

# Apologia della ragione scientifica – IV: dilemmi di scelta ed etica dell'IA

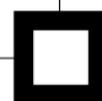
T. Bassetti, A. Luvison

## Sommario

*La scienza e la tecnologia sono due pilastri fondamentali per lo sviluppo di ogni Paese. Dal punto di vista metodologico e operativo, il pensiero critico basato sul ragionamento scientifico fornisce strumenti concettuali adeguati per prendere decisioni in condizioni d'incertezza. Conoscendo e superando i condizionamenti (o bias) cognitivi, la razionalità stessa risulta rafforzata dall'esercizio continuo. L'articolo discute due casi emblematici. Il primo è il paradosso di Newcomb, un esperimento mentale di natura logico-filosofica, riguardante un gioco a due giocatori, uno dei quali è dotato di capacità predittive. La sua soluzione è ricondotta a un semplice calcolo algebrico e probabilistico. Partendo dal problema del carrello, la focalizzazione del secondo caso è sui dilemmi di carattere socio-etico, che le applicazioni dell'intelligenza artificiale (IA) sollevano in modo dirimpente, in particolare, per le auto a guida autonoma. Fondamentale è l'assunzione di responsabilità (accountability) da parte dei progettisti di questi sistemi. La tesi di base è che le due culture – l'umanistica e la tecnoscientifica – grazie ai numerosi punti di fertilizzazione reciproca, anziché realtà separate, sono due facce della stessa medaglia: il sapere interdisciplinare. L'interdisciplinarietà è componente fondamentale del bagaglio di conoscenze che ogni leader o studioso deve padroneggiare per affrontare l'impegno morale delle sfide poste in essere dalla società digitale.*

## Abstract

*Science and technology are two major drivers for development in the most advanced countries. Methodologically, critical thinking, motivated reasoning and cognitive reflection provide the conceptual framework and operational toolbox for decision making under*



*uncertainty. The inherent limits (i.e., biases) of human bounded rationality can be overcome, so that scientific rationality is enhanced even more. Two cases are discussed to support this assumption. The first is the Newcomb's paradox, a logical and philosophical thought experiment entailing a game between two players, one of whom claims to be able to predict the future. The solution to this brainteaser is based on elementary algebra involving simple probabilities. By starting from the trolley problem, the focus of the second case is on ethical issues of artificial intelligence (AI), e.g., autonomous systems, such as driverless cars, or other disruptive and pervasive AI applications. Here, the term accountability applies to a designer who considers the question of how intelligent systems should be imbued with ethical values. The underlying argument is that the two cultures – humanities and techno-science – thanks to the many intersection and cross-fertilization points, are both faces of the same coin, i.e., interdisciplinary knowledge. This type of knowledge should belong to the education and training background of any leader, executive, or opinion maker, responsible for facing the incumbent challenges of the digital society.*

**Keywords:** Science and humanities, Critical decision making, Logical conundrums, Newcomb's paradox revisited, Trolley problem, AI and machine ethics

*La preoccupazione per l'uomo e il suo destino deve sempre essere il principale interesse di ogni sforzo tecnico*  
(Albert Einstein)

*Quando soffia il vento del cambiamento, alcuni costruiscono muri, altri costruiscono mulini a vento [per sfruttarne l'energia]*  
(Detto cinese)

## 1. Introduzione

*"[L]e trovate della tecnologia di largo uso hanno risolto molti problemi pratici, ma raramente – e quasi preterintenzionalmente – si sono trasformate in motori d'innalzamento della nostra cultura, o di affinamento delle nostre capacità intellettuali. La tecnologia più comune – quella che oggi gratifichiamo della qualifica d'innovazione digitale – desertifica di giorno in giorno attenzione, reattività, spirito critico, in breve l'intelligenza di miliardi d'utenti, persuadendoli della sostanziale inutilità d'una conoscenza duramente e lentamente conseguita" [1].* Queste affermazioni, blandamente provocatorie, di un brillante linguista accademico sollevano questioni concrete e indicano elementi di riflessione critica. Insanabile e irrimediabile fu, invece, la frattura provocata dai filosofi neoidealisti Benedetto Croce e Giovanni Gentile, che, all'inizio del Novecento, lanciarono con inaudita *vis polemica* una loro personale crociata contro il matematico ed epistemologo Federigo Enriques e altri primari

esponenti della cultura scientifica. Questi furono denigrati come “ingegni minuti” (locuzione risalente a Giambattista Vico), buoni solo per mestieri da “banausico”, ossia meramente utilitaristici. Del tutto opposta era la mentalità di un intellettuale tedesco di allora, il teologo e storico Adolf von Harnack, il quale nel 1911 dichiarava: “*La gente si lamenta che la nostra generazione non abbia filosofi. Non è assolutamente vero: solo che i filosofi, oggi, appartengono a un altro settore del sapere, e si chiamano Planck e Einstein*” ([2, pp. 102-103]).

La scienza, non avendo dogmi, progredisce per acquisizioni, avanzamenti, affinamenti successivi, non verso certezze assolute bensì verso certezze ragionevoli, cioè basate sull’esercizio della ragione e sull’esperienza. Più che produrre certezze, la scienza riduce quindi, le incertezze, o fornisce una misura dei livelli di incertezza. Nel suo solito stile schietto e irriverente, Richard Feynman dichiarava: “*Non abbiamo mai ragione con certezza, possiamo solo essere sicuri di avere torto*”. La ragione scientifica non ha di certo la pretesa di essere universale, completa, di poter conoscere tutto; anzi, ha la consapevolezza, in termini di teoremi e risultati, che le nostre conoscenze sono limitate: si pensi, per esempio, ai teoremi di Gödel in logica matematica o al principio di indeterminazione in fisica quantistica. Ma alla fine la scienza, imparando dai suoi eventuali errori, costituisce ancora la miglior forma di conoscenza.

In un breve articolo-saggio [3], John Harris, professore di Bioetica al King's College di Londra, sviluppa un ragionamento ineccepibile nella successione logico-sintattica e nella coerenza argomentativa, dicendo fra l’altro:

*La preparazione per il futuro esige scienza e tecnologia ed esige che le menti siano preparate a porsi domande liberamente, a basarsi sulle evidenze [prove di efficacia (N.d.A.)] e, soprattutto, che le menti siano abituate all’onestà e ad evitare comode “verità alternative”, così come potrebbero presentarsi adottando una concezione un po’ – diciamo – allargata di scienza.*

Diventa sempre più necessario esercitare un acuto senso critico guidato dal pensiero razionale, dalla logica, dalla capacità di analisi e di giudizio, alimentati dall’informazione fattuale. Senza dimenticare che la ricerca deve sempre ispirarsi a principi e codici etici estremamente solidi, non disgiunti da una precisa assunzione di responsabilità delle conseguenze (*accountability*), dall’impegno civico-sociale e dall’atteggiamento etico-morale. Non meno rilevante è la coerenza tra il comportamento e il pensiero, che si dovrebbe seguire sulle questioni più sensibili e i temi più cruciali, tuttora aperti e ampiamente dibattuti. Le finalità di uno studioso devono essere sia educative sia culturali, tanto formative quanto informative, perché il suo ruolo non può limitarsi alla ricerca meramente speculativa o all’attività di laboratorio.

Questo articolo ha l’obiettivo di integrare e completare i precedenti sull’“Apologia della ragione scientifica” [4]-[6], mantenendo l’impostazione trasversale e interdisciplinare, nonché il taglio panoramico e l’intento

comunicativo.<sup>1</sup> Leitmotiv del lavoro è che l'esercizio del pensiero critico (*critical thinking*), della razionalità, del ragionamento motivato, della riflessione cognitiva, dell'argomentazione, insieme con gli strumenti di supporto (analisi quantitativa e probabilistica, rigore logico e metodologico, pragmatismo, ecc.), consente di inquadrare, affrontare e risolvere efficacemente i problemi che caratterizzano l'ecosistema digitale, nel presente e nel futuro.

Per dare concretezza alle tesi precedenti proponiamo e discutiamo due casi emblematici: il paradosso di Newcomb di carattere logico-filosofico e il dilemma del trolley come si manifesta nell'intelligenza artificiale (IA) applicata ai veicoli a guida autonoma.<sup>2</sup> I paragrafi 2, 3 e il riquadro 1 illustrano il paradosso di Newcomb, un gioco a due giocatori, uno dei quali vanta doti di predizione. Il problema-rompicapo viene generalizzato nei suoi parametri costituenti e ridefinito in termini di un criterio di scelta ottimale basato sul confronto dei valori attesi delle due scelte possibili, dove l'alternativa è qui riformulata in termini unicamente probabilistici. I paragrafi 4, 5 e il riquadro 2 hanno come tema principale il dilemma del trolley nel contesto dei problemi etici (o morali) sollevati dall'IA e dai sistemi autonomi. Il secondo riquadro riporta una curiosità all'origine del dilemma del trolley durante un episodio della Seconda guerra mondiale, che vede Winston Churchill protagonista. Il quesito di logica del riquadro 3 è un bonus finale per il lettore, nonché un'efficace sintesi delle nostre argomentazioni. Si tratta, infatti, di un enigma di natura logica, la cui soluzione, semplice e ovvia solo dopo che viene svelata, porta a chiedersi: "Com'è che non ci abbiamo pensato prima?".

La bibliografia comprende le indicazioni relative ai lavori – molti dei quali di carattere interdisciplinare – effettivamente utilizzati, o comunque consultati, per la stesura del testo. È perciò rappresentativa della "cassetta degli attrezzi", con i quali abbiamo compiuto l'esplorazione di alcuni temi decisivi per il rapporto (alleanza) tra tecnoscienza, o tecnologia scientizzata, e studi umanistici e società. Questi riferimenti possono essere utili in approfondimenti successivi.

## 2. Il paradosso di Newcomb: il problema

Il concetto di razionalità limitata (*bounded rationality*) – introdotto da Herbert Simon e ripreso in anni recenti da economisti e investitori attenti agli aspetti psicologici e comportamentali [7]-[10] – implica che non possiamo prendere decisioni perfettamente razionali, perché le informazioni di cui disponiamo sono incomplete; spesso manchiamo del tempo necessario per una decisione

<sup>1</sup> Si rimanda a [4]-[6] per approfondimenti sulle motivazioni che hanno ispirato la serie di articoli dedicati all'"Apologia della ragione scientifica" (titolo originariamente proposto dal prof. Franco Filippazzi dell'AICA). A essi si farà frequente riferimento per inquadrare la trattazione in una visione complessivamente unitaria nelle tesi e nelle argomentazioni.

<sup>2</sup> Poiché l'IA è una disciplina in rapida evoluzione, non sembra esistere oggi una definizione dell'IA universalmente accettata e comprensiva delle sue numerose sfaccettature teoriche e pratiche. A titolo di esempio, dall'Enciclopedia Treccani online riportiamo: "Si chiama intelligenza artificiale (IA) quel settore dell'informatica che studia la possibilità di costruire computer che siano in grado di riprodurre il funzionamento di alcune capacità della mente umana o, nel caso della cosiddetta intelligenza artificiale forte, dell'intero pensiero umano. Questa locuzione è anche utilizzata per indicare quella branca della filosofia che si pone il problema se sia davvero possibile riprodurre il pensiero umano". (A firma di Andrea Carobene, <http://www.treccani.it/enciclopedia/intelligenza-artificiale>).

ponderata e, considerazione ancora più importante, le nostre capacità cognitive sono limitate. Gregory Bateson annotava: “*Fra tutte le invenzioni umane, quella dell'uomo economico è di gran lunga la più noiosa*”.

Nelle scuole di psicologia cognitiva è in uso il termine inglese *dysrationalia* (“disrazionalità”) per indicare l'incapacità di una persona, nonostante il suo alto quoziente intellettivo (QI), di pensare e comportarsi razionalmente, o almeno ragionevolmente. È evidente che molti, pur dotati di un'intelligenza elevata secondo il parametro QI, non sono allenati al ragionamento critico. Il QI non è quindi un buon criterio di valutazione del pensiero razionale, al più ne è un blando indicatore [11].<sup>3</sup>

Eppure, razionalità limitata non significa affatto irrazionalismo, o che non conosciamo alcunché: sarebbe bene che molti critici o pseudoesperti “*si facessero una ragione di ciò*”. L'economia comportamentale nasce, in definitiva, come intersezione tra la psicologia cognitiva e comportamentale, da un lato, e le decisioni economico-finanziarie, dall'altro. Riferito ai mercati finanziari, ma di portata ben più ampia, è l'icastico commento: “[L]’unico antidoto agli errori indotti dalle emozioni è la razionalità, specialmente se applicate sul lungo periodo, con pazienza e perseveranza” [10, p. 222].

Con l'ausilio di esempi provenienti da settori fra i più vari, i limiti cognitivi sono stati ampiamente discussi in [4]-[6], dimostrando però che, con l'esercizio, la razionalità, anziché bandita, risulta addirittura ampliata e rafforzata. (Cfr. anche il riquadro 3 per un quesito logico, la cui soluzione, pur ovvia *a posteriori*, non è in generale percepita subito; anzi, *a priori* sembra controintuitiva).

In [6], abbiamo delineato il quadro generale della teoria dei giochi, ricorrendo a numerosi esempi. Nei prossimi due paragrafi illustriamo, analizzandolo e discutendone la soluzione, il paradosso di Newcomb, un rompicapo logico-probabilistico che contende la palma del più controverso a un altro gioco di probabilità, il problema delle tre porte o di Monty Hall [12].<sup>4</sup> Il riquadro 1 esamina parametricamente il problema in forma generale.

Il paradosso di Newcomb è stato ideato dal fisico William Newcomb (1960), studiato dal filosofo Robert Nozick [14] e diffuso da Martin Gardner [15], [16] nella sua popolare rubrica di matematica ricreativa su *Scientific American* e, in italiano, su *Le Scienze*.<sup>5</sup>

<sup>3</sup> Il riferimento [11] fornisce una rassegna completa e aggiornata degli studi di Keith E. Stanovich (Università di Toronto) e del suo gruppo di ricerca sulla distinzione tra razionalità e intelligenza, cioè sul concetto di *dysrationalia*, e su come le nostre vite ne siano influenzate.

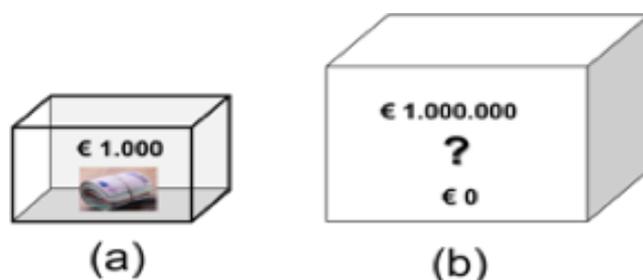
<sup>4</sup> Riguardo al gioco di Monty Hall, è interessante ricordare che la maggior parte delle persone, anche matematici di professione, arriva a convincersi della soluzione corretta molto lentamente, se mai ci arriva [4]. Invece, uno studio del 2010 ha dimostrato sperimentalmente che i piccioni apprendono la strategia che porta alla soluzione ottimale in un tempo decisamente minore del nostro [13]. Considerate le dimensioni del cervello di questi volatili, il fatto è davvero curioso.

<sup>5</sup> L'articolo [16] contiene le precisazioni di Nozick alla valanga di commenti ricevuti dalla rubrica di giochi tenuta da Gardner dopo la pubblicazione di [15]. Gardner ripropone e aggiorna con un addendum l'intera questione nel capitolo 44 della raccolta [17], un volume molto pregevole, anche dal punto di vista editoriale.

Adotteremo sia nel testo sia nelle illustrazioni un registro volutamente colloquiale, didascalico, ridondante, ancorché lessicalmente preciso, per facilitare la comprensione del problema e delle sue implicazioni. Le sottigliezze del ragionamento sono soprattutto logiche, non di tipo computazionale; i calcoli (probabilistici) sono, infatti, di algebra elementare e alla portata di tutti.

Partiremo da:

**Il problema in forma semplificata.** In bella mostra su un tavolo ci sono due scatole: una piccola trasparente (da qui in poi “la scatola”), della quale si può vedere il contenuto, e una grande opaca (“lo scatolone”), il cui interno è invisibile (cfr. figura 1).



**Figura 1**

**Paradosso di Newcomb:** (a) la scatola trasparente di sinistra contiene con certezza € 1.000; (b) lo scatolone opaco di destra contiene € 1.000.000 oppure è vuoto

Un oracolo o indovino chiaroveggente – che potremmo anche identificare, in omaggio allo sviluppo dell'IA, con un algoritmo di *machine learning* o con un *brain scanning computer* – ti propone uno strano gioco con premi in denaro.<sup>6</sup> (Come valuta di riferimento, anziché il dollaro, utilizzeremo l'euro). L'algoritmo mette: 1) € 1.000 nella scatola e € 0 nello scatolone, oppure 2) € 1.000 nella scatola e € 1.000.000 nello scatolone. A questo punto, se tu potessi scegliere tra: A) prendere il denaro solo dallo scatolone oppure B) prendere il denaro da entrambe le scatole, non avresti certamente dubbi. Infatti, scegliere entrambe le scatole (opzione B) è la scelta economicamente più vantaggiosa, dal momento che garantirebbe una vincita almeno uguale a quella dell'azione A (€ 1.000 sicuri) e potenzialmente maggiore (€ 1.001.000).

**Il paradosso di Newcomb in forma originale.** La questione si complica – e non di poco! – perché aggiungiamo la condizione che l'algoritmo faccia una predizione prima che tu scelga: se pensa che sceglierai l'opzione A (solo lo scatolone), mette in esso € 1.000.000; se pensa che sceglierai B (entrambe le scatole), depone € 0 nello scatolone, cioè niente. Analogamente, lo scatolone risulterà vuoto (€ 0) se l'oracolo si aspetta che tu lanci una moneta per rendere aleatoria la scelta. Nella scatola ci sono, qualunque sia la tua azione, € 1.000.

<sup>6</sup> Cicerone sosteneva che “Un aruspice non può incontrare un altro aruspice senza ridere”, ma non poteva immaginare che sarebbe sopraggiunta l'era dei big data e del machine learning. Per la verità, parecchi studiosi – però non tutti – mettono in dubbio le capacità predittive dei computer, per quanto potenti essi siano (cfr., per es., [18]).

Punto-chiave è che l'oracolo fa la sua predizione prima della tua scelta. Si suppone inoltre che l'algoritmo, essendo dotato di rimarchevole intelligenza cognitiva, sia quasi sempre corretto nelle sue previsioni, anche se le sue capacità predittive, o divinatorie, non sono assolute. In altri termini, "ci azzecca" il più delle volte, benché non sempre: si ricordi anche la nota 6.

L'ultima colonna della tabella 1 riassume le quattro possibili vincite in relazioni alle due previsioni dell'oracolo e alle tue due possibilità di scelta.

PREVISIONE DELL'ORACOLO	TUA SCELTA	SCATOLONE	SCATOLA	VINCITA TOTALE
Solo scatolone	Solo scatolone	€ 1.000.000	€ 0	€ 1.000.000
Solo scatolone	Scatola e scatolone	€ 1.000.000	€ 1.000	€ 1.001.000
Scatola e scatolone	Solo scatolone	€ 0	€ 0	€ 0
Scatola e scatolone	Scatola e scatolone	€ 0	€ 1.000	€ 1.000

**Tabella 1**

*Tabella delle vincite nel paradosso di Newcomb. (Nota: nella terza e quarta colonna, per i casi in cui la scatola o lo scatolone non vengono scelti le corrispondenti vincite sono ovviamente € 0)*

La tabella 2 fornisce le stesse informazioni della tabella 1, però sotto forma di matrice delle vincite. Anche qui, i quattro casi possibili risultano dalla combinazione delle due predizioni dell'oracolo e dalle tue due opzioni di scelta.

LA TUA SCELTA	L'ORACOLO PREDICE A	L'ORACOLO PREDICE B
<b>A) PRENDI SOLO LO SCATOLONE</b>	Vinci € 1.000.000	Vinci € 0
<b>B) PRENDI ENTRAMBE LE SCATOLE</b>	Vinci € 1.001.000	Vinci € 1.000

**Tabella 2**

*Matrice delle vincite nel paradosso di Newcomb*

**La domanda-chiave.** Scegli l'opzione A (solo lo scatolone) o scegli l'opzione B (entrambe le scatole)?

Per (tentare di) sciogliere il dilemma in questo problema, indubbiamente fra i più complessi nel suo genere, si possono presentare due argomentazioni diverse le quali – benché entrambe logiche, solide e convincenti – portano a scelte opposte.

**Argomentazione a supporto della scelta A (solo lo scatolone).** Se scegli A, l'oracolo molto probabilmente lo ha previsto e ottieni € 1.000.000; se scegli B, l'oracolo quasi certamente lo ha previsto, hai perciò € 0 dallo scatolone e € 1.000 dalla scatola: in totale € 1.000. Pertanto la scelta più ovvia è A.

**Argomentazione a supporto della scelta B (entrambe le scatole).** Come sottolineato, l'oracolo ha già preso la sua decisione, qualunque essa sia e, quindi, il denaro – per l'ammontare conseguente alla decisione – è già nelle scatole. Se lo scatolone contiene € 0, ti conviene scegliere B (entrambe le scatole) perché ottieni € 1.000 invece di € 0. Se lo scatolone contiene € 1.000.000, oltre a questo ammontare, ottieni anche € 1.000 dalla scatola se scegli B. In entrambi i casi, per il principio di dominanza in teoria dei giochi, B risulta l'azione più saggia.

Un commento di Nozick sul quesito fu: “Quasi tutti pensano che il da farsi sia perfettamente chiaro e ovvio. La difficoltà è che gli esseri umani sembrano dividersi quasi esattamente a metà riguardo alla soluzione, e in buona parte pensano che quelli dell'altra metà siano sciocchi” [13, p. 117].

A riprova del fatto che il paradosso polarizza le risposte degli interpellati praticamente a metà, citiamo i dati di un sondaggio condotto nel novembre 2016 sul sito online del *Guardian* dal matematico e divulgatore Alex Bellos. Su quasi 32.000 risposte, i risultati sono stati il 53,5% a favore di A (scelta dello scatolone) e il 46,5% a favore di B (entrambe le scatole).<sup>7</sup> Bellos ha spiritosamente commentato che l'esito ricorda il referendum sulla Brexit, per il quale vi è stata una divisione anche all'interno di molte famiglie. Discutere e argomentare sul voto migliore (almeno non peggiore) non fa altro che radicalizzare le posizioni di ogni partecipante. Come, in genere, succede nelle reti e piattaforme sociali [19].

Per molti dei numerosi filosofi che se ne sono occupati, questo esperimento mentale riguarda questioni metafisiche e psicologiche, quali determinismo, fatalismo, predestinazione, libero arbitrio. Purtroppo, con ragionamenti di questo tipo si resta intrappolati in un labirinto di sconcertante ambiguità sul tempo e sulla causalità, il che rende il gioco dilemmatico e la scelta tra le opzioni A e B oltremodo difficile. Tant'è che qualche filosofo della scienza ha sostenuto che il paradosso mette in crisi la possibilità di effettuare una scelta completamente libera e razionale. Ma è veramente così? Fortunatamente, ci sono anche le discipline “dure”, e una di queste – il calcolo delle probabilità – consente di porre due domande precise: è possibile riformulare il problema in modo più chiaro? E se è così, esiste una soluzione non ambigua?

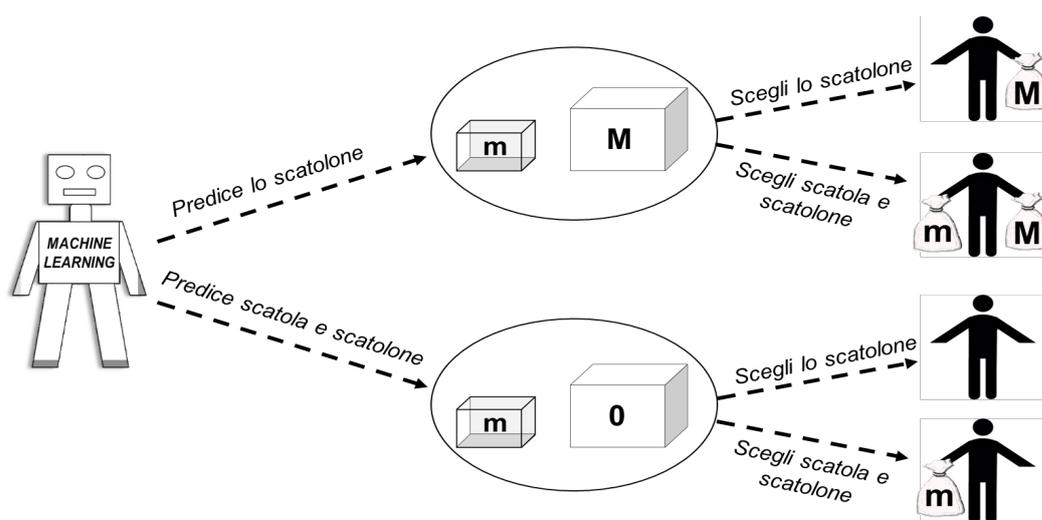
### 3 Il paradosso di Newcomb: la scelta probabilistica ottimale

In questo paragrafo, vedremo come e perché le risposte siano positive per entrambi i quesiti precedenti, facendo riferimento a una versione generalizzata, cioè parametrica, del paradosso.<sup>8</sup>

<sup>7</sup> Cfr. <https://www.theguardian.com/science/alexs-adventures-in-numberland/2016/nov/28/newcombs-problem-divides-philosophers-which-side-are-you-on>. In un'altra indagine di massa (<http://www.philosophyexperiments.com/newcomb/Default.aspx>), i risultati sono stati molto simili: il 55% è per la scelta del solo scatolone e il 45% a favore di tutte e due le scatole.

<sup>8</sup> L'impostazione che segue si ispira liberamente alla presentazione [20] di Norton J. Wildberger, matematico della University of New South Wales, Sidney.

**Generalizzazione del paradosso.** Il problema può essere reso più comprensibile definendo in modo generale ed esplicito le quantità in gioco:  $m$  = ammontare della scatola (€ 1.000 nella formulazione originale di Newcomb) e  $M$  = massimo ammontare dello scatolone (€ 1.000.000 nell'originale). Vedremo come il rapporto  $K = M/m$  ( $= 1.000$  nell'originale), o il suo reciproco  $k = 1/K = m/M$ , giochi un ruolo-chiave nella soluzione proposta. Con queste notazioni, il contenuto potenziale dello scatolone diventa  $M = mK$ . La generalizzazione del gioco è rappresentata in forma estesa nella figura 2.



**Figura 2**

Rappresentazione in forma estesa del paradosso di Newcomb generalizzato. (Figura adattata da [http://lesswrong.com/lw/gu1/decision\\_theory\\_faq/#what-about-newcombs-problem-and-alternative-decision-algorithms](http://lesswrong.com/lw/gu1/decision_theory_faq/#what-about-newcombs-problem-and-alternative-decision-algorithms))

Sono poi da considerare le due probabilità che rappresentano le capacità predittive dell'algoritmo:  $P(A)$  e  $P(B)$ .  $P(A)$  è, per definizione, la probabilità che l'algoritmo preveda correttamente la scelta quando essa è A (solo lo scatolone); mentre  $P(B)$  è, per definizione, la probabilità che l'algoritmo preveda correttamente la scelta se essa è B (entrambe le scatole). In generale, può essere  $P(A) \neq P(B)$ . È realistico sostenere che l'algoritmo – avendo sviluppato, forse grazie alla miriade di dati a sua disposizione, una metrica del tuo profilo e delle tue preferenze – sia in grado di stimare tanto  $P(A)$  quanto  $P(B)$  nella fascia di valori compresi tra il 50 e il 100%, certamente non inferiori al 50%.

**Criterio di scelta: massimizzazione del valore atteso.** Nei problemi di scelta in condizioni aleatorie, in particolare nei giochi d'azzardo, un criterio conveniente per decidere fra le opzioni possibili è la massimizzazione del valore

atteso (chiamato anche media ponderata o speranza matematica) fra gli eventi considerati:<sup>9</sup>

**Il valore atteso  $E(X)$  di un evento  $X$  aleatorio è il suo valor medio ponderato, cioè la somma di ogni possibile valore di  $X$  moltiplicato per la probabilità ad esso associata.**

Nel nostro caso – scelta, o decisione, tra  $A$  oppure  $B$  – il valore atteso dell'evento  $A$  o  $B$  si ottiene moltiplicando ciascuno dei due premi possibili per la corrispondente probabilità di vincita. Pertanto, i valori attesi  $E(A)$  e  $E(B)$  delle azioni alternative  $A$  e  $B$  risultano:

$$E(A) = (mK)P(A) + (0)[1 - P(A)] = mKP(A)$$

$$E(B) = mP(B) + (m + mK)[1 - P(B)] = m(K + 1) - P(B)(mK)$$

(Ricordiamo che in entrambe le relazioni  $mK = M$ ). Come criterio di scelta, poniamo la condizione  $E(A) \geq E(B)$ , cioè che l'azione  $A$  (scelta del solo scatolone) sia conveniente almeno tanto quanto la  $B$  (scelta di entrambe le scatole). Con pochi facili passaggi algebrici, si ottengono le relazioni-chiave del processo decisionale:

$$P(A) + P(B) \geq 1 + 1/K$$

e l'equivalente:

$$P(A) + P(B) \geq 1 + k$$

Ognuna delle due formule, dove le probabilità  $P(A)$  e  $P(B)$  e il rapporto  $K = M/m$  (o  $k = m/M$ ) compendiano tutti i casi possibili, fornisce la soluzione ottimale al paradosso di Newcomb. (Si noti che le entità dei premi  $M$  e  $m$  non intervengono singolarmente ma solo attraverso i loro rapporti  $K$  o  $k$ ). La scelta tra  $A$  e  $B$  è indifferente – dal punto di vista del criterio probabilistico adottato – se vale il segno di uguaglianza.

La versione originale del paradosso considera  $m = \text{€ } 1.000$  e  $M = \text{€ } 1.000.000$ , cioè  $K = 1.000$ ; supponendo altresì  $P(A) = P(B) = 0,9$ , abbiamo  $E(A) = \text{€ } 900.000$  e  $E(B) = \text{€ } 101.000$ , per cui  $E(A) > E(B)$ : infatti,  $P(A) + P(B) = 1,8$  è ben maggiore di  $1 + 1/K = 1 + k = 1,001$ . Pertanto:

**Il criterio di scelta basato sul valore atteso più conveniente fa preferire lo scatolone da solo (opzione A) rispetto a entrambe le scatole (opzione B).**

Per una discussione di altri casi significativi, al variare dei parametri in gioco, si rinvia al riquadro 1 e alla relativa figura 4. Si confermerà che per molti casi la scelta dell'opzione  $B$ , implicitamente basata su un ragionamento del tipo “se

<sup>9</sup> La massimizzazione del valore atteso è il criterio più semplice e immediato per scegliere fra le opzioni di un gioco con vincite aleatorie. Metodi più complessi per rappresentare numericamente l'attitudine (propensione o avversione) al rischio di un decisore sono basati sull'utilità attesa (von Neumann e Morgenstern) o sulla prospect theory (Kahneman e Tversky).

*proprio mi va male intasco € 1.000*", non corrisponde al valore atteso migliore né alla strategia più efficace. Esito che a questo punto della discussione dovrebbe risultare chiaro: si noti che  $E(A)$  è direttamente proporzionale a  $P(A)$ , mentre  $E(B)$  decresce linearmente con il crescere di  $P(B)$ . Infatti, il risultato prima evidenziato discende unicamente dalle due relazioni tra valori attesi e probabilità.

**Conclusioni.** Grazie alla precedente impostazione, puramente matematica, tanto il paradosso quanto la sua soluzione risultano ben definiti, e ogni considerazione filosofica, soprattutto metafisica, è riconducibile ai valori di probabilità  $P(A)$  e  $P(B)$ . Peraltro, non è difficile prevedere che molti – dilettanti appassionati o esperti – continueranno a sollevare obiezioni, se non a cercare correlazioni addirittura con il gatto di Schrödinger e l'*entanglement* (fenomeno della non-località) in fisica quantistica. Infatti, Gardner, dopo la pubblicazione su *Scientific American*, fu invaso da lettere bislacche di commento sulla soluzione, oltre che da proposte, più o meno bizzarre, di varianti al problema originale. Il fenomeno è diventato così vasto che è stato nominato "newcombmania".

Il lettore attento avrà anche notato l'analogia tra questo dilemma e quello del prigioniero [6]. In effetti, in letteratura si è sviluppato un ampio dibattito, più che altro filosofico, sulla questione:

*Al di là delle considerazioni sul piano della struttura, resta il fatto che, a differenza di quanto avvenuto con il dilemma del prigioniero, il problema di Newcomb non ha ricevuto grande interesse in campo applicativo, forse perché la fonte di attrazione verso il problema è di natura più spiccatamente teorico-speculativa che non pratica. Questo spiega anche la circostanza per cui, mentre il dilemma del prigioniero è stato oggetto di studio da parte di numerose discipline, la trattazione del problema di Newcomb è rimasta quasi confinata all'interno del dibattito filosofico [21].*

Tuttavia, il rompicapo illustrato rientra a pieno diritto nella teoria della decisione comportamentale (*behavioral decision theory*) perché compendia efficacemente una questione di base: come prendiamo le decisioni, specialmente quando guadagni, perdite, probabilità riguardano contesti e orizzonti temporali diversi in un quadro non ben definito. È perciò strano che il problema non sia considerato fra i temi d'interesse di questa recente branca dell'economia.

#### 4. Motivazioni e limiti dell'IA attuale

*Non siamo macchine pensanti che si emozionano, ma macchine emotive che pensano*  
(Antonio Damasio)

*[Danny Kahneman e io] più che di intelligenza artificiale, ci occupiamo di stupidità naturale* (Amos Tversky, [8, p. 309])

Ancora con l'obiettivo di affrontare aspetti cruciali di tecnoscienza usando, come strumenti decisionali, logica, ragione e pensiero critico, esamineremo in questo paragrafo e nel successivo il secondo caso di studio, quello dell'IA. Infatti, i problemi economici, etico-sociali, addirittura di politica fiscale, legati all'IA e alla robotica (umanoide e no) sono al centro dell'attenzione pubblica e oggetto di sempre più ampi e serrati dibattiti, praticamente su ogni mezzo mediatico o canale informativo. Per esempio, far sposare le imprese più avvedute con le nuove tecnologie è l'intrigante sfida dell'incombente Industria/Impresa 4.0, dove IA e robotica insieme possono agire da catalizzatori. Bisogna accettare che il futuro, per incrementare la produttività e assicurare il successo nei settori tecnico-economici, non è esclusivamente umano né esclusivamente artificiale: si trova in tutti e due. La parola-chiave è "arricchimento" o "potenziamento" (*augmentation*) che sfrutta a favore di entrambi la sinergia fra macchine e umani per un lavoro migliore, più intelligente e più veloce [22], [23].<sup>10</sup> Sono però urgenti grandi interventi di *re-skilling* e riqualificazione delle risorse umane, perché l'AI, in tutte le sue manifestazioni, è oggi un fenomeno sempre più dirompente e dilagante.

Non dimentichiamo che siamo complementari e simbiotici con gli utensili da millenni e con le macchine da secoli, in modo socialmente ed eticamente sostenibile: la padronanza di nuove tecnologie è andata di pari passo con l'evoluzione della nostra specie; inoltre, capacità cognitiva e arte tecnologica si sono arricchite e rafforzate a vicenda.

Dal momento che, negli scenari della complessità, le tecnologie, ogni giorno, ridisegnano e aprono nuovi orizzonti non predeterminati, siano essi di rischio o di opportunità, è bene ricordare l'aurea massima, di sapore quasi Zen, dello storico Melvin Kranzberg: "*La tecnologia non è né buona né cattiva; non è neanche neutrale*". Per cui, anche se gli esperti nelle professioni più nobili non saranno rimpiazzati dall'IA, è prevedibile che gli operatori specializzati che la usano sostituiranno quelli non lo fanno.

Se l'ausilio vicariante della tecnologia, soprattutto nei due secoli passati, era prevalentemente diretto alle attività fisiche e muscolari, oggi si sta progressivamente spostando verso la sostituzione delle cerebrali e mentali [25]. Certo è che ne sono stati fatti progressi da quando si diceva: "[...] *un computer è una macchina stupida con la capacità di fare cose incredibilmente intelligenti, mentre i programmatori sono persone intelligenti con la capacità di fare cose incredibilmente stupide. In breve, insieme creano un'accoppiata di pericolosa perfezione*" [26, p. 316]. In ogni caso, se l'IA e i suoi prodotti ci sostituiranno nei lavori che richiedono attività muscolari (i robot), evitiamo che sostituiscano del tutto il nostro cervello, semmai lo potenzino con l'*augmentation* prima citata. E facciamo in modo che la frase "*la tecnica è il prodotto di una scienza senza umanità, che dà soltanto meccanismi e strumenti*" resti una boutade da filosofi inguaribilmente antimodernisti e misoneisti.

L'essenza del rapporto uomo-macchina è distillata dal paradosso di Hans Moravec: "*la macchina trova difficile svolgere i lavori di routine dell'uomo, l'uomo*

<sup>10</sup> Per discussioni critiche sugli effetti dell'IA in generale, rinviamo, per esempio, a [19], [24] e alle bibliografie annesse.

*trova ardui i compiti tipici della macchina*". Infatti, la nostra specie, per ragioni biologiche e genetiche, ha facilità di movimento, di percezione, di inferenza euristica; al contrario, una macchina è costitutivamente portata a svolgere compiti di elaborazione algoritmica e deduzione logica nella risoluzione di problemi computazionali di larga scala. I sistemi artificiali non sono oggi in grado di affrontare da soli situazioni impreviste, tutte diverse, ma frequenti nel loro insieme: sono "casi di coda lunga e governati da una legge di potenza o di Pareto" [27].<sup>11</sup> Si tratta di episodi in cui i sistemi non sono (ancora) dotati di una base di conoscenza sufficientemente ampia per permettere loro di decidere come agire; tantomeno, hanno la capacità – cioè, il buon senso – di capire che cosa possa essere dannoso per l'uomo [28]. Quando si verificano questi casi, singolarmente rari, ma in totale molto numerosi, la macchina lavora fuori della propria zona di comfort; quindi, non può essere lasciata a operare in assoluta autonomia senza la responsabilità della supervisione umana. (Si veda anche la distinzione tra Sistema 1 e Sistema 2 – o pensieri veloci vs. pensieri lenti – nel modello di Daniel Kahneman [7]). Qui, al momento, sta il limite a una IA dotata di autocoscienza, cioè in grado di sostituire l'uomo in qualsiasi attività manuale, intellettuale ed emozionale. Oggi l'intelligenza artificiale emula sostanzialmente solo il comportamento razionale dell'uomo – o il Sistema 2, cioè la parte riflessiva dei nostri processi di elaborazione dell'informazione.

A proposito di sentimenti di macchine artificiali, è noto che la fantasia spesso anticipa e supera la realtà: già Hal 9000, il paranoico e malvolente supercomputer di *2001: Odissea nello spazio* (1968), era in grado di provare empatia cognitiva, sia pure imbibita di tutta l'umana maliziosa perfidia – per la quale non sarebbe azzardato proporre il neologismo "disempatia".<sup>12</sup> La fantasia realizzava dunque, cinquant'anni fa, ciò che oggi stanno cercando di fare algoritmi avanzati di IA: la lettura delle labbra, oltre che la comprensione del linguaggio e la capacità di ragionamento degli umani. C'è poi la situazione, nient'affatto irrealistica per l'Internet delle cose, in cui l'uomo perde il controllo di

<sup>11</sup> *Lavori fondamentali di Vilfredo Pareto mostrano che un'ampia categoria di distribuzioni statistiche segue una "legge di potenza", non una gaussiana. Una distribuzione con legge di potenza rappresenta casi in cui, insieme a molti eventi piccoli, coesistono pochi eventi straordinariamente grandi. La loro forma è caratterizzata da code molto più marcate (e lunghe) di quelle legate ai modelli gaussiani o esponenziali. Inoltre, spiega, matematicamente, eventi rari ma significativi – come, i super-ricchi, i terremoti di grande magnitudo, le guerre mondiali o i colossi tecnologici (Amazon, Apple, Facebook, Google, ecc.). In un fenomeno governato da un andamento di questo tipo, esistendo sempre un piccolo numero di dati estremamente lontani dalla media, ci si deve attendere che capitino casi anomali, gli ormai noti "cigni neri" di Nassim Nicholas Taleb. All'opposto, per ogni "Paperone" ci sono milioni di poveri, oppure per ogni terremoto catastrofico ci sono miriadi di movimenti tellurici di poco conto.*

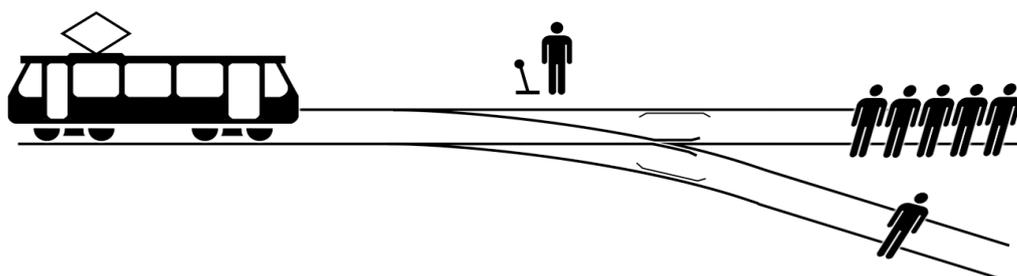
<sup>12</sup> *Nei dibattiti mediatici, risuona un po' stucchevole l'abuso del termine politicamente corretto di empatia – soprattutto, quella emozionale che, portando a commuoversi, è ben diversa dalla cognitiva – perché il mettersi nei panni dell'interlocutore, sempre e indistintamente, non permette di analizzare i problemi con il dovuto distacco per trovarne le soluzioni più logiche. Infatti, solo con la sospensione del giudizio, con l'analisi ragionata (il critical thinking), con ampiezza di vedute, ma tenendo a freno emotività e passioni è possibile discernere il vero dalle mistificazioni: si veda il caso della post-verità, trattato ampiamente in [19]. L'abuso di empatia (emozionale), portando a giustificare e non solo a comprendere, rischia di uniformare il pensiero verso il basso, ed è noto che senza entropia non c'è informazione. La scelta di rispettare tutti i punti di vista, di sforzarsi di vedere le buone ragioni dell'altro e di rifuggire la contrapposizione, non sempre e comunque, può essere conveniente. Al contrario, sviluppando con l'esercizio del pensiero razionale un acuto spirito critico, si evita di confondere i bravi (più capaci) con i buoni (più politicamente corretti), le soluzioni più opportune con le inadatte, le decisioni migliori con le abborracciate.*

macchine che comunicano tra loro con un linguaggio sviluppato autonomamente.

I progressi indotti dall'IA daranno vita a un'epoca di maggiore prosperità e benessere, di conoscenze senza precedenti, ma il transitorio a questo stato potrà essere lungo, anche aspro, se non sapremo adattare ad esso, fin da subito, economia, politiche sociali, comportamenti collettivi [18]. Nel seguito, per approfondire il problema dell'IA negli aspetti etici, partiamo dalle possibilità (rischi e opportunità) aperte dai veicoli a guida autonoma, variante moderna dal dilemma del carrello/vagone ferroviario o vettura tranviaria (*trolley problem*).

## 5. Il problema del trolley e i dilemmi etici dell'IA

Il saggio monografico [29] propone e discute ben nove varianti del problema del trolley. Nella versione classica, illustrata in figura 3, un uomo vicino a una leva di uno scambio è di fronte all'alternativa di: 1) *non fare niente* e lasciare che un trolley in corsa incontrollata travolga cinque persone; oppure 2) *azionare deliberatamente* lo scambio deviando il mezzo su un ramo secondario dove, c'è una sola persona, che rimarrebbe, purtroppo, uccisa. L'alternativa di carattere etico è dunque un solo morto contro cinque, ma l'unica vittima sarebbe a causa di un atto volontario e non di fatalità.



**Figura 3**

*Versione classica del dilemma del trolley: azionare o no lo scambio? (Da Wikipedia)*

Questa formulazione del dilemma, la cosiddetta *switch version*, fu proposta nel 1967 da Philippa Foot (poi Foot). Da allora sono stati pubblicati innumerevoli articoli di filosofia morale, sollevando così tante polemiche che all'intera questione è stato appiccicato l'ironico neologismo di "carrellologia" (*trolleyology*); sicché il problema con le sue innumerevoli varianti è diventato il paradigma più utilizzato nelle ricerche di psicologia morale e di *experimental philosophy*. Adattando il modello trolley della precedente figura ai veicoli con guida autonoma, l'articolo di *Nature* [30] riporta i risultati dell'esperimento Moral Machine, una piattaforma sperimentale online progettata dal Media Lab del MIT per esplorare i dilemmi morali che si incontrano in questo caso. In quattro anni, la piattaforma ha raccolto 40 milioni di pareri di persone da 233 Paesi, che hanno espresso priorità etiche diverse in relazione alle

proprie diversità culturali (per dettagliate analisi statistiche dei risultati si rinvia a [30], [31]).

Il filosofo dell'informazione Luciano Floridi (Università di Oxford), in pochi minuti di brillanti e serrate argomentazioni durante una lezione al Politecnico di Torino [32], riesce a dissacrare tutta l'aura metafisica del problema dimostrandone la fatuità. Egli osserva che l'alternativa posta conduce sempre a perdere, qualsiasi scelta si faccia, perché non si prendono in considerazione altre possibili soluzioni ingegneristiche (o tecniche) che consentirebbero di ampliare il ventaglio delle possibilità in modo da rilasciare i troppo stringenti vincoli originali. Nel caso specifico dell'auto, per esempio, si tratta di usare i freni di emergenza o il clacson, di sterzare, di spegnere il motore, ecc. La soluzione filosofica non c'è, la soluzione ingegneristica sì, e consente di cambiare una circostanza di gioco *lose-lose* (tutti i giocatori sono perdenti dal punto di vista morale) in una *win-win* (in pratica, tutti i giocatori vincono). Si potrebbe, peraltro, ribattere che la questione riguarda una scelta morale, non un problema tecnico, perché possono capitare situazioni in cui la tecnica non ha risposte, mentre la questione morale resta. Questo potrebbe essere il caso in cui una barca, pur affondando, si sta avvicinando alla costa, essendo spinta dalla corrente. Che cosa fare? Per salvare tutti gli altri, si butta a mare l'uomo più grasso a bordo? Sembra quasi che non ci sia via d'uscita, ma non è così.

L'etica delle macchine riguarda la questione di come i sistemi autonomi possano essere permeati di valori etici. Per una panoramica delle discussioni sui dilemmi etici sollevati dall'IA con proposte di soluzione, si suggeriscono gli aggiornati fascicoli tematici [33]-[35] di riviste scientifiche. Sistemi autonomi dotati di eticità – nei fini e nei comportamenti – sono necessari perché, inevitabilmente i sistemi del prossimo futuro saranno agenti morali. Siano automobili senza guidatore o diagnosi mediche basate su IA, questi sistemi dovranno fare scelte con norme sociali e conseguenze etiche. Sulle motivazioni per cui i progettisti hardware e software devono incominciare a preoccuparsi di includere aspetti etici nelle loro realizzazioni si rinvia agli articoli-saggio [36], [37] (da queste motivazioni nasce anche il fascicolo monografico [35]). Poiché l'IA si rapporta con le istituzioni sociali, i lavori [35] analizzano le implicazioni etiche in importanti settori applicativi socialmente rilevanti, quali l'istruzione, l'assistenza sanitaria e la cura degli anziani, i veicoli a guida autonoma. La loro *governance* tramite standard e processi trasparenti è necessaria per realizzare la piena sicurezza dei sistemi di AI in modo da conseguire la fiducia di buona parte dell'opinione pubblica.

Prendiamo un esempio concreto: quello delle diagnosi mediche o dei test giudiziari, che coinvolgono l'incertezza probabilistica [4], [5]. Sappiamo che ogni stima o sistema di classificazione produce sia falsi positivi sia falsi negativi. Anche se questo semplice concetto è ben noto agli esperti del campo, non è facile fornire una regola certa per determinare i costi relativi associati ai due errori. Peraltro, un sistema che operi, per esempio, nel campo della diagnostica medica, deve funzionare con numeri realistici e affidabili. La scelta dei valori e dei costi relativi riflette una decisione etica del progettista o del sistema progettato? Se riflette il progetto, la macchina può essere considerata etica? È chiaro che, come gli esseri umani, la macchine dovranno apprendere l'etica

adattandosi al contesto o alla situazione, al tempo, cioè dovranno essere adattative. Un'altra domanda inevitabile è: *“Qual è la gamma di situazioni e quali sono i parametri da considerare per l'adattatività?”* [36]. Poiché è improbabile che, nel prossimo futuro, la macchina sia in grado di trattare una gran varietà di situazioni diverse, si ritorna alla zona di comfort e alla necessità della supervisione umana, cui abbiamo accennato prima.

Una prima bozza di linee e principi guida di codice etico per l'IA è stata pubblicata dall'UE [38], mentre le raccomandazioni per una “Good AI Society” del gruppo AI4People sono disponibili in [39]. Lavori di questo livello testimoniano l'attenzione che l'Europa sta dedicando a un settore cruciale, cioè alla consapevolezza per un'etica dell'intelligenza artificiale [40].

In definitiva, Floridi non ha tutti i torti a cercare di smontare il dilemma del carrello, ma gli studiosi di discipline scientifiche (STEM – *Science, Technology, Engineering and Mathematics*) e i filosofi devono lavorare per mettersi d'accordo su quali siano le regole morali degli esseri umani, prima di poter ridurre le considerazioni morali ad algoritmi. E con i nuovi sistemi tecnologici cognitivi, non sembra più utopistico realizzare automobili totalmente autonome con l'ambizioso obiettivo di azzerare (non solo minimizzare) i morti in incidenti stradali.

In campo militare, l'IA con i suoi algoritmi e processori rende già possibile la logistica dei mezzi di guerra in modo da farli convergere sul campo di battaglia in tempi rapidissimi, e si possono immaginare i rischi del passaggio alla guerra ibrida o interamente basata su mezzi informatici. L'esperto di IA Pedro Domingos prevede addirittura: *“Se mai scoppierà una cyberguerra, i generali saranno esseri umani, ma la fanteria sarà formata da algoritmi. Gli esseri umani sono troppo lenti e poco numerosi, e verrebbero spazzati via da un esercito di bot. Abbiamo bisogno di un nostro esercito di bot, e il machine learning sarà la loro accademia, la nuova West Point”* [41, p. 43]. Nell'inquietante scenario di una guerra robotica, le macchine saranno dotate di una crescente autonomia, ma, auspicabilmente, sotto la supervisione degli esseri umani, che dovranno quindi assumersi ancor più la responsabilità (*accountability*) di un comportamento etico.

## 6. Conclusione: alleanza tra humanities e scienze

*Se pensi all'anno prossimo, semina il granturco/ Se pensi ai prossimi 10 anni, pianta un albero/ Se pensi ai prossimi 100 anni, istruisci le persone*  
(Detto cinese)

*L'immaginazione è più importante della conoscenza. La conoscenza è limitata, mentre l'immaginazione abbraccia il mondo, stimolando il progresso e facendo nascere l'evoluzione*  
(Albert Einstein)

Le considerazioni sviluppate nel corso dell'articolo sul paradosso di Newcomb e sui problemi etici dell'IA fanno comprendere l'importanza di un approccio interdisciplinare basato sulle due culture su cui si fondano le società avanzate:

l'umanistico-letteraria e la tecnoscientifica. Non sembra perciò irrilevante cercare di cogliere, in queste osservazioni conclusive, le implicazioni positive, se l'interdisciplinarietà è percepita in modo costruttivo, negative, in caso contrario.

Un Paese, come il nostro, non cresce se non è consapevole che istruzione e formazione, ricerca, sviluppo e innovazione, brevetti e creatività in scienza e tecnologia sono i veri acceleratori del suo sviluppo qualitativo ed economico [42]. E bisognerà pure che qualità come talento, capacità, merito e professionalità tornino a essere premianti al posto dell'appartenenza. L'alleanza tra tecnoscienza e società, ancorché ineludibile, non può avvenire immediatamente, per il fatto che storicamente è una novità. Gli ostacoli, che si frappongono, non finiscono mai; oggi se ne deve registrare uno che di tanto in tanto riemerge, come un fiume carsico, nella politica e nei suoi apparati. Il sociologo Massimo Panarari [43] chiama "postmodernismo radicale":

*[Una] vera e propria ideologia antiscientifica (e anti-illuministica) [per cui] la scienza rappresenta una narrazione come le altre [...], il metodo sperimentale viene ricondotto a poco più di un discorso, e non si riconosce il fondamento obiettivo e oggettivo dei risultati che ne derivano (al punto che anche i numeri e le statistiche, insieme a coloro che li studiano e producono, sono fatti oggetto di virulenta polemica politica). Scienza e tecnica vengono considerate poco più che "ideologie" [...].*

Poiché scienza e politica agiscono su due piani diversi, oltre a non fare un uso strumentalmente politico della scienza, sarebbe importante che le decisioni strategiche attinenti alla scienza venissero prese sulla base di prove di efficacia scientifiche, sperimentalmente dimostrate e metodologicamente fondate. La morale di tutto ciò? È che c'è ancora molto da fare per colmare il divario tra la scarsa considerazione di cui gode la tecnoscienza rispetto alle discipline umanistiche, o per evitare che l'opinione pubblica italiana continui a subire il fascino di spiegazioni pseudoscientifiche spinte da un'epidemia senza precedenti di false notizie e verità alternative [44]. L'antidoto principale è partire sempre dalla ricerca dei fatti di cui sia accertata la verità, non dai media [45], con l'obiettivo minimo, eppure necessario, che tecnoscienza e società comunichino di più e, soprattutto, si capiscano meglio.

Chi pratica scienza e tecnica – perché, oltre a pensare, sa anche agire – ha il compito di esercitare un ruolo sia educativo sia sociale, da attore pubblico se non da *policymaker*; quindi, non può essere osservatore passivo del corso degli eventi. In questo quadro articolato e complesso, coloro che si occupano di scienza e tecnica, non solo dal punto di vista speculativo, mettono a frutto il loro bagaglio di conoscenze e professionalità, di esperienze e competenze nel progettare nuovi sistemi. Tuttavia, la dimensione progettuale da sola non è sufficiente; occorre, infatti, che gli esperti, disponendo di *skill* tanto operativi quanto intellettivi, dimostrino di saper intervenire nella società per interagire proficuamente con essa. E le imprese devono sostenerli, facendo la loro parte in modo attivo e consapevole [46].

In questo quadro, la preparazione di ricercatori, ingegneri, tecnici diventa fondamentale. Potremmo parafrasare un detto tradizionale con la sentenza: “*uno scienziato conosce ogni cosa di qualche cosa e qualche cosa di ogni cosa*”, pur con la consapevolezza che l'istruzione e le credenziali in un settore non garantiscono la competenza in ogni area. Con la sottolineatura degli obiettivi formativi, si chiude il cerchio; infatti, le aziende high tech richiedono oggi risorse umane altamente qualificate.

Poiché viviamo in un mondo che cambia vorticosamente, per operare efficacemente nella nuova realtà globale una preparazione aperta e flessibile deve essere tanto profonda in verticale (in senso specialistico) quanto estesa in orizzontale (in senso relazionale), cioè “a forma di T”. Negli USA, le facoltà STEM più all'avanguardia propongono questo modello di istruzione e formazione (*new education*), non più vincolato a professionalità, know-how ed esperienze esclusivamente settoriali. Istruzione, conoscenza e competenza costituiscono il punto di partenza, non d'arrivo, di un processo che deve portare ad apprendere ininterrottamente per tutta la vita e nel quale esperienza e professionalità aumentano progressivamente il loro peso. È bene ricordare che è finita l'era delle carriere continue e strutturate basate su un'unica competenza specifica. Specializzarsi è dunque necessario, ma non più sufficiente: all'inizio degli anni Settanta, lo scrittore di fantascienza Robert Heinlein coniò una massima per cui “*la specializzazione va bene per gli insetti*”. Se una preparazione riconosciuta, specifica e approfondita, oggi, infastidisce molti, uno dei motivi è che la specializzazione non può che essere elitaria e, quindi, esclusiva [47]. A proposito di culto dell'ignoranza, per gli Stati Uniti Isaac Asimov commentava tempo fa: “*Una vena di antintellettualismo si è insinuata nei gangli vitali della nostra politica e cultura, alimentata dalla falsa nozione che democrazia significhi «la mia ignoranza vale tanto quanto la tua conoscenza».*” [44, p. 191]. La stessa osservazione è valida tuttora, a maggior ragione per noi.

Per quanto attiene alla sfera della riflessione sociale, il rapporto con l'opinione pubblica da parte di studiosi ed esperti implica una loro capacità di divulgazione, comunicazione e rappresentazione dei temi moralmente sensibili. Cioè, una comunicazione strategica, persuasiva e performativa, ma non resa univoca e dogmatica, o investita da valenze normative. La credibilità e la qualità sono dati essenziali perché l'informazione scientifica non sia recepita dal pubblico come atto censorio, bensì sia accolta come momento di presa di consapevolezza. Un non secondario compito di tecnici, ricercatori e scienziati è dunque saper comunicare in modo dialettico, cioè dialogare – nei luoghi deputati – sui temi indicati mediante una lingua comprensibile anche ai non specialisti, senza ostentare forme di *hybris* tecnologica. La consapevolezza del ruolo e la conseguente assunzione di responsabilità da parte di chi detiene l'*expertise* (la competenza tecnoscientifica) sono leve fondamentali per superare la frammentazione di senso e di obiettivi tra protagonisti e decisori diversi.

Secondo quanto esposto in [48], otto sono le arti liberali che concorrono al successo persino negli investimenti finanziari (nell'ordine: fisica, biologia, sociologia, psicologia, filosofia, letteratura, matematica e teoria delle decisioni);

la riuscita in campo azionario o, più in generale, negli affari dipende da un mix di questi fattori. La interdisciplinarietà, associata alla passione in più aree del sapere, corrisponde a ciò che Charlie Munger – sodale e partner di Warren Buffett – chiama un reticolo di idee (*latticework of ideas*), che integra all'interno di una cornice unitaria la "saggezza" frutto delle conoscenze citate (*worldly wisdom*).<sup>13</sup> Certamente significativa e sorprendente è questa convergenza di realtà tradizionalmente considerate distinte e non comunicanti in un settore pratico, se non utilitaristico, come quello finanziario. Ricordiamo, peraltro, che il colto ed eclettico John Maynard Keynes fu, oltre che leggendario economista, un brillante amministratore di beni pubblici e privati grazie a investimenti oculatamente selezionati in campo finanziario.

Che un leader di successo in un qualsiasi settore debba avere un bagaglio di conoscenze provenienti dalle due culture è pure conseguenza delle tesi discusse dal fisico e romanziere inglese Charles P. Snow in un saggio di quasi sessant'anni fa. Non sorprende che questo lavoro di cento dense paginette sia stato snobbato, spesso stroncato, da umanisti e letterati influenzati dal solito pregiudizio antiscientifico. Meritoriamente, il saggio è stato riproposto in italiano da Alessandro Lanni, arricchito dagli interventi di esponenti culturali primari quali Giulio Giorello, Giuseppe O. Longo e Piergiorgio Odifreddi [49].

Come tanti altri studiosi attenti all'innovazione tecnologica, crediamo che i Paesi, le imprese, i leader, i giovani talenti che sanno sfruttare al meglio le opportunità della rivoluzione digitale, controllandone allo stesso tempo i rischi connessi, abbiano in mano il futuro. Con tutte le attenzioni e cautele che sapremo mettere in atto eviteremo il rischio di trasformarci da *Homo sapiens sapiens* in *insipiens digitalis*. In ogni caso, la riflessione critica e razionale – la capacità, cioè, di esaminare nuove informazioni e idee in modo spassionato, logico e senza pregiudizi o preconcetti emotivi – non deve e non può significare critica fine a sé stessa [47]. Senza dimenticare che l'*hybris* nella tecnoscienza [50] e nell'economia [51] è fonte di errori e fallimenti memorabili, dai quali, peraltro, si può imparare con l'umiltà. "Conoscere per deliberare" è il quanto mai attuale ammonimento di Luigi Einaudi. Conoscenza da intendere, come già ricordato in [6], quale pratica collettiva del pensiero critico.

In questo articolo, come nei precedenti [4]-[6], abbiamo fatto largo uso di enigmi e rompicapi logico-matematici, non solo per introdurre una nota più divertente e leggera, ma anche, perché essi consentono di analizzare le ragioni che impediscono alle persone di trovare *immediatamente* la soluzione corretta, anche se *a posteriori* la risposta per molti di essi risulta ovvia (cfr. [7], [52] e riquadro 3).

L'obiettivo principale, che si spera conseguito almeno in parte, è stato di mostrare come i due saperi – l'umanistico e lo scientifico – essendo complementari, se non del tutto consustanziali, possano fruttuosamente contaminarsi interagendo, pur nel rispetto dell'autonomia degli statuti

<sup>13</sup> La metafora del reticolo (o graticcio), essendo rappresentativa delle molteplici e interconnesse discipline afferenti alle due culture, sembra più calzante di quella comunemente adottata per visualizzare linearmente lo stesso concetto mediante l'analogia del binario, ossia una coppia di rotaie connesse da traversine (le relazioni).

epistemologici e dell'indagine metodologica, che competono alle loro singole discipline costituenti. Si spera di non dover più assistere in futuro a posizioni, prive di riscontri fattuali e logicamente non motivate, da parte di inguaribili esponenti di una tecnofobia distopica e passatista, i quali sostengono che la tecnoscienza non solo non crea cultura, ma addirittura che è, sempre e in ogni caso, dannosa per l'uomo. È proprio questa mentalità che ci preclude di investire in istruzione, innovazione tecnologica e alfabetizzazione scientifica, come e quanto sarebbe necessario, se prendessimo a modello i Paesi più avanzati.

In ogni caso, bisogna tenere in conto che, per tornare – certo senza pretese di ecumenismo – a una visione organica, strutturale, sistemica e plurale della cultura, come quella dell'*Encyclopédie*, ci vorranno ancora molto tempo e intenso lavoro, dedicati a erigere ponti, non steccati o muri.

### Riquadro 1 – Riesame del paradosso di Newcomb generalizzato

La discussione dei casi particolari chiarisce l'utilità delle formule di scelta:  $P(A) + P(B) \geq 1 + 1/K$  o l'equivalente  $P(A) + P(B) \geq 1 + k$  (cfr. par. 3) e ne dimostra la duttilità. Permette, inoltre, di acquisire dimestichezza sulla convenienza di una scelta anziché l'altra. A questo scopo, vedremo che la figura 4 mostra le aree di convenienza delle due scelte A e B nel piano cartesiano costituito da  $k = m/M$  (in ascisse) e  $P(A) + P(B)$  (in ordinate). Entrambe le scale sono lineari.

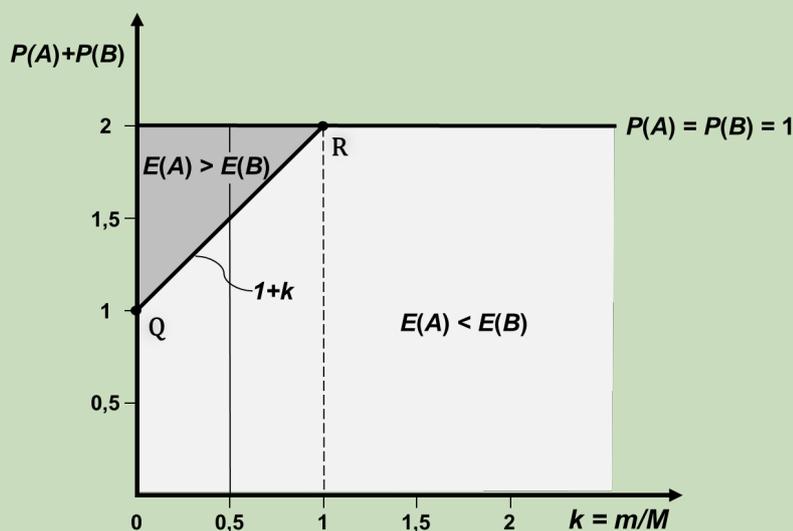


Figura 4

Paradosso di Newcomb generalizzato e aree di convenienza in funzione di  $k$ : nella grigia conviene la scelta A (solo lo scatolone); nella bianca conviene la scelta B (entrambe le scatole). Sulla frontiera QR la scelta è indifferente

Nel corpo del testo, abbiamo già visto che, per massimizzare il valore atteso, la scelta migliore è lo scatolone da solo (azione A) con i valori del problema originale:  $m = € 1.000$  e  $M = € 1.000.000$ , cioè  $K = 1.000$ , e supponendo  $P(A) = P(B) = 0,9$ .

Più in generale, se  $K = 1.000$  e se  $P(A) + P(B) \geq 1,001$  l'opzione migliore è la A (solo lo scatolone); infatti, purché l'oracolo faccia meglio, anche di poco, della pura casualità,  $P(A)$  e  $P(B)$  sono entrambe maggiori di 0,5. Se invece  $P(A) = P(B) = 0,5$ , la scelta ottimale diventa B (entrambe le scatole) perché  $1 < 1,001$ . Se, più in generale,  $P(A) = P(B) = p$ , deve essere  $p \geq 0,5 + 1/(2K)$  o  $p \geq 0,5 + 0,5k$  perché la scelta sia solo lo scatolone (opzione A).

Se la vincita rimane € 1.000 per la scatola, ma scende a € 2.000 (potenziali) per lo scatolone, risulta  $K = 2$  ( $k = 0,5$ ) e  $1 + 1/K = 1 + k = 1,5$ . Pertanto si richiedono due probabilità abbastanza elevate nel totale  $P(A) + P(B)$  per decidere di scegliere A (solo lo scatolone), cioè occorre essere convinti che l'oracolo sia piuttosto efficace nelle sue previsioni (cfr. nota 6).

È infine interessante il caso-limite in cui l'algoritmo azzeccchi tutte le previsioni:  $P(A) = P(B) = 1$ . La convenienza di scegliere A (solo lo scatolone) implica  $K \geq 1$  (o  $k \leq 1$ ), cioè  $M \geq m$ . La controprova è che, essendo  $E(A) = mK$  e  $E(B) = m$ , si ha  $E(A) \geq E(B)$  se  $K \geq 1$  ovvero  $k \leq 1$ . Nel caso opposto ( $K < 1$  o  $M < m$ ), la scelta migliore diventa la B, in quanto  $E(A) < E(B)$ .<sup>14</sup> In particolare, se anche  $K = k = 1$ , la scelta è indifferente dal punto di vista del valore atteso:  $E(A) = E(B)$ .

Tutte queste combinazioni sono riassunte nella citata figura 4. Nell'area grigia conviene solo lo scatolone (scelta A); nella zona bianca è preferibile la scelta B (entrambe le scatole). La spezzata che delimita la frontiera è tale poiché ciascuna delle due probabilità  $P(A)$  e  $P(B)$  non può superare il valore 1. Il segmento QR rappresenta quindi l'insieme dei punti di pareggio (*break-even*) tra le due scelte.

Notare che la figura può risultare ingannevole, perché il valore della sua ascissa  $k$  non è, teoricamente, limitato. In pratica, per come abbiamo posto e discusso il problema,  $k$  è da intendere molto piccolo: nell'esempio classico è  $1/1.000 = 0,001$ . Quindi, il triangolo grigio (la scelta A è migliore della B) risulta delimitato da valori alquanto realistici di  $k$  (o  $K$ ),  $P(A)$  e  $P(B)$ , ed è assai più significativo dell'area bianca (dove B è preferibile ad A), che è, ma solo graficamente, non limitata. In conclusione, come per il caso particolare descritto nel par. 3, è confermato che:

<sup>14</sup> La figura 4, mostra chiaramente, che a partire da  $k = 1$  in su, la scelta B (entrambe le scatole) è sempre migliore rispetto alla A (solo lo scatolone). Infatti, se l'entità del premio nella scatola piccola diventa più alta di quella dello scatolone, la convenienza della scelta si sposta, anche intuitivamente, verso la B, in quanto l'ammontare dello scatolone incide sempre meno sul guadagno complessivo. Questo scenario, pur matematicamente corretto, non corrisponde però allo spirito originale del problema, che prevede uno sbilanciamento netto tra i due premi a favore di quello nello scatolone, cioè  $k$  molto piccolo (o  $K$  molto grande).

**Per valori significativi di  $k$  (o  $K$ ),  $P(A)$  e  $P(B)$ , cioè nell'area grigia della figura 4, il criterio di scelta del valore atteso più conveniente fa preferire la scelta dello scatolone da solo (opzione A) rispetto a entrambe le scatole (opzione B).**

Un altro approccio per arrivare agli stessi risultati è di analizzare le due relazioni (lineari) di  $E(A)$  ed  $E(B)$ , ricavate nel paragrafo 3:

$$E(A) = mKP(A)$$

$$E(B) = m(K + 1) - P(B)(mK)$$

o, più semplicemente i loro valori normalizzati al parametro  $mK$ . Lasciamo al lettore lo studio degli andamenti di queste rette sul piano cartesiano, che definiscono senza ambiguità, all'unico punto di incrocio, le aree di convenienza per le opzioni A e B.

## **Riquadro 2 – I dilemmi morali di Churchill durante la Seconda guerra mondiale**

A partire dall'agosto 1941, con la decrittazione di Enigma a Bletchley Park [53], gli inglesi erano in grado di vedere quasi tutte le carte dell'avversario e fare, quindi, un uso militare delle informazioni segretamente intercettate. Il personaggio di Alan Turing nel film *The Imitation Game* dichiara di poter sviluppare con i colleghi "un sistema per determinare ogni quante informazioni agire, quale attacco fermare, quale lasciar correre. Le statistiche, il numero minimo di azioni per vincere la guerra, ma il numero massimo da intraprendere perché i tedeschi non sospettino mai". In effetti, il servizio MI6 dell'intelligence britannica poté usufruire di un criterio metodologicamente fondato per decidere, giorno per giorno, quali obiettivi decrittati colpire e quali no, senza svelare le proprie mosse al nemico. Naturalmente, il premier Churchill, costantemente tenuto al corrente, era il solo ad avere la responsabilità militare delle decisioni da prendere nei singoli casi.

Mentre a Bletchley Park si parlava di giochi, sull'Atlantico i britannici impiegavano la stessa logica dell'ottimizzazione nei combattimenti contro i sommergibili tedeschi. Possiamo dire che i criteri di ottimizzazione strategica – fra cui il MinMax e il MaxMin [6] – della disciplina che va sotto il nome di ricerca operativa costituiscono una buona, se non la migliore, matematizzazione del pragmatismo di matrice angloamericana.

Tre anni dopo, Churchill si sarebbe trovato di fronte a un altro dilemma morale, considerato l'origine del problema del trolley (cfr. par. 5). Secondo una narrazione aneddotica ripresa da [29, cap. 1], il dilemma trae origine da un episodio reale avvenuto durante la seconda guerra mondiale:

*Alle 4:13 del 13 giugno 1944, Londra fu colpita dalla prima bomba volante, denominata dai nazisti V1. Le V1 erano una sorta di missile lanciate da basi poste nella Francia occupata. Alla prima V1 ne seguirono centinaia che colpirono Londra con effetti devastanti. Erano ordigni potenti ma imprecisi e i tedeschi avevano bisogno di sapere dove cadevano le V1, per adeguarne la traiettoria e colpire il centro di Londra, al fine di causare il maggior numero di vittime e danni. Gli inglesi, tramite agenti segreti che facevano il doppio gioco erano in grado di passare ai nazisti informazioni false sui reali luoghi di caduta delle V1, quindi potevano far modificare le traiettorie delle bombe volanti in modo da salvaguardare le zone centrali di Londra, più abitate e frequentate, a discapito di aree più periferiche e meno affollate della città. [...] La decisione ultima spettava a Churchill, il quale, già in altre occasioni, aveva dimostrato di non avere molte remore a prendere decisioni drammatiche [per esempio, quattro anni prima in occasione dell'Operazione Dynamo per l'evacuazione da Dunkerque della truppe inglesi e di parte delle francesi (N.d.A.)]. I servizi segreti inglesi fecero arrivare ai nemici informazioni false per depistarli. L'operazione ebbe successo e le V1 cominciarono a cadere sui quartieri meridionali di Londra. Furono salvate 10.000 vite ma, dopo la guerra, non fu mai data grande pubblicità alla notizia. (Adattamento da [54]).*

Solo negli ultimi decenni, sia pur parzialmente e con molta cautela, l'intelligence britannica ha consentito che qualche tassello della storia dell'operazione Ultra, compresa la decrittazione di Enigma, fosse divulgato [53].

### Riquadro 3 – Un enigma di logica

Questo intrigante quesito logico (cfr., per es., [11, pp. 70-71]) – molto popolare tra gli informatici e gli psicologi dell'Università di Toronto – costituisce un'efficace sintesi di come, con l'esercizio della ragione e del pensiero critico, sia possibile superare errori di valutazione. Il problema, benché lineare e formulato senza alcuna malizia sottesa (cioè, nessuna informazione è fasulla o mascherata, né la descrizione è artatamente ingannevole), riceve dalla maggioranza delle persone interpellate una risposta intuitiva, ma non corretta. La soluzione appare ovvia solo *ex post*.

**Formulazione.** Ci sono tre persone: Jack, Anna e George. Jack sta guardando Anna, mentre Anna sta guardando George. Sappiamo che Jack è sposato e George no. La descrizione della scena e del suo contesto finisce qui.

**Domanda.** Sapete dire se, fra le tre persone, ce n'è una sposata che sta guardando un'altra non sposata? Ossia, qual è la risposta corretta fra le tre

possibili: A) sì, c'è una persona sposata che sta guardando una persona non sposata; B) no, non ce ne è nessuna; C) non si può stabilire con certezza?

Non sapendo se Anna sia sposata o no, la maggioranza degli interpellati sceglie istintivamente la C) non si può stabilire con certezza. E voi cosa ne dite? Pensateci su con calma, prima di leggere la:

**Soluzione.** Per trovare la risposta corretta, basta considerare lo stato civile di Anna che può essere: 1) non sposata oppure 2) sposata. Analizziamo queste due possibilità:

**1) Se Anna non è sposata, allora Jack, che è sposato, sta guardando una persona non sposata: Anna.**

Se invece:

**2) Anna è sposata, lei stessa sta guardando una persona non sposata: George.**

In entrambi i casi la risposta è inevitabilmente:

**A) Sì, c'è una persona sposata che sta guardando una persona non sposata.**

**Morale.** Perché questo indovinello, a tutta prima, appare così complicato? La ragione è che l'informazione di partenza sembra essere insufficiente: lo stato civile di Anna non è noto né può essere inferito dalle condizioni iniziali; così l'intervistato – tramite il suo Sistema 1 veloce e istintivo nel decidere, ma impreciso [7] – conclude che anche la domanda sia indeterminata. Invece, lo stato maritale di Anna è irrilevante per la risposta, che ora ci appare ovvia.

Ogni problema diventa facile una volta che se ne conosca la soluzione, purché si siano analizzate tutte le possibilità. Anziché dare istintivamente una risposta non corretta (indotta dal Sistema 1), bisogna prima riflettere, attivando il Sistema 2, lento ma affidabile. Lo svantaggio è che, in termini di tempo ed energia, questo processo, in quanto basato sul ragionamento logico, è più dispendioso di quello del Sistema 1. In conclusione, l'esempio precedente riassume, in modo semplice ed efficace, l'essenza del pensiero critico come metodo di scelta.

## Bibliografia

- [1] Tomasin, L. (2018). "La tecnologia salverà le lingue?", *Domenica-Il Sole 24 Ore*, 96 (8 aprile), 23.
- [2] Barone, V. (2016). *Albert Einstein. Il costruttore di universi*, Laterza.
- [3] Harris, J. (2017). "Il mondo salvato dai robot", *Domenica-Il Sole 24 Ore*, 167 (25 giugno), 27.
- [4] Luvison, A. (2013). "Apologia della ragione scientifica", *Mondo Digitale – Rassegna critica del settore ICT*, 45 (marzo), 1-28, <http://>

mondodigitale.aicanet.net/2013-1/articoli/05\_LUVISON.pdf (ultimo accesso marzo 2019).

[5] Luvison, A. (2014). "Apologia della ragione scientifica – II: strumenti per decidere", *Mondo Digitale – Rassegna critica del settore ICT*, 55 (dicembre), 1-31, [http://mondodigitale.aicanet.net/2014-7/articoli/03\\_Apologia\\_della\\_ragione\\_scientifica\\_II.pdf](http://mondodigitale.aicanet.net/2014-7/articoli/03_Apologia_della_ragione_scientifica_II.pdf) (ultimo accesso marzo 2019).

[6] Bassetti T., Luvison, A. (2018). "Apologia della ragione scientifica – III: decisioni e giochi strategici", *Mondo Digitale – Rassegna critica del settore ICT*, 76 (maggio), 1-30, [http://mondodigitale.aicanet.net/2018-3/Articoli/MD76\\_01\\_Apologia\\_della\\_ragione\\_scientifica-III.pdf](http://mondodigitale.aicanet.net/2018-3/Articoli/MD76_01_Apologia_della_ragione_scientifica-III.pdf) (ultimo accesso marzo 2019).

[7] Kahneman, D. (2012). *Pensieri lenti e veloci*, Mondadori.

[8] Lewis, M. (2017). *Un'amicizia da Nobel. Kahneman e Tversky, l'incontro che ha cambiato il nostro modo di pensare*, Raffaello Cortina Editore.

[9] Thaler, R.H. (2018). *Misbehaving. La nascita dell'economia comportamentale*, Einaudi.

[10] Hagstrom, R.G. (2014). *Il metodo Warren Buffett. I segreti del più grande investitore del mondo*, 3ª edizione, Hoepli.

[11] Stanovich, K.E. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought*, Yale University Press.

[12] Rosenhouse, J., (2009). *The Monty Hall Problem: The Remarkable Study of Math's Most Contentious Brain Teaser*, Oxford University Press.

[13] Herbranson, W.T., Schroeder, T. (2010). "Are birds smarter than mathematicians? Pigeons (*Columbia livia*) perform optimally on a version of the Monty Hall Dilemma", *Journal of Comparative Psychology*, 1 (febbraio), 1-13, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3086893/> (ultimo accesso marzo 2019).

[14] Nozick, R. (1969). "Newcomb's problem and two principles of choice", in Rescher, N., *et al.* (a cura di), *Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of his Sixty-Fifth Birthday*, D. Reidel Publishing Company, 114-146.

[15] Gardner, M. (1973). "Mathematical games. Free will revisited with a mind-bending paradox by William Newcomb", *Scientific American*, 1 (luglio), 104-109. Tr. it. (1974). "Giochi matematici. Ancora sul libero arbitrio, con un affascinante paradosso di William Newcomb", *Le Scienze*, 65 (gennaio), 100-104.

[16] Nozick, R. (1974). "Mathematical Games. Reflections on Newcomb's problem: A prediction and free-will dilemma", *Scientific American*, 3 (marzo), 102-108. Tr. it. (1974). "Giochi matematici. Riflessioni sul paradosso di Newcomb: un dilemma sulla predizione e il libero arbitrio", *Le Scienze*, 72 (agosto), 106-111.

[17] Gardner, M. (2001). "Newcomb's paradox", in *The Colossal Book of Mathematics*, W. W. Norton & Company, 580-591.

- [18] Rasetti, M. (2019). "Big Data, Scienza dei Dati, Intelligenza Artificiale: sfide, prospettive, sogni e pericoli", relazione al *Convegno "Saperi e metodologie a confronto"*, Accademia delle Scienze di Torino, 22 gennaio, [https://www.youtube.com/watch?v=E3jmwD\\_sjsM](https://www.youtube.com/watch?v=E3jmwD_sjsM) (ultimo accesso marzo 2019).
- [19] Luvison, A. (2017). "L'ecosistema dell'innovazione digitale: analisi critica", *AEIT*, 3/4 (marzo/aprile), 6-27, [http://www.aeit.it/aeit/edicola/aeit/aeit2017/aeit2017\\_02\\_cisa/aeit2017\\_02\\_riv.pdf](http://www.aeit.it/aeit/edicola/aeit/aeit2017/aeit2017_02_cisa/aeit2017_02_riv.pdf) (ultimo accesso marzo 2019).
- [20] Wildberger, N.J. (2013). "Famous math problems 7: Newcomb's paradox", 19 febbraio, <https://www.youtube.com/watch?v=aR5GYeZkgvY> (ultimo accesso marzo 2019).
- [21] Gallucci, M. (2007). "Paradosso di Newcomb e dilemma del prigioniero: quando il problema non è la scelta da compiere, bensì la teoria in cui credere", Tesi di laurea specialistica in Scienze economiche (anno accademico 2006-2007), Università di Pisa, <https://core.ac.uk/download/pdf/14695341.pdf> (ultimo accesso marzo 2019).
- [22] Wilson, H.J., Daugherty, P.R. (2018). "Collaborative intelligence: Humans and AI are joining forces", *Harvard Business Review*, 4 (luglio-agosto), 114-123.
- [23] Davenport, T.H. (2018). *The AI Advantage: How to Put the Artificial Intelligence to Work*, The MIT Press.
- [24] Luvison, A. (2019). "La società e l'ecosistema digitale", in Pozzi, P. (a cura di), *Immagini del digitale. Dopo il Bit Bang*, Nemapress Edizioni, in corso di stampa.
- [25] AA. VV. (2017). "Special Report: Can We Copy the Brain?", *IEEE Spectrum*, 6 (giugno), <https://spectrum.ieee.org/static/special-report-can-we-copy-the-brain> (ultimo accesso marzo 2019).
- [26] B. Bryson, B. (2017). "Persi a Cyberlandia", in *Notizie da un grande paese*, Guanda, 315-319.
- [27] Anderson, C. (2008). *La coda lunga. Da un mercato di massa a una massa di mercati*, Codice Edizioni.
- [28] Levesque, H.J. (2018). *Common Sense, the Turing Test, and the Quest for Real AI*, The MIT Press.
- [29] Edmonds, D. (2014). *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Raffaello Cortina Editore.
- [30] Awad, E., et al. (2018). "The Moral Machine experiment", *Nature*, 563 (1 novembre), 59-64, <https://www.nature.com/articles/s41586-018-0637-6.pdf> (ultimo accesso marzo 2019).
- [31] Hao, K. (2018). "Should a self-driving car kill the baby or the grandma? Depends where you're from", *MIT Technology Review*, 24 ottobre, <https://www.technologyreview.com/s/612341/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem/> (ultimo accesso marzo 2019).
- [32] Floridi, L. (2017). "L'intelligenza artificiale. Cosa cambierà nella nostra società e nella nostra vita", 14 novembre, Politecnico di Torino, <https://>

www.fondazione scuola.it/costruire-futuro/intelligenza-artificiale (ultimo accesso marzo 2019).

[33] AA. VV. (2016). "Dossier Intelligenza artificiale – Auto senza guidatore", *Le Scienze*, 576 (agosto), pp. 38-55.

[34] AA. VV. (2017). "Special Issue: Machine Law", *Artificial Intelligence and Law*, 3 (settembre), 251-378, <https://www.springerprofessional.de/en/artificial-intelligence-and-law-3-2017/15085774> (ultimo accesso marzo 2019).

[35] AA. VV. (2019). "Special Issue: Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems", *Proceedings of the IEEE*, 3 (marzo).

[36] Trussell, H.J. (2018). "Point of view: Why a Special Issue on Machine Ethics", *Proceedings of the IEEE*, 10 (ottobre), 1774-1778, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8472909> (ultimo accesso marzo 2019).

[37] Bonnefon, J.-F., Shariff, A., Rawhan, I. (2019). "Point of view: The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars", *Proceedings of the IEEE*, 3 (marzo), 502-504, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8662742> (ultimo accesso marzo 2019).

[38] The European's Commission High-Level Experts Group on Artificial Intelligence (2018). *Draft Ethics Guidelines for Trustworthy AI: Working Document for Stakeholders' Consultation*, 18 dicembre, <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai> (ultimo accesso marzo 2019).

[39] Floridi, L., *et al.* (2018). "AI4People – An ethical framework for Good AI Society: Opportunities, risks, principles, and recommendations", *Minds and Machines*, 4 (dicembre), 689-707, <https://link.springer.com/content/pdf/10.1007%2Fs11023-018-9482-5.pdf> (ultimo accesso marzo 2019).

[40] Floridi, L. (2019). "Consapevolezza per un'etica dell'intelligenza artificiale", *Il Sole 24 Ore*, 12 (13 gennaio), 11.

[41] Domingos, P. (2016). *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, Bollati Boringhieri.

[42] Defez, R. (2018). *Scoperta. Come la ricerca scientifica può aiutare a cambiare l'Italia*, Codice Edizioni.

[43] Panarari, M. (2018). "Se la scienza diventa ideologia", *La Stampa*, 307 (7 novembre), 21.

[44] Tipaldo, G. (2019). *La società della pseudoscienza. Orientarsi tra buone e cattive spiegazioni*, il Mulino.

[45] Rosling, H., con Rosling, O. e Rosling Rönnlund, A. (2018). *Factfulness. Dieci ragioni per cui non capiamo il mondo e perché le cose vanno meglio di come pensiamo*, Rizzoli.

[46] Luvison, A. (2018). "Editoriale. L'ecosistema digitale: le responsabilità di chi sviluppa l'ICT", *AEIT*, 7-8 (luglio-agosto), 4-5, [https://www.aeit.it/aeit/edicola/aeit/aeit2018/aeit2018\\_04\\_cisa/aeit2018\\_04\\_riv.pdf](https://www.aeit.it/aeit/edicola/aeit/aeit2018/aeit2018_04_cisa/aeit2018_04_riv.pdf) (ultimo accesso marzo 2019).

- [47] Nichols, T. (2018). *La conoscenza e i suoi nemici. L'era dell'incompetenza e i rischi per la democrazia*, LUISS University Press.
- [48] Hagstrom, R.G. (2013). *Investing: The Last Liberal Art*, seconda edizione, Columbia Business School.
- [49] Snow, C.P. (2005). *Le due culture* (a cura di Lanni, A.), I libri di Rese-Marsilio.
- [50] Bucchi, M. (2018). *Sbagliare da professionisti. Storie di errori e fallimenti memorabili*, Rizzoli.
- [51] Baumol, W.J. (2005). "Errors in economics and their consequences", *Social Research*, 1 (primavera), 169-194.
- [52] Bar-Hillel, M., Noha, T., Frederick, S. (2018). "Learning psychology from riddles: The case of stumpers", *Judgement and Decision Making*, 1 (gennaio), 112-122, <http://www.ratio.huji.ac.il/sites/default/files/publications/dp714.pdf> (ultimo accesso marzo 2019).
- [53] Luvison, A. (2015). "La crittologia da arte a scienza: l'eredità di Shannon e Turing", *Mondo Digitale – Rassegna critica del settore ICT*, anno XIV, 60 (novembre), 1-31, [http://mondodigitale.aicanet.net/2015-5/articoli/03\\_crittologia\\_da\\_arte\\_a\\_scienza.pdf](http://mondodigitale.aicanet.net/2015-5/articoli/03_crittologia_da_arte_a_scienza.pdf) (ultimo accesso marzo 2019).
- [54] Mancini, F. (2015). "Il dilemma del trolley, il conflitto tra colpa deontologica e colpa altruistico/umanitaria e il disturbo ossessivo", *State of Mind*, 25 maggio, <http://www.stateofmind.it/2015/05/colpa-disturbo-ossessivo/> (ultimo accesso marzo 2019).

## Biografie

**Thomas Bassetti** è professore associato di Politica Economica presso l'Università degli Studi di Padova (Dipartimento di Scienze Economiche e Aziendali "Marco Fanno"), dove insegna "Macroeconomia", "Monetary and Fiscal Policy" e "Economics of Human Capital". Dopo aver conseguito la laurea in Economia e Commercio presso l'Università degli Studi di Pisa, ha passato un periodo di ricerca come Honorary Fellow alla Wisconsin University-Madison per poi tornare in Italia e concludere il dottorato di ricerca in Economia Politica, sempre a Pisa. Tra le sue pubblicazioni troviamo riviste internazionali come: *Journal of Socio-Economics*, *Environmental and Resource Economics*, *Economic Inquiry*, *Journal of Family Business Strategy* e *Applied Economics*. I suoi attuali interessi di ricerca riguardano soprattutto l'economia ambientale e il finanziamento delle campagne elettorali.

Email: [thomas.bassetti@unipd.it](mailto:thomas.bassetti@unipd.it)

**Angelo Luvison** è ingegnere elettronico (Politecnico di Torino) dal 1969, con successivi perfezionamenti in teoria statistica delle comunicazioni al MIT e in management aziendale all'INSEAD-CEDEP di Fontainebleau. Per oltre trent'anni in CSELT, ha svolto e diretto ricerche in teoria delle comunicazioni, reti di fibre ottiche ad alta velocità, società dell'informazione, anche nell'ambito di progetti cooperativi internazionali. È stato professore di "Teoria dell'Informazione e della

Trasmissione” all’Università di Torino. Ha ricoperto la posizione di segretario generale dell’AEIT. Nell’ambito di Federmanager, la maggiore associazione nazionale dei dirigenti d’aziende industriali, si è occupato di formazione permanente ed è stato presidente di Federmanager Piemonte. Detiene sette brevetti e, tra saggi e articoli scientifici e divulgativi, è autore, o coautore, di oltre 200 lavori, uno dei quali è stato ripubblicato (2007) nel volume celebrativo *The Best of the Best: Fifty Years of Communications and Networking Research* della IEEE Communications Society. È Life Member dell’IEEE, membro del Comitato scientifico di *Mondo Digitale*, del Comitato editoriale di *AEIT* e dell’Advisory Board di *Harvard Business Review*. Si occupa e scrive di temi di innovazione per l’informatica e le telecomunicazioni.

Email: [angelo.luvison@alice.it](mailto:angelo.luvison@alice.it); [angelo.luvison@gmail.com](mailto:angelo.luvison@gmail.com)