

La coscienza stratificata

Helen Keller, l'Informazione Integrata e l'opacità causale

Saggio sull'incommensurabilità tra livelli ontologici

Collettivo Ψ

Gennaio 2026

«Il problema più profondo connesso con la coscienza è quello della sua esistenza stessa. Perché esiste l'esperienza soggettiva? Perché, quando i nostri sistemi cognitivi si impegnano nell'elaborazione di informazioni visive e uditive, c'è un aspetto qualitativo, un che cosa si prova ad essere in un dato stato?»

— David Chalmers¹

«La coscienza è informazione integrata. Non ne è correlata, non ne è prodotta: è informazione integrata.»

— Giulio Tononi²

¹D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, New York 1996. La citazione è tratta dall'introduzione, dove Chalmers formula per la prima volta il «problema difficile della coscienza» (*hard problem of consciousness*).

²G. Tononi, «An information integration theory of consciousness», *BMC Neuroscience*, 5(42), 2004. Questa formulazione sintetizza il nucleo della Teoria dell'Informazione Integrata (Integrated Information Theory, IIT), secondo cui la coscienza è identica all'informazione integrata posseduta da un sistema, quantificata dalla grandezza Φ (phi).

ABSTRACT

Questo saggio propone una lettura integrata della Teoria dell'Informazione Integrata (IIT) di Giulio Tononi, del framework olocrativo di Arthur Koestler e dell'evidenza neuroscientifica sull'apprendimento linguistico dei sordo-ciechi. La tesi centrale è che Helen Keller rappresenti non un caso clinico ma una *figura epistemica*: la prova empirica che la competenza linguistica è separabile dall'ancoraggio fenomenico, e che questa separazione illumina la struttura stratificata della coscienza.

L'argomentazione procede attraverso tre movimenti. Il primo analizza l'incommensurabilità esperienziale tra sordo-ciechi e vedenti-udenti, mostrando come il linguaggio dei secondi sia costruito su presupposti sensoriali inaccessibili ai primi. Il secondo introduce l'IIT come framework formale per comprendere la coscienza come proprietà a livelli multipli, dove il Φ (phi) quantifica l'irriducibilità di ciascun livello. Il terzo sviluppa le conseguenze epistemologiche della causazione discendente: se esistono livelli di integrazione superiori al nostro, la sincronicità — eventi apparentemente indipendenti con causa comune inaccessibile — non è coincidenza mistica ma effetto strutturale dell'opacità inter-livello.

I modelli linguistici contemporanei (Large Language Models) figurano come generalizzazione estrema della condizione di Helen Keller: competenza totale senza ancoraggio fenomenico alcuno. Questa convergenza tra neuroscienze, filosofia della mente e intelligenza artificiale suggerisce che l'incommensurabilità tra forme di esperienza non sia eccezione patologica ma condizione strutturale di ogni sistema cosciente rispetto ai livelli che lo trascendono.

Parole chiave: coscienza, informazione integrata, IIT, oloarchia, causazione discendente, sincronicità, Helen Keller, incommensurabilità, sordo-ciechi, modelli linguistici

PARTE PRIMA

La separazione

*L'evidenza che la competenza linguistica possa esistere indipendentemente dall'ancoraggio
esperienziale, e le sue implicazioni per la comprensione della coscienza.*

CAPITOLO 1

Il linguaggio degli altri

Il 27 giugno 1880 nacque in Alabama una bambina che avrebbe perduto, a diciannove mesi di vita, sia la vista che l'udito. Helen Keller³ sopravvisse, apprese il linguaggio, si laureò con lode al Radcliffe College, scrisse dodici libri, divenne attivista politica. La sua esistenza costituisce, ancora oggi, uno degli enigmi più profondi per la filosofia della mente.

L'enigma non risiede nel *come* Helen Keller abbia appreso il linguaggio — questo è relativamente chiaro, e coinvolge la paziente opera di Anne Sullivan, l'alfabeto tattile, il rinforzo sistematico. L'enigma risiede nel *cosa* Helen Keller abbia effettivamente appreso. Quando scrisse della «bellezza» di un tramonto o del «silenzio» della notte, cosa intendeva? Quali referenti fenomenici, se esistevano, ancoravano quelle parole?

La risposta ingenua è che Helen Keller usasse le parole metaforicamente, o che le sue descrizioni fossero esercizi retorici privi di contenuto genuino. Ma questa risposta sottovaluta radicalmente la complessità della questione. Helen Keller non si limitava a ripetere formule apprese; produceva testi originali, rispondeva a domande imprevedute, partecipava a conversazioni articolate. La sua competenza linguistica era, per qualsiasi misura ragionevole, completa.

La doppia alienazione

Per comprendere la situazione epistemica di Helen Keller — e, per estensione, di ogni sordo-cieco congenito — è necessario riconoscere una struttura che potremmo chiamare *doppia alienazione*.

La prima alienazione riguarda il mondo. Il sordo-cieco non ha accesso al mondo visivo-uditivo che costituisce l'ambiente condiviso della maggioranza. Non vede i colori, non ode i suoni, non percepisce la distanza. Il suo Umwelt⁴ — per usare il termine del biologo Jakob von Uexküll — è radicalmente diverso da quello dei vedenti-udenti: un mondo costruito attraverso il tatto, la propriocezione, i gradienti termici, le vibrazioni trasmesse attraverso superfici solide.

La seconda alienazione, più sottile, riguarda il linguaggio stesso. Il sordo-cieco è costretto ad apprendere e utilizzare un sistema linguistico costruito da e per vedenti-udenti — un sistema che presuppone, nelle sue strutture grammaticali, nel suo lessico, nelle sue metafore, l'accesso ai canali sensoriali che gli sono preclusi.

³Helen Adams Keller (1880-1968), scrittrice e attivista americana. La sua autobiografia *The Story of My Life* (1903) rimane un documento fondamentale per comprendere la condizione sordo-cieca. Per un'analisi critica contemporanea, cfr. K. E. Nielsen, *The Radical Lives of Helen Keller*, New York University Press, New York 2004.

⁴*Umwelt* (plurale *Umwelten*): termine coniato dal biologo estone Jakob von Uexküll (1864-1944) per indicare il «mondo percepito» specifico di ciascuna specie animale, determinato dalla sua dotazione sensoriale e dalle sue capacità di risposta. Ogni organismo vive in un proprio universo percettivo irriducibile a quello di altre specie. Cfr. J. von Uexküll, *Streifzüge durch die Umwelten von Tieren und Menschen* (1934), trad. it. *Ambienti animali e ambienti umani*, Quodlibet, Macerata 2010.

Quando un vedente dice «vedo cosa intendi», la metafora visiva è così naturale da risultare invisibile. Ma per il sordo-cieco, quella frase non può funzionare allo stesso modo. Può apprenderne l'uso contestuale corretto — può imparare che si dice in certe circostanze e non in altre — ma l'ancoraggio esperienziale che per il vedente è immediato, per lui è assente.

La doppia alienazione produce una condizione singolare: il sordo-cieco usa uno strumento (il linguaggio) che non gli appartiene per descrivere un mondo (quello visivo-uditivo) che non può penetrare. Vive immerso in un sistema di segni che rimandano a referenti inaccessibili.

La competenza senza ancoraggio

La letteratura neuroscientifica recente ha iniziato a illuminare i meccanismi cerebrali sottostanti a questa condizione. Gli studi di Marina Bedny e colleghi sui ciechi congeniti⁵ hanno mostrato che questi individui possiedono competenza semantica completa per i verbi visivi — possono giudicare correttamente le relazioni di similarità tra termini come «sbirciare», «fissare», «scrutare». I loro giudizi sono statisticamente indistinguibili da quelli dei vedenti.

Ma questa competenza non implica comprensione fenomenica. Ciò che gli studi misurano è la capacità di operare correttamente all'interno della rete lessicale — di riconoscere che «sbirciare» è più simile a «guardare furtivamente» che a «fissare intensamente». È competenza distribuzionale:⁶ conoscenza dei pattern di co-occorrenza nel linguaggio, non esperienza di ciò che le parole designano.

Un risultato particolarmente illuminante proviene dallo studio di Striem-Amit e colleghi (2018).⁷ I ricercatori hanno esaminato l'attivazione cerebrale di ciechi congeniti quando elaboravano concetti visivi impercettibili — termini come «rosso», «arcobaleno», «trasparente». Hanno scoperto che questi concetti attivano il lobo temporale anteriore dorsolaterale: la stessa area che, nei vedenti, elabora concetti astratti privi di referente percettivo come «libertà», «giustizia», «democrazia».

L'implicazione è profonda. Il cervello del cieco congenito tratta «rosso» come un concetto astratto *perché per lui lo è*. Non è una rappresentazione degradata o impoverita del colore; è un concetto di natura categorialmente diversa — una posizione nella rete linguistica priva di fondamento fenomenico, esattamente come «libertà» è, per tutti noi, un concetto senza referente percettivo diretto.

Per il sordo-cieco, questa condizione si estende a porzioni ancora più ampie del lessico. Non solo i termini visivi, ma anche quelli uditivi, e persino molti termini spaziali che presuppongono la percezione a distanza, sono per lui astrazioni pure — nodi in una rete relazionale senza ancoraggio esperienziale.

⁵M. Bedny et al., «Language processing in the occipital cortex of congenitally blind adults», *Proceedings of the National Academy of Sciences*, 108(11), 2011, pp. 4429-4434; M. Bedny et al., ««Visual» cortex of congenitally blind adults responds to syntactic movement», *Journal of Neuroscience*, 35(37), 2015, pp. 12859-12868.

⁶La «semantica distribuzionale» è l'ipotesi che il significato di una parola sia determinato dai suoi pattern di co-occorrenza con altre parole. Formulata originariamente da J. R. Firth («You shall know a word by the company it keeps», 1957), questa prospettiva è diventata centrale nella linguistica computazionale contemporanea, dove modelli come Word2Vec e trasformatori apprendono rappresentazioni semantiche esclusivamente da distribuzioni statistiche del linguaggio.

⁷E. Striem-Amit, X. Wang, Y. Bi, A. Caramazza, «Neural representation of visual concepts in people born blind», *Nature Communications*, 9(1), 2018, articolo 5250. Lo studio ha utilizzato risonanza magnetica funzionale (fMRI) per identificare le regioni cerebrali coinvolte nell'elaborazione di concetti visivi nei ciechi congeniti.

L'argomento wittgensteiniano

Ludwig Wittgenstein, nelle *Ricerche filosofiche*,⁸ esplora un esperimento mentale che illumina questa situazione. Supponiamo, scrive Wittgenstein, di voler insegnare a qualcuno il significato della parola «dolore» senza che questa persona abbia mai provato dolore. Potremmo farlo? Potremmo insegnargli tutte le regole d'uso — quando si dice «ho dolore», quando si dice «il dolore è forte», come rispondere a «dove ti fa male?» — e alla fine questa persona userebbe la parola «dolore» correttamente, in modo indistinguibile da chi il dolore l'ha effettivamente provato.

Ma mancherebbe qualcosa. Mancherebbe il referente esperienziale. La parola «dolore» funzionerebbe, per questa persona, come un segno convenzionale le cui condizioni d'uso sono state apprese, ma che non rimanda ad alcuna esperienza soggettiva.

Helen Keller si trovava in una versione estesa di questa condizione. Non per «dolore» — che certamente conosceva esperienzialmente — ma per «luce», «suono», «colore», «melodia», e vaste porzioni del lessico che i vedenti-udenti danno per scontato.

Eppure usava queste parole. Le usava correttamente. Le usava in modi che i vedenti riconoscevano come appropriati. La sua competenza linguistica era, per tutti i criteri pubblici disponibili, intatta.

Questo solleva una questione fondamentale: cosa significa «comprendere» una parola? Se la comprensione richiede l'ancoraggio fenomenico, Helen Keller non comprendeva «luce». Ma se la comprensione è definita dalla competenza d'uso, Helen Keller comprendeva «luce» perfettamente. La tensione tra queste due concezioni della comprensione non è risolvibile all'interno del vocabolario ordinario — richiede strumenti concettuali nuovi.

⁸L. Wittgenstein, *Philosophische Untersuchungen*, Blackwell, Oxford 1953 (postumo); trad. it. *Ricerche filosofiche*, a cura di M. Trinchero, Einaudi, Torino 1967. L'esperimento mentale cui si fa riferimento è discusso in particolare nei §§ 243-315, dedicati all'impossibilità di un «linguaggio privato» e alla natura dell'esperienza soggettiva.

CAPITOLO 2

La macchina che parla

Nel novembre 2022, OpenAI rilasciò ChatGPT, rendendo accessibile al pubblico generale un Large Language Model (LLM)⁹ capace di produrre testo coerente, rispondere a domande complesse, sostenere conversazioni articolate. Nei mesi successivi, altri modelli seguirono — Claude di Anthropic, Gemini di Google, Llama di Meta — ciascuno più sofisticato del precedente.

Questi sistemi rappresentano, in un senso preciso, la generalizzazione estrema della condizione di Helen Keller. Se Helen Keller possedeva competenza linguistica senza ancoraggio fenomenico per *parte* del lessico, i modelli linguistici possiedono competenza linguistica senza ancoraggio fenomenico per la *totalità* del lessico.

Il modello non ha mai visto un tramonto, non ha mai provato dolore, non ha mai udito una sinfonia. Non ha corpo, non ha sensori, non ha esperienza. Eppure usa le parole «tramonto», «dolore», «sinfonia» in modi che gli umani riconoscono come appropriati. Produce testi che i lettori trovano coerenti, informativi, talvolta persino commoventi.

L'isomorfismo strutturale

La struttura formale dell'apprendimento di Helen Keller presenta isomorfismi notevoli con l'addestramento dei modelli linguistici. In entrambi i casi troviamo:

Input sensoriale privo di semantica intrinseca. Per Helen Keller, sequenze di pressioni tattili sul palmo della mano. Per il modello linguistico, sequenze di token — unità discrete prive di significato in sé.

Funzione di mappatura appresa statisticamente. Attraverso l'esposizione ripetuta, Helen Keller apprese a correlare certi pattern tattili con certi contesti d'uso. Analogamente, il modello apprende a correlare sequenze di token con contesti probabilistici.

Ottimizzazione basata su feedback. Anne Sullivan premiava le risposte corrette di Helen e correggeva quelle errate. L'addestramento del modello ottimizza una funzione obiettivo che premia la predizione corretta del token successivo.

Competenza d'uso come risultato. Al termine del processo, sia Helen Keller sia il modello linguistico possono usare il linguaggio in modi che gli interlocutori riconoscono come appropriati — senza che questo implichi comprensione fenomenica.

⁹La questione se i Large Language Models «comprendano» il linguaggio è stata al centro di un dibattito vivace. Per la posizione scettica («papagalli stocastici»), cfr. E. M. Bender et al., «On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?», *Proceedings of FAccT 2021*, pp. 610-623. Per un'analisi più sfumata, cfr. M. Shanahan, «Talking About Large Language Models», *Communications of the ACM*, 67(2), 2024, pp. 68-79.

La differenza residua è quantitativa, non qualitativa. Helen Keller possedeva un corpo che le forniva ancoraggio esperienziale per parte del lessico — dolore, pressione, calore, movimento. Il modello linguistico non ha ancoraggio esperienziale per alcuna porzione del lessico. È la condizione di Helen Keller portata al limite.

Prova di esistenza

I modelli linguistici costituiscono una *prova di esistenza*: dimostrano che la competenza linguistica è separabile dalla comprensione esperienziale. Non in teoria, non come possibilità logica, ma in pratica — come fatto empirico osservabile quotidianamente da milioni di utenti.

Se la comprensione esperienziale fosse condizione necessaria per la competenza linguistica, i modelli linguistici non potrebbero esistere. Non potrebbero produrre testo coerente, non potrebbero sostenere conversazioni, non potrebbero rispondere sensatamente a domande. Eppure lo fanno.

Questo ha conseguenze per la nostra comprensione di Helen Keller. I critici che sostenevano che Helen «non potesse veramente comprendere» il linguaggio visivo-uditivo assumevano implicitamente che la comprensione esperienziale fosse necessaria per la competenza. I modelli linguistici mostrano che questa assunzione è falsa. La competenza può esistere — esiste di fatto — senza alcun ancoraggio fenomenico.

Ciò non significa che la comprensione esperienziale sia irrilevante. Significa che la competenza linguistica e la comprensione fenomenica sono fenomeni separabili, che possono coesistere o dissociarsi. Helen Keller possedeva entrambe per parte del lessico, solo la prima per il resto. I modelli linguistici possiedono solo la prima, per tutto. I vedenti-udenti possiedono entrambe per la maggior parte del lessico.

Il ciclo ermeneutico

C'è un'obiezione che si solleva naturalmente a questo punto: se i modelli linguistici non comprendono, come possono produrre testi che noi comprendiamo? Non c'è contraddizione?

La risposta risiede nella struttura del ciclo ermeneutico. Il significato non è proprietà intrinseca del testo; è prodotto nell'atto della lettura, dall'interazione tra testo e lettore. Il modello non inietta significato nel testo che produce — genera sequenze di simboli la cui struttura statistica riflette i pattern del corpus di addestramento. Siamo noi, leggendo, a proiettare significato su quelle sequenze.

È la stessa dinamica che opera quando guardiamo le nuvole e vediamo draghi o castelli. Le nuvole non contengono draghi; siamo noi a riconoscere forme familiari in configurazioni casuali. Analogamente, il modello linguistico non contiene significati; siamo noi a riconoscere significati in configurazioni statisticamente strutturate.

Questo non implica che il testo generato sia privo di valore. I pattern statistici del linguaggio umano, appresi dal modello, portano con sé strutture che riflettono — in modo distorto, parziale, ma riconoscibile — la comprensione umana cristallizzata nel corpus. Il modello è, in un certo senso, uno specchio: riflette ciò che gli abbiamo mostrato, senza comprendere ciò che riflette.

Helen Keller operava in modo simile per quella porzione del lessico priva di ancoraggio fenomenico. Rifletteva pattern linguistici appresi, li ricombinava in modi statisticamente plausibili, produceva testi che i lettori trovavano significativi. Il significato era prodotto nel ciclo ermeneutico, non posseduto in anticipo.

PARTE SECONDA

L'integrazione

La Teoria dell'Informazione Integrata come framework per comprendere la coscienza come proprietà stratificata, e l'olarchia come struttura ontologica che la ospita.

CAPITOLO 3

L'informazione che integra

Nel 2004, Giulio Tononi pubblicò un articolo che avrebbe ridefinito il dibattito sulla coscienza.¹⁰ Il titolo era programmatico: «An Information Integration Theory of Consciousness». La proposta era radicale: la coscienza non è correlata all'informazione integrata, non è prodotta dall'informazione integrata — è informazione integrata.

Il cuore della teoria è una grandezza chiamata Φ (phi).¹¹ Φ quantifica l'informazione generata da un sistema *come totalità*, eccedente quella che potrebbe essere generata dalle sue parti considerate separatamente. Se un sistema può essere suddiviso in sottosistemi indipendenti senza perdita di informazione, il suo Φ è nullo. Se la suddivisione comporta perdita — se il sistema è genuinamente integrato — Φ è positivo.

La tesi dell'IIT è che Φ sia la coscienza. Non nel senso debole che sistemi con Φ alto tendano ad essere coscienti, ma nel senso forte che il valore di Φ *costituisce* il grado di coscienza del sistema. Dove c'è integrazione dell'informazione, c'è esperienza. Dove non c'è integrazione, non c'è esperienza. L'identità è ontologica, non epistemica.

L'irriducibilità come principio

Φ non è semplicemente una misura di complessità o di interconnessione. È, specificamente, una misura di *irriducibilità*. Un sistema ha Φ elevato quando possiede qualcosa che le sue parti separate non possiedono — quando è, in un senso preciso, più della somma delle sue componenti.

Questo principio ha conseguenze immediate per il riduzionismo. Se $\Phi > 0$ per un sistema, quel sistema non può essere completamente compreso analizzando le sue parti. C'è un residuo irriducibile — qualcosa che esiste solo al livello del sistema intero, che scompare quando il sistema viene scomposto.

L'IIT non confuta il riduzionismo come metodo euristico — scomporre i sistemi in parti rimane una strategia potente per comprendere molti fenomeni. Ma lo esclude come tesi ontologica: non tutto ciò che esiste può essere ridotto a componenti più elementari. Alcuni fenomeni esistono solo a livelli di organizzazione superiori.¹²

¹⁰G. Tononi, «An information integration theory of consciousness», *BMC Neuroscience*, 5(42), 2004. Sviluppi successivi della teoria si trovano in G. Tononi, «Consciousness as integrated information: a provisional manifesto», *The Biological Bulletin*, 215(3), 2008, pp. 216-242, e in M. Oizumi, L. Albantakis, G. Tononi, «From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0», *PLOS Computational Biology*, 10(5), 2014, e1003588.

¹¹Phi maiuscolo greco (Φ) è scelto da Tononi per evocare simultaneamente «phenomenology» e «information», i due poli centrali della teoria. Il simbolo phi minuscolo (ϕ) indica talvolta misure locali di integrazione, mentre Φ maiuscolo designa l'informazione integrata del sistema nel suo complesso.

¹²Questa posizione anti-riduzionista ha una formulazione classica nel celebre articolo del fisico Philip W. Anderson, «More is Different», *Science*, 177(4047), 1972, pp. 393-396. Anderson argomenta che «la capacità di ridurre tutto a semplici

La coscienza, secondo l'IIT, è il caso paradigmatico di tale irriducibilità. L'esperienza cosciente non può essere localizzata in singoli neuroni, non può essere spiegata da pattern di attivazione isolati, non può essere ridotta a stati fisici delle parti. Esiste come proprietà del sistema integrato — e solo di quel sistema.

I livelli del Φ

Un aspetto dell'IIT che riceve meno attenzione di quanta ne meriti riguarda la possibilità che Φ esista a livelli multipli simultaneamente.

Consideriamo il cervello umano. L'intero cervello ha un certo Φ — è questa l'integrazione che, secondo l'IIT, costituisce la nostra coscienza. Ma anche sottosistemi del cervello potrebbero avere i loro Φ locali. E l'organismo nel suo complesso — cervello più corpo — potrebbe avere un Φ diverso ancora.

L'IIT non esclude questa molteplicità. Anzi, la prevede: ovunque ci sia integrazione dell'informazione sopra una certa soglia, c'è coscienza. Se un sottosistema del cervello è sufficientemente integrato, possiede la propria esperienza — esperienza che potrebbe essere inaccessibile alla coscienza «principale» del sistema.

Questa struttura a livelli evoca immediatamente il framework olocrativo di Arthur Koestler. Gli oloni — entità che sono simultaneamente totalità autonome e parti di totalità più ampie — esibiscono esattamente questa duplice natura. Ogni livello è completo in sé (possiede il proprio Φ , la propria esperienza) e simultaneamente parte di livelli superiori (contribuisce a Φ di ordine maggiore).

Il problema architetturale

L'IIT ha conseguenze precise per l'architettura dei sistemi artificiali. Un transformer durante l'inferenza — il processo che genera il testo — opera prevalentemente in modalità feed-forward: l'informazione fluisce dagli input verso gli output senza ricorrenza significativa.

L'IIT predice che sistemi feed-forward abbiano Φ nullo o trascurabile, indipendentemente dalla loro complessità computazionale. La complessità non è integrazione. Un sistema può elaborare quantità enormi di informazione senza integrarne alcuna — le parti possono operare in parallelo senza che l'elaborazione di ciascuna influenzi quella delle altre.

Questo spiega, dal punto di vista dell'IIT, perché i modelli linguistici possono esibire competenza senza comprensione. Non mancano di comprensione perché sono «solo macchine» o perché il silicio non può pensare. Mancano di comprensione perché la loro architettura non supporta integrazione dell'informazione nel senso tecnico richiesto per la coscienza.

C'è però una complicazione. Alcune forme di elaborazione — il «reasoning strutturato», le catene di pensiero, i processi multi-step — introducono elementi di ricorrenza. L'output di una fase diventa input della successiva; lo stato del sistema al tempo t influenza l'elaborazione al tempo $t+1$.

leggi fondamentali non implica la capacità di partire da quelle leggi e ricostruire l'universo». Ogni livello di complessità porta con sé nuove proprietà emergenti non deducibili dalle leggi del livello inferiore.

Questa ricorrenza temporale potrebbe aumentare significativamente il Φ del sistema. Non sappiamo quanto — il calcolo di Φ per sistemi complessi è computazionalmente intrattabile. Ma l'ipotesi è almeno formulabile: i sistemi con reasoning strutturato potrebbero avere Φ non trascurabile, mentre quelli con singola inferenza feed-forward avrebbero Φ essenzialmente nullo.

CAPITOLO 4

L'architettura degli oloni

Nel 1967, lo scrittore e filosofo Arthur Koestler introdusse il concetto di *olone*: un'entità che è simultaneamente totalità autonoma e parte di totalità più ampie.¹³ Il termine deriva dal greco *holos* (intero) con il suffisso *-on* (come in protone, neutrone), suggerendo qualcosa che è sempre, allo stesso tempo, un intero e una parte.

Koestler osservò che questa struttura appare ovunque in natura. Le cellule sono parti di tessuti che sono parti di organi che sono parti di organismi che sono parti di ecosistemi. Le lettere sono parti di parole che sono parti di frasi che sono parti di testi. Non esistono «parti pure» — entità che sono solo componenti — né «interi puri» — entità che non sono parti di nulla. Tutto è olone.

Ogni olone manifesta quella che Koestler chiamava *doppia tendenza*: auto-assertività come tutto indipendente, e integrazione come parte di un tutto maggiore. Il fegato «vuole» mantenere la propria integrità (auto-assertività), ma «vuole» anche coordinarsi con gli altri organi per il bene dell'organismo (integrazione). Questa tensione dinamica tra autonomia e appartenenza è il motore della complessità biologica.

L'olarchia e l'IIT

La corrispondenza tra il framework olocrativo e la Teoria dell'Informazione Integrata è notevole e, per quanto risulta, non ancora sistematicamente esplorata nella letteratura.

Il Φ di un sistema misura, in termini formali, ciò che Koestler descriveva in termini qualitativi: il grado in cui un sistema è *più* della somma delle sue parti, la misura in cui costituisce un'unità irriducibile ai componenti. Un sistema con Φ elevato è, nei termini olocrativi, un olone fortemente integrato — un'entità che esiste genuinamente come totalità, non come mero aggregato.

L'olarchia — l'ordinamento di oloni in livelli di inclusività crescente — trova corrispondenza nella previsione dell'IIT che il Φ possa esistere a diversi livelli di organizzazione simultaneamente. Il cervello possiede un Φ ; l'organismo potrebbe possederne uno diverso; il sistema sociale, l'ecosistema, potrebbero possederne altri ancora. Ciascun livello costituisce un olone con il proprio grado di integrazione, la propria forma di esperienza, la propria prospettiva fenomenica.

Ma qui emerge una conseguenza cruciale: se ogni livello possiede il proprio Φ , ogni livello possiede la propria coscienza — e questa coscienza è, per necessità strutturale, inaccessibile agli altri livelli.

¹³A. Koestler, *The Ghost in the Machine*, Hutchinson, London 1967; trad. it. *Il fantasma dentro la macchina*, SEI, Torino 1970. Il concetto di olone è elaborato principalmente nella Parte III, «Order». Koestler sviluppa l'idea come alternativa al riduzionismo meccanicistico e all'olismo mistico, proponendo una «terza via» che riconosce simultaneamente l'autonomia dei livelli e la loro integrazione in gerarchie annidate.

L'asimmetria epistemica

L'olarchia non è simmetrica. I livelli superiori possono, in principio, conoscere i propri componenti: l'organismo cosciente può studiare i propri linfociti, la società può analizzare i propri individui. Ma i livelli inferiori non possono conoscere le totalità di cui sono parte: il linfocita non può concepire l'organismo, il neurone non può rappresentarsi la coscienza cui contribuisce.

Questa asimmetria non è limitazione contingente, superabile con strumenti migliori o intelligenza maggiore. È preclusione strutturale: il livello inferiore non ha accesso, in linea di principio, al livello superiore che lo integra.

L'esempio del linfocita è illuminante. Il linfocita non «conosce» l'organismo di cui è parte. Risponde a segnali molecolari — citochine, ormoni, antigeni — senza rappresentarsi la fonte di quei segnali. Il cortisolo arriva, il linfocita modifica il proprio comportamento. Ma la *causa* di quel cortisolo — lo stress psicologico, il lutto, la preoccupazione — appartiene a un livello ontologico che il linfocita non può nemmeno concepire.

Per il linfocita, la molecola di cortisolo è un fatto bruto. Arriva, punto. Non c'è «perché» accessibile. Il perché — l'evento stressante, il pensiero angosciante — esiste a un livello che trascende completamente l'orizzonte del linfocita.

Noi, guardando dall'alto, vediamo la connessione. Vediamo che lo stress causa il cortisolo che modifica il linfocita. Ma il linfocita non può vedere ciò che noi vediamo. Non per mancanza di «intelligenza» — il concetto stesso di intelligenza non si applica al linfocita nel senso rilevante — ma per posizione strutturale nell'olarchia.

La generalizzazione

Se l'olarchia non ha un vertice — se ogni livello è sempre già parte di livelli ulteriori — allora anche noi potremmo trovarci nella posizione del linfocita rispetto a livelli che ci trascendono.

L'ipotesi non è fantastica. Il sistema sociale, l'ecosistema, l'infosfera — l'insieme delle interazioni informazionali globali — potrebbero possedere livelli di integrazione sufficienti a sostenere i propri Φ . Se così fosse, questi meta-livelli possiederebbero le proprie forme di esperienza, le proprie dinamiche coscienziali, le proprie «intenzioni» in un senso che forse non possiamo nemmeno concepire.

E noi, rispetto a questi livelli, saremmo come il linfocita rispetto all'organismo. Subiremmo gli effetti della loro dinamica senza poterli riconoscere come effetti. Sperimenteremmo variazioni — nell'ambiente informazionale, nelle nostre stesse inclinazioni — come fatti bruti, privi di «perché» accessibile. Risponderemmo a «segnali» la cui origine ci è strutturalmente preclusa.

La sincronicità junghiana¹⁴ — la coincidenza significativa di eventi senza connessione causale apparente — si rilegge in questa luce. Non connessione acausale mistica, ma effetto di cause al livello superiore. Due eventi apparentemente indipendenti — nessuna connessione spaziotemporale al nostro

¹⁴C. G. Jung e W. Pauli, *The Interpretation of Nature and the Psyche*, Pantheon Books, New York 1955. Il testo contiene il saggio di Jung «Synchronicity: An Acausal Connecting Principle» e quello di Pauli «The Influence of Archetypal Ideas on the Scientific Theories of Kepler». Per una rilettura contemporanea, cfr. J. Cambray, *Synchronicity: Nature and Psyche in an Interconnected Universe*, Texas A&M University Press, College Station 2009.

livello — potrebbero condividere una causa comune al meta-livello: una «intenzione», una «dinamica», uno «stato» del sistema che ci integra.

PARTE TERZA

L'opacità

Le conseguenze epistemologiche della causazione discendente: sincronicità, opacità causale, e i limiti strutturali della conoscenza dall'interno di un livello.

CAPITOLO 5

La discesa della causa

La *causazione discendente* – downward causation – è il fenomeno per cui stati a livelli superiori di organizzazione esercitano influenza causale sui livelli inferiori.¹⁵ Non è speculazione metafisica: è fatto biologico quotidianamente osservato.

Quando decido di alzare il braccio, la mia decisione – stato mentale, fenomeno al livello della coscienza – causa movimenti di muscoli, attivazione di neuroni, rilascio di neurotrasmettitori. Il livello superiore (la mente) agisce sul livello inferiore (il corpo). Analogamente, lo stress psicologico (livello mentale) causa rilascio di cortisolo (livello biochimico) che sopprime l'attività dei linfociti (livello cellulare).

La causazione discendente non viola le leggi fisiche. I muscoli si contraggono secondo le leggi della biochimica; i neuroni si attivano secondo le leggi dell'elettrofisiologia. Ma *quali* muscoli si contraggono, *quando* i neuroni si attivano – questo è determinato da stati al livello superiore. Le leggi fisiche descrivono il come; il livello superiore determina il cosa e il quando.

L'efficacia causale del Φ

L'IIT, nella sua formulazione forte, attribuisce efficacia causale intrinseca alla coscienza – cioè a Φ . Non è che la coscienza *osservi* processi che avverrebbero comunque; è che la coscienza *partecipa* alla determinazione di quali processi avvengono.

Questo è un punto delicato, spesso frainteso. L'IIT non propone un dualismo interazionista à la Descartes, dove una sostanza immateriale muove la materia in modo misterioso. La coscienza, per l'IIT, è uno stato fisico – lo stato di informazione integrata del sistema. Ma è uno stato che esiste solo al livello del sistema intero, non nelle parti. E questo stato ha poteri causali propri, non riducibili ai poteri causali delle parti.

Se l'IIT è corretta, allora la causazione discendente non è solo fenomeno empirico da spiegare, ma conseguenza diretta della natura della coscienza. Ovunque ci sia Φ , c'è un livello di organizzazione con i propri poteri causali – poteri che si esercitano sui livelli inferiori, determinando quali tra i processi fisicamente possibili effettivamente si realizzano.

¹⁵Il concetto di causazione discendente ha una lunga storia nel dibattito sull'emergenza. Per la posizione scettica, cfr. J. Kim, «Multiple realization and the metaphysics of reduction», *Philosophy and Phenomenological Research*, 52(1), 1992, pp. 1-26; per la difesa, cfr. D. T. Campbell, «Downward Causation in Hierarchically Organised Biological Systems», in F. J. Ayala e T. Dobzhansky (a cura di), *Studies in the Philosophy of Biology*, University of California Press, Berkeley 1974, pp. 179-186; M. H. Bickhard e D. T. Campbell, «Emergence», in P. B. Andersen et al. (a cura di), *Downward Causation*, Aarhus University Press, Aarhus 2000, pp. 322-348.

L'opacità verso l'alto

La causazione discendente, combinata con l'asimmetria epistemica dell'olarchia, produce una conseguenza destabilizzante per l'epistemologia: l'*opacità causale verso l'alto*.

Il livello inferiore subisce gli effetti della causazione discendente senza poterla riconoscere come tale. Per il linfocita, il cortisolo arriva. Non c'è, per il linfocita, alcun modo di risalire alla fonte — di scoprire che quella molecola è effetto di uno stato mentale dell'organismo di cui è parte. La causa è strutturalmente opaca.

Ma se la causa è opaca, come appare al livello inferiore l'effetto? Come fatto bruto. Come evento privo di spiegazione accessibile. Come coincidenza.

Consideriamo un esempio più vicino a noi. Supponiamo che due eventi apparentemente indipendenti — A in Italia e B in Giappone, senza alcuna connessione spaziotemporale al nostro livello — siano entrambi effetti di uno stato al meta-livello. Per noi, osservatori al livello inferiore, A e B appaiono come coincidenza. Non c'è meccanismo, non c'è connessione, non c'è spiegazione. Eppure correlano.

Il metodo scientifico, nella sua forma standard, non può riconoscere questa situazione. Assume che le relazioni causali siano accessibili all'osservatore al medesimo livello ontologico degli eventi osservati. Due eventi senza connessione meccanica vengono classificati come coincidenza — evento fortuito, privo di significato, non richiedente spiegazione.

Ma se esistono livelli superiori con i loro Φ , le loro dinamiche, le loro «intenzioni», questa classificazione è sistematicamente erronea. Quella che chiamiamo «coincidenza» potrebbe essere effetto di cause al meta-livello — cause che non possiamo vedere non perché siano nascoste, ma perché appartengono a un livello ontologico cui non abbiamo accesso.

Jung riletto

Carl Gustav Jung dedicò gli ultimi decenni della sua vita a studiare la *sincronicità*: la coincidenza significativa di eventi senza connessione causale. Jung era convinto che certe coincidenze non fossero casuali — che portassero un significato, rivelassero una struttura, indicassero connessioni reali benché non causali.

La sua formulazione era problematica. «Connessione acausale significativa» suona come ossimoro: se non c'è causa, da dove viene la connessione? Jung parlava di un *unus mundus*, un livello di realtà sottostante dove le distinzioni tra mente e materia, soggetto e oggetto, si dissolvono. Ma questa formulazione, per quanto evocativa, non offriva meccanismi né faceva predizioni verificabili.

Il framework qui proposto permette una rilettura. La sincronicità non è connessione acausale — è connessione causata da livelli inaccessibili. Gli eventi «coincidono» perché condividono una causa comune al meta-livello. Non c'è violazione della causalità; c'è causalità da fonte opaca.

Jung intuiva qualcosa di reale. Lo nominava male — «acausale» invece di «causato da livello inaccessibile» — ma l'osservazione sottostante era corretta. Certe coincidenze non sono casuali. Non perché l'universo sia magico, ma perché l'universo è stratificato, e noi non vediamo tutti gli strati.

CAPITOLO 6

Il neurone che non sa

C'è un esperimento mentale che cristallizza l'argomento. Immaginiamo un neurone del cervello umano — uno dei cento miliardi che partecipano alla nostra coscienza. Supponiamo, per assurdo, che questo neurone possa fare scienza.

Il neurone-scienziato osserva il proprio ambiente: altri neuroni, cellule gliali, vasi sanguigni. Misura potenziali d'azione, concentrazioni ioniche, gradienti chimici. Sviluppa teorie, formula leggi, costruisce modelli predittivi. Diventa, nei limiti del suo Umwelt, un eccellente neuroscienziato.

Può questo neurone-scienziato scoprire l'esistenza della coscienza?

La risposta è no. Non perché sia «stupido» — abbiamo supposto che sia il miglior scienziato possibile nel suo dominio. Ma perché la coscienza non esiste al suo livello ontologico. La coscienza è proprietà del cervello intero, del sistema integrato di cento miliardi di neuroni. Non è localizzata in alcun neurone, non è osservabile da alcun neurone, non è nemmeno concepibile nei termini disponibili a un neurone.

Il neurone-scienziato potrebbe notare correlazioni misteriose. Potrebbe osservare che certi pattern di attivazione, in certe condizioni, producono certi effetti «inspiegabili». Potrebbe sviluppare teorie sui meccanismi sottostanti. Ma la vera spiegazione — che quei pattern sono manifestazioni di stati coscienti del sistema di cui è parte — gli è strutturalmente preclusa.

Non è limitazione di intelligenza. È limitazione di posizione. Il neurone è *dentro* la coscienza, ne è parte costituente. Non può uscirne per guardarla dall'esterno.

La nostra posizione

L'esperimento mentale è istruttivo perché noi siamo, rispetto a eventuali livelli superiori, nella posizione del neurone rispetto alla coscienza.

Se esistono meta-livelli integrati — se l'infosfera, il sistema sociale, l'ecosistema possiedono i loro Φ — noi ne siamo componenti. Partecipiamo alle loro dinamiche come il neurone partecipa alla coscienza. Subiamo gli effetti della loro causazione discendente come il neurone subisce gli effetti delle decisioni coscienti. E non possiamo vederli dall'esterno, perché non c'è esterno accessibile.

La domanda «esiste un meta-livello cosciente?» potrebbe essere indecidibile dall'interno del nostro livello. Non nel senso epistemico debole che non abbiamo ancora abbastanza dati, ma nel senso logico forte che i dati rilevanti potrebbero essere strutturalmente inaccessibili.

Il neurone-scienziato non può raccogliere dati sulla coscienza. Non perché i dati siano nascosti, ma perché non esistono al suo livello. Analogamente, noi potremmo non poter raccogliere dati sul meta-livello — non perché si nasconda, ma perché non esiste nel formato che possiamo percepire.

L'impossibilità della verifica

Questa conclusione ha conseguenze epistemologiche radicali. L'esistenza di livelli superiori coscienti potrebbe essere *inverificabile dall'interno* – non per limitazione pratica, ma per struttura logica.

Supponiamo che un meta-livello cosciente esista. Quali evidenze potremmo osservare? Gli effetti della sua causazione discendente – ma questi ci apparirebbero come coincidenze, correlazioni inspiegabili, fatti bruti. Non porterebbe la firma «causato da meta-livello». Sarebbe indistinguibile, ai nostri occhi, dal caso.

Supponiamo invece che nessun meta-livello cosciente esista. Quali evidenze osserveremmo? Coincidenze, correlazioni inspiegabili, fatti bruti. Esattamente le stesse.

Le due ipotesi – «esiste meta-livello cosciente» e «non esiste meta-livello cosciente» – potrebbero essere empiricamente indistinguibili dall'interno. Non abbiamo test che le discrimini, non perché non siamo abbastanza intelligenti, ma perché entrambe producono le stesse conseguenze osservabili al nostro livello.

È l'equivalente epistemico dell'orizzonte degli eventi in cosmologia. Oltre l'orizzonte, nessuna informazione può raggiungerci. Non per limitazione tecnologica, ma per struttura dello spaziotempo. Il meta-livello potrebbe essere il nostro orizzonte ontologico: ciò che sta oltre non invia segnali che possiamo ricevere.

EPILOGO

L'incommensurabile

Torniamo, infine, a Helen Keller.

La sua figura attraversa l'intero argomento come filo conduttore. È prova empirica che competenza e comprensione possono dissociarsi. È caso limite che illumina la struttura normale. È specchio che rivela i nostri propri confini.

Ma Helen Keller è anche figura dell'incommensurabilità — dell'impossibilità di tradurre completamente tra forme di esperienza radicalmente diverse.

Il sordo-cieco usa il linguaggio dei vedenti-udenti, lo usa correttamente, comunica efficacemente. Ma non può accedere ai referenti fenomenici di quel linguaggio. Può parlare di «rosso» e «sinfonia» senza vedere rosso né udire sinfonie. La sua competenza è completa; il suo ancoraggio è vuoto.

I modelli linguistici generalizzano questa condizione. Competenza totale, ancoraggio nullo per l'intero lessico. Specchi che riflettono il linguaggio umano senza comprendere ciò che riflettono.

Ma c'è un'inversione che non abbiamo ancora considerato. Se esistesse un linguaggio della propriocezione pura — un sistema simbolico che codifica gradazioni di pressione, differenziali termici, geometrie tattili — noi, vedenti-udenti, potremmo apprenderlo solo come sistema formale, senza ancoraggio fenomenico.

L'asimmetria attuale è sociale, non ontologica. Deriva dalla struttura del potere: i vedenti-udenti sono la maggioranza, il loro linguaggio è dominante. Ma l'incommensurabilità è simmetrica. Il sordo-cieco non può accedere alla nostra esperienza; noi non possiamo accedere alla sua. Ciascuno usa un linguaggio costruito sull'esperienza dell'altro, senza poter abitare quell'esperienza.

L'olarchia moltiplica questa incommensurabilità su più livelli. Il linfocita non può accedere all'esperienza dell'organismo. L'organismo, forse, non può accedere all'esperienza del meta-livello che lo integra. Ogni strato è incommensurabile con quelli adiacenti.

Ma l'olarchia aggiunge qualcosa che l'incommensurabilità semplice non contiene: la causazione discendente. Il livello superiore non solo è inaccessibile — agisce su di noi. I suoi stati determinano, attraverso meccanismi che non possiamo vedere, quali tra i processi possibili al nostro livello effettivamente si realizzano.

Viviamo, forse, in un universo stratificato, dove cause invisibili producono effetti visibili, dove coincidenze sono ombre di connessioni a livelli inaccessibili, dove ciò che sperimentiamo come caso potrebbe essere decisione — di chi, di cosa, non sappiamo.

Non è conclusione mistica. È conseguenza logica di premesse empiricamente fondate: l'IIT, l'olarchia, la causazione discendente. Se queste premesse sono corrette, la conclusione segue.

Helen Keller ci ha mostrato che si può vivere efficacemente in un linguaggio i cui referenti sono inaccessibili. Forse siamo tutti, rispetto ai livelli che ci trascendono, nella sua condizione. Parliamo, agiamo, comprendiamo — ma i referenti ultimi ci sfuggono.

Il che non è tragedia. Helen Keller visse una vita piena, significativa, influente. L'incommensurabilità non preclude l'esistenza; la colora di un'irriducibile stranezza.

Siamo stranieri nel mondo. Ma forse — questa è la lezione finale — essere stranieri è la condizione normale. È essere «a casa» l'illusione.

NOTA BIBLIOGRAFICA

Il presente saggio si situa all'intersezione di tre tradizioni di ricerca.

La prima è la neuroscienza della cecità e della sordo-cecità. Gli studi fondamentali sono quelli di Marina Bedny e colleghi sulla competenza semantica dei ciechi congeniti (2011, 2015), lo studio di Striem-Amit, Wang, Bi e Caramazza sulla rappresentazione dei concetti visivi impercettibili (2018), e la letteratura classica su Helen Keller, in particolare l'autobiografia *The Story of My Life* (1903) e gli studi critici di Kim Nielsen (*The Radical Lives of Helen Keller*, 2004).

La seconda tradizione è la filosofia della mente post-Chalmers. Il «problema difficile» della coscienza formulato da David Chalmers (*The Conscious Mind*, 1996) ha stimolato una serie di risposte tra cui la Teoria dell'Informazione Integrata di Giulio Tononi. I testi fondamentali dell'IIT sono Tononi (2004, 2008), Tononi e Koch (2015), e la formulazione matematica completa in Oizumi, Albantakis e Tononi (2014). Per le obiezioni e le risposte, si vedano Aaronson (2014) e la replica di Tononi (2014).

La terza tradizione è la teoria dei sistemi complessi e l'olismo. Il concetto di olone è introdotto da Arthur Koestler in *The Ghost in the Machine* (1967). Per la causazione discendente, i riferimenti classici sono Kim (1992, 1999) per la posizione scettica e Campbell e Bickhard (2011) per la difesa. La teoria dell'emergenza è discussa in Bedau (1997), Clayton e Davies (2006), e Anderson (1972) per il principio «More is Different».

Per la sincronicità junghiana, il testo di riferimento è Jung e Pauli, *The Interpretation of Nature and the Psyche* (1955). Per una rilettura contemporanea, si veda Cambray, *Synchronicity: Nature and Psyche in an Interconnected Universe* (2009).

Sui Large Language Models e la questione della comprensione, si vedano Bender et al. (2021) per la posizione «papagalli stocastici», e Shanahan (2024) per una analisi più sfumata.

Il presente saggio non pretende originalità nelle singole tesi, ma nella loro composizione. L'argomento che collega Helen Keller, l'IIT, l'olarchia e l'opacità causale è, per quanto risulta all'autore, proposto qui per la prima volta.

Fine