratteri nato poco dopo di te, molto potente e molto schematico. Cosa succede ora? Quello che succede sempre nelle lettere: si cerca una chiusura. Oggi si usano le *emoticon*, dopo tutti questi caratteri inventati siamo tornati ai disegni. Ma io cercherò una chiusura classica, perché sono abituata all'eternità dei caratteri matematici, alla ricerca di regole nella complessità. Questo rende meno oscuro il labirinto del sapere. Ci incontreremo lì, perché in fondo, a tutti gli scienziati piace quel labirinto. Parafrasando il finale della tua lettera: per minor tedio di V. S. Illustrissima e mio, ho risolto più tosto di rimmetterla a tante altre letture dello stesso libro nei secoli futuri.

Luigi Martina

Fisico Teorico

Nelle parole di Galilei confluiscono il "*De Divina Proportione*" e il "*Mysterium Cosmographicum*": rinascimento italiano e riforma luterana. Forgiando la cultura europea, il tempo ha prodotto neoplatonismo e positivismo. Il dramma prima scientifico, e poi personale, di Boltzmann si collocò esattamente al centro dello scontro tra tali concezioni totalitarie, che produssero poi gli orrori successivi. Le riflessioni sui fondamenti della Matematica (il modo di pensare) e della Fisica (il modo di percepire l'universo), accompagnarono le contemporanee scomposizioni dell'anima (Freud) e dell'arte (Picasso), verso una deviazione permanente nel cammino del *logos* creatore. Le magnifiche sorti e progressive vennero sepolte assieme al vitalismo romantico. Matematica, Scienza ed Arte condividono ora l'incertezza e l'ansia dell'umanità. Non più triangoli, ma frattali. Non ragioniamo più con l'orbita, ma in termini di orbitali. Il banditore del re è sostituito dal web e dagli algoritmi di profilazione. Il cinquecentesco ritratto della matematica famiglia Neudorffer è sostituito dallo sberleffo al mondo di Einstein.

Marco Mazzeo

Fisico Sperimentale

Per comprendere in fondo la famosa frase di Galilei, che potremmo per certi versi definire il fondamento dello statuto epistemologico della scienza moderna, essa va divisa in due parti, inserendola, inoltre, nel suo contesto storico.

Nella prima lo scienziato pisano afferma che la filosofia naturale, pur essendo una attività meramente umana, è scritta in un libro, che non è però cartaceo ma è niente meno che l'universo stesso. Il senso di questa affermazione è soprattutto fondativo: Galilei intende rendere indipendente la scienza da altre discipline, dalla teologia alla filosofia in particolare, e dalla tradizione. Ovviamente Galilei con questo non voleva affermare che tutta la tradizione passata fosse da buttare. Voleva piuttosto sostenere che è lo studio diretto del mondo naturale la fonte della vera conoscenza, e che al più i testi del passato possono avere la loro validità a patto che siano supportati dai fatti osservativi e che siano studiati con senso critico. È evidente qui l'attacco alla filosofia aristotelico-scolastica di quel periodo, la quale era divenuta una sequenza di commenti dei testi aristotelici.

La seconda parte della frase Galilei sostiene che l'universo stesso, si badi bene non la filosofia naturale ma la stessa realtà, sia scritto in linguaggio matematico. Essa va contestualizzata nel dibattito tra due diversi modi di intendere il ruolo della matematica nelle scienze naturali e ha enormi implicazioni circa il rapporto che decidiamo debba esserci tra scienza e verità. Questi due punti di vista sono la visione modellistica e quella platonista. Nel primo caso il rapporto che c'è tra matematica e realtà è lo stesso che passa tra una mappa geografica dell'Italia e la penisola vera e propria: ovvero una ottima rappresentazione ma pur sempre una rappresentazione, la quale può migliorare con rappresentazioni più efficaci e dettagliate. In quest'ottica il rapporto tra scienza e verità è solo un rapporto di efficacia: le teorie semplificano il mondo mediante enti geometrici o algebrici e consentono di fare previsioni utili mediante teoremi. Questi enti vivono in un mondo virtuale (culturale), entro cui abbiamo stabilito certe regole matematiche e logiche, e da cui possiamo, mediante teoremi, estrarre certe conclusioni da confrontare, mediante

regole di corrispondenza, con la realtà naturale (fenomeni osservati), artificiale (tecnologia) o tecno-naturale (esperimento). Per un modellista dunque non esiste una teoria vera, ma solo una teoria efficace. Per i platonisti (moderni pitagorici) la matematica ha un ruolo differente: essa è intesa come il linguaggio della struttura stessa dell'universo e della sua dinamica, la quale va esplorata come si esplora la geografia di un nuovo ambiente. Le teorie in questo senso sono da scoprire e non da inventare: è il caso della famigerata teoria del Tutto, nei confronti della quale vi è la fede (platonica) che esista una struttura matematica da scoprire per unificare le forze e che essa sia una e necessaria.

Galilei era un modellista o un platonista? Per un modellista il modello copernicano è più efficace nel descrivere i fenomeni. Per un platonista invece la teoria copernicana è vera in virtù dei fenomeni che riesce a spiegare. Ricordiamo che il moto retrogrado del pianeta Marte poteva essere descritto adeguatamente sia dal modello tolemaico che da quello copernicano. Tuttavia Galilei sosteneva che la teoria copernicana fosse vera e quella tolemaica falsa. Come mai, visto che lui fu il campione dei relativisti per i quali, in nome proprio della relatività, è equivalente dire che la terra è ferma con il sole che gli si muove attorno o viceversa? In realtà le osservazioni bocciarono il modello matematico di Tolomeo, il quale, come prontamente dimostrò lo stesso Galilei mediante le osservazioni effettuate al suo cannocchiale, non poteva spiegare le fasi di Venere.

Galilei era evidentemente un platonista, sia perché, come lui stesso affermò più volte si definì pitagorico, sia perché la mentalità cristiano-monoteista, supponendo l'esistenza di una intelligenza divina la quale crea e ordina il cosmo, è più pronta a ritenere che esplorare la matematica dell'universo equivalga a capire la struttura con cui Dio ha fondato il mondo. La posizione platonista ha la sua legittimità. Tuttavia essa non tiene conto del fatto che le teorie evolvono nel tempo non tanto col risultato di creare modelli sempre più precisi di un canovaccio consolidato, quanto con quello di creare nuove strutture matematiche ed enti che scalzano completamente quelli precedenti. È la lezione che abbiamo imparato un secolo fa, quando i fisici, gran parte platonisti, si meravigliarono non poco quando videro che

i modelli classici erano incapaci a spiegare certi fenomeni.

Il mio punto di vista, da modellista, è che il platonismo, riportato in auge da Galilei, è stato di indubbia efficacia per 400 anni e lo sarà ancora, come una "finzione che funziona", ovvero come un modo euristico di avvicinarci al mondo. Tuttavia se si prende troppo sul serio si rischia di ritenere assolutamente vere idee che sono solo "dinamicamente e temporaneamente vere", in quanto meri modelli contingenti e culturali del mondo, pur essendo vincolati ai fenomeni osservati.

Margherita Primavera

Fisica Sperimentale

Con tutto il rispetto per Galileo e la matematica (che adoro), a me sembra che il linguaggio di cui scrive Galileo vada inteso come l'abecedario, fondamentale certamente ... ma da lì a scrivere "I fratelli Karamazov" (solo per fare un esempio) ce ne passa: c'è di mezzo la comprensione della complessità dell'animo umano, la sensibilità di coglierlo nel profondo eccetera eccetera eccetera. Ho usato una metafora, ma spero che si colga cosa voglio esprimere!!

Marco Ticozzi

Ingegnere

Ad un ingegnere civile, qual io sono, il corso degli studi offre nei primi anni una sicurezza ontologica: la matematica e la geometria "sono", esistono apoditticamente perché servono a fare poi altre cose concrete. Sono a tutti gli effetti l'A-B-C da imparare per elaborare modelli interpretativi della realtà che poi si concretizzano in progetto.

Cito (a braccio) la conclusione di un testo di Giuseppe Grandori, uno dei pochi maestri che io abbia conosciuto all'università, nel quale egli asseriva che il tutto (la Scienza delle Costruzioni) era mirato all'interpretazione dei reconditi disegni della Provvidenza (la natura). In effetti, un ingegnere da giovane vede la realtà, a partire dalla composizione vettoriale delle forze nella Statica, come una successione, una ragnatela di forme geometriche che si compongono alla ricerca di un equilibrio e, in fin dei conti, di un

senso. La natura così è rappresentata *a fortiori* da geometrie connesse da relazioni matematiche.

Non vi sono dubbi epistemologici, semmai, dopo, dubbi pratici: "funzionerà?"; "starà in piedi?". Come un novello, sebbene erudito, "buon selvaggio" davanti all'asserzione di Galileo: "è così, e non c'è altro da dire", se non progredire e valutare criticamente il prodotto.

E l'imprinting iniziale spesso resta il parametro esperienziale per valutare l'evoluzione delle tecniche di calcolo: chi è nato con il metodo di De Saint Venant, assolutamente geometrico, interpreta le formulazioni di calcolo più raffinate, che si sono succedute nel tempo, usando quelle imparate agli esordi come contro-verifica istintiva.

Ora, l'uso della geometria come complesso di sintagmi di un linguaggio comune e convenuto, comunicativo ed efficiente per l'indagine, resta, quantomeno in me, la forma mentis dell'Ego (buon selvaggio), della supposta razionalità, ma anche chiave di lettura della natura nel bisogno di controllare e leggere anche l'emozione: un dipinto, una scultura, un'architettura appaiono ordinate secondo ordini geometrici che ne contemplano anche varianti e digressioni.

L'esperienza esistenziale della contraddizione, della varianza, della confutazione degli schemi propongono (all'Es-inconscio) l'ipotesi che non tutto possa essere descritto e che il lessico possa variare a seconda della cultura e dei punti di vista. Fino alla domanda ingenua se davvero esistano o no altri modi, altri linguaggi rispetto alla matematica, che è stata inventata, più che scoperta, nei tempi antichi. Perché il tutto nella sua molteplice ma unitaria diversità sembra sfuggirne.

Al di là delle fantasie infantili in campo fantascientifico e delle senili derive spiritual-esoteriche, la visione geometrica dialoga con la contezza delle contraddizioni, proponendo il tarlo creativo del dubbio. "Tutto si tiene" e si inizia con una lingua per approfondire o avvicinarsi al tutto. La certezza del linguaggio matematico e geometrico, assieme alla coscienza di derivare dalla cultura di origine Greco-Mediterranea, restano le chiavi della mia fascinazione verso la geometria. Come in un film di fantascienza, tutto quello che vedo si scompone in elementi geometrici cosiccome tutto quello che amo in quanto bel-

lo: ad esempio la pittura che, fino all'astrattismo del Novecento, si basa ancora (anche nel confutarla) sulla composizione geometrica rinascimentale secondo le forme che Galileo intravede nella natura.

Mi piace sempre pensare a come una costruzione di geometria applicata, quale è la prospettiva per la pittura e il disegno, venisse vissuta da Leon Battista Alberti e da Leonardo, per esempio, come strumento scientifico di conoscenza e rappresentazione della realtà, o meglio della natura; mentre nelle opere d'arte si applicavano i canoni di bellezza a loro volta figli di quelle proporzioni geometrico-numeriche teorizzate dei Greci nell'antichità.

E qui sta il primo dubbio costruttivo: questa bellezza, matematica, aveva trovato nei Greci giustificazione in quanto espressione di bontà e di giustizia, secondo visioni dunque relative al contesto culturale e alle necessità di ordine, anche sociale.

Il richiamo ai miti antichi, come forieri di una psicoanalisi collettiva, da parte dei padri della psicoanalisi moderna - individuale -, scardinando le certezze della modernità e permeando la coscienza del sé nel contemporaneo, rafforzano poi nel "buon selvaggio" il dubbio che possano esistere altre formulazioni e che il tutto non possa essere compreso soltanto in questa sintassi.

Ma come miti di bellezza estetica e di pensiero, nella mia visione geometrica del mondo restano affascinanti il quinto postulato di Euclide - cosiesteticamente raffinato e astruso, ma icasticamente risolutivo in un ambito di riferimento allo spazio piano a due coordinate - e l'assurda ma esteticamente efficacissima ricerca della sezione aurea, che tutte le volte mi sorprende con la domanda "*cui prodest?*", ma arriva con un numero impossibile - 1,6 e tanti decimali - a definire delle proporzioni estetiche che, tuttora radicate nello spirito, si dimostrano efficaci sia nel loro riprodursi ordinato che nel generare varianti.

Dunque sì, Galileo mi ha formato secondo l'asserzione in causa, e questo si è sedimentato con convinzione, ma non basta forse per la molteplicità naturale e culturale, e ne ammetto il limite personale.

- [1] G. Galilei: *Il Saggiatore, nel quale con bilancia esquisita e giusta si ponderano le cose contenute nella Libra astronomica e filosofica di Lotario Sarsi Sigensano, scritto in forma di lettera all'ill.mo et rever.mo mons.re d. Virginio Cesarini acc.o linceo m.o di camera di N.S.*, Giacomo Mascardi, Roma (1623).
- [2] E. Wigner: *The unreasonable effectiveness of mathematics in the natural sciences*, Communications on Pure and Applied Mathematics, 13 (1960) 1.
- [3] H. E. Huntley: *The divine proportion: a study in mathematical beauty*, (Dover Publications, città (1970), Serge Lang: *The beauty of doing mathematics*, Springer, Berlin (1985), G.-C. Rota: *The phenomenology of mathematical beauty*, Synthese (1977) 171.
- [4] G. Carli Ballola: *Beethoven: la vita e la musica*, Bompiani, Milano (2001).
- [5] B. Russell: *The study of mathematics in Mysticism and logic*, Penguin, città (1953).
- [6] A. Berque: *Ecoumène. Introduction à l'étude des milieux humains*, Belin, Paris (2000). p. 65



Numero XVII Anno 2021

Viaggio nella Scienza

Ithaca

Dare i numeri: (ab)usi della Matematica (A)

