Risonanza cognitiva

Due menti che sbagliano in versi opposti possono valere, insieme, molto più di ciascuna. Nel modello del turing la coppia segue la stessa curva di un circuito risonante, con la fiducia a fare da manopola, e accanto al picco c'è un punto in cui due menti capaci, insieme, tirano a indovinare. Lo abbiamo messo alla prova su reti vere, con un protocollo scritto prima: con pregiudizi opposti si arriva a 7,8 turing, come previsto. Costruire la controfase addestrando una macchina sugli errori di una mente è un'altra storia: una ricetta fallisce, una funziona, e la macchina più brava da sola resta il compagno migliore.

endjneer · paithon.it · 11 novembre 2026 · versione 1
Articolo: https://paithon.it/risonanza-cognitiva/
Codice e dati: github.com/paithon-it/cognitive-resonance · doi.org/10.5281/zenodo.24140357
Testo con licenza CC BY 4.0.

Chi ha girato almeno una volta la manopola di una vecchia radio conosce il fenomeno prima di conoscerne il nome. Si gira, e c'è solo fruscio. Poi, in un tratto di pochi millimetri, una voce esce nitida; un altro millimetro, ed è di nuovo fruscio. Dietro la manopola c'è un circuito con due componenti che reagiscono in modi opposti: una bobina, che si oppone al segnale tanto più quanto più è rapido, e un condensatore, che fa il contrario. A una sola frequenza le due opposizioni sono uguali, e si annullano. Per quella frequenza, e per un tratto strettissimo intorno, il circuito risponde molte volte più forte che per tutte le altre: la stazione che trasmette lì, debolissima all'antenna, esce dal fruscio. Si chiama *risonanza*.

La risonanza ha una proprietà che la rende interessante anche fuori dall'elettrotecnica: **non aggiunge energia**. La voce era già nell'aria, coperta. Due difetti di segno opposto, messi nello stesso circuito e regolati l'uno sull'altro, tolgono di mezzo ciò che la copriva.

Nell'articolo precedente, «Intelligenza Apparente», abbiamo proposto un'unità di misura, il *turing*, per dire quanto una macchina aggiunge a chi deve decidere. Un turing vuol dire che con la macchina si vale quanto da soli. Due turing vogliono dire che è raddoppiata la capacità di distinguere la risposta giusta da quella sbagliata, non il numero delle risposte giuste: chi da solo, fra due risposte, ne indovina 60 su cento, a due turing arriva a 69. Insieme all'unità abbiamo introdotto un angolo, quello fra gli errori della persona e quelli della macchina. Su quattro raccolte pubbliche di giudizi umani l'angolo stava in mediana fra 59 e 81 gradi: persone e macchine sbagliano in parte sugli stessi casi. A 90 gradi sbaglierebbero su casi diversi, e la coppia ideale varrebbe ciò che dice il teorema di Pitagora. Di ciò che sta oltre abbiamo dato soltanto l'annuncio.

Questo articolo va dall'altra parte dell'angolo retto. Che cosa succede se gli errori di una persona e quelli di una macchina non cadono soltanto su casi diversi, ma sono opposti? La formula dell'articolo precedente ha già una risposta: quando l'angolo va verso i 180 gradi, i turing che la coppia promette crescono senza un tetto. È la condizione che abbiamo chiamato **risonanza cognitiva**. Il

nome non è una metafora: la curva che descrive la coppia è, simbolo per simbolo, quella del circuito dietro la manopola.

Le domande sono tre. Se la formula mantenga la promessa su reti vere e immagini vere, o soltanto sulla carta. Se la controfase si possa costruire, addestrando una macchina sugli errori di una mente. E dove stiano, su questa scala, le coppie che esistono. Il protocollo, con le ipotesi e i criteri per dirle confermate o smentite, è stato scritto e registrato prima degli esperimenti, e le aggiunte fatte dopo sono dichiarate come tali: ciò che non è passato sta scritto accanto a ciò che è passato.

In breve

- **L'idea.** Quando gli errori di due menti vanno in versi opposti, la formula dei turing apparenti promette più di Pitagora, e verso i 180 gradi non ha un tetto. Al variare della fiducia data alla macchina la coppia segue la curva di un circuito risonante: la fiducia è la sintonia, e l'angolo decide quanto il picco è alto e stretto.
- **Il nullo.** La stessa curva ha uno zero: un peso a cui due menti capaci, insieme, tirano a indovinare. Chiede di fidarsi al contrario, oppure una macchina che su certi casi sbaglia per sistema.
- **Il banco.** Su 140 coppie di reti con un pregiudizio fabbricato, i turing misurati stanno sulla curva prevista fino a 164 gradi e **7,8 turing**, e il nullo sta dove deve. Gli undici criteri fissati prima di guardare i dati passano tutti.
- **Il limite.** Fra le coppie in controfase i turing più alti sono di quelle che in assoluto ne sanno meno: la risonanza restituisce ciò che il pregiudizio copriva, non aggiunge niente.
- **Costruire la controfase.** Punire una macchina perché somiglia alla persona non funziona: capovolge tutto. Addestrarla sull'errore della somma sì: 132 gradi, 1,37 turing, anche con una macchina che del compito non sa niente. Ma la macchina più brava da sola resta il compagno migliore, anche quando è la macchina stessa a imparare la correzione: sul banco la persona non sa niente che la macchina non veda.
- **Le coppie vere.** Con i giudizi umani la punizione apre l'angolo da 68 a 75 gradi e i turing apparenti scendono; la ricetta della somma non sposta niente. Le coppie misurate finora non sono in risonanza.

Oltre l'angolo retto

Due periti stimano lo stesso quadro. Se sbagliano ciascuno per conto suo, la media delle due stime sbaglia meno di ciascuna: a parità di bravura, di un fattore $\sqrt{2}$. Se sbagliano nello stesso verso, la media non corregge niente. Se uno sbaglia sempre per eccesso e l'altro per difetto, della stessa quantità, la media è esatta: l'errore del primo è l'errore del secondo con il segno cambiato, e nella somma sparisce. Sbagliare su casi diversi e sbagliare in versi opposti non sono la stessa cosa: nel primo caso gli errori si compensano in media, nel secondo uno per uno.

Fra una persona e una macchina un caso così, oggi, nessuno l'ha misurato, ma si può immaginare. Un radiologo guarda una lastra dopo aver letto la cartella del paziente, e la cartella lo tira: con una storia clinica pesante vede un po' più malattia di quanta ce n'è, con una leggera un po' meno. Un programma che legge la stessa lastra senza la cartella sbaglia per ragioni sue, per esempio per i difetti dell'apparecchio. Sono errori su casi diversi: è l'angolo retto, e insieme i due valgono Pitagora. Per andare oltre servirebbe un programma che la cartella la legge, e ne è tirato nel verso opposto: davanti a una storia pesante tende a vedere un po' meno malattia. Da solo sarebbe un difetto. **In coppia è ciò che cancella la spinta della cartella.**

Per passare dai periti a un numero basta un'idea. Chi deve dire «cane» o «gatto» davanti a una fotografia non ha in testa una risposta secca: ha un'impressione più o meno forte, un «quanto cane», e risponde secondo il lato da cui l'impressione pende. L'impressione in parte sbaglia, e in

una coppia conta come sbagliano *insieme* le due impressioni. Se quando la prima pende troppo verso «cane» tende a farlo anche la seconda, gli errori sono allineati. Se l'una non dice niente dell'altra, sono indipendenti. Se quando la prima esagera verso «cane» la seconda esagera verso «gatto», sono opposti. Questa tendenza è una correlazione, un numero fra −1 e 1, e l'angolo è un altro modo di scriverla: **0 gradi per gli stessi errori, 90 per errori indipendenti, 180 per errori opposti**. Di una rete l'impressione si legge direttamente: è il numero che calcola prima di rispondere. Di una persona si vede soltanto la risposta, e l'angolo si ricostruisce contando quante volte due giudici sbagliano sugli stessi casi, rispetto a quante ne spiegherebbe il caso.

In simboli. L'impressione è un *indizio* graduato, disturbato dal rumore. La bravura di chi decide è la *sensibilità* d', che vale zero per chi tira a indovinare e non ha un tetto. Una persona e una macchina hanno sensibilità d_H e d_M sugli stessi casi; $k = d_M / d_H$ dice quante volte la macchina è più sensibile, ρ è la correlazione fra i rumori dei loro indizi, e l'angolo è $\varphi = \arccos \rho$. Il **turing**, Tu, è il rapporto fra la sensibilità della persona con la macchina e quella della stessa persona da sola. La coppia ideale, che mette in comune tutti gli indizi e li pesa come vanno pesati, promette

$$\mathrm{Tu}_{app} = \frac{\sqrt{1 + k^2 - 2k\cos\varphi}}{\sin\varphi}$$

turing apparenti. A 90 gradi il coseno si annulla e resta $\sqrt{1 + k^2}$, Pitagora. Sotto i 90 la promessa si accorcia, perché una parte di ciò che dice il secondo ripete il primo (fino all'angolo utile dell'articolo precedente: sotto quell'angolo la formula risale, ma con un peso negativo). Sopra i 90 succede l'inverso: il numeratore cresce, il denominatore torna verso zero, e **a 180 gradi la frazione non ha più un valore finito**.

> **I precedenti.** Due stime che cadono ai lati opposti del vero le hanno studiate Richard Larrick e Jack Soll nel 2006, e hanno chiamato la cosa *bracketing*: in quel caso la media fa sempre meglio del giudice medio, anche se non sempre del migliore dei due. Nei codici dei neuroni la regola si chiama *sign rule*: un rumore correlato con il segno opposto a quello del segnale aumenta l'informazione, fino al caso limite in cui il codice vale quanto uno senza rumore (Hu, Zylberberg e Shea-Brown, 2014). Negli insiemi di reti è la *negative correlation learning* di Liu e Yao (1999). La combinazione lineare di due indizi con errori correlati, con il suo peso migliore, l'hanno scritta Oruç, Maloney e Landy nel 2003. La statistica conosce il fenomeno dal 1941 con un altro nome, *variabile soppressore* (Paul Horst): un predittore che migliora la previsione perché è legato all'errore di un altro. Friedman e Wall, nel 2005, hanno disegnato le zone in cui due predittori insieme spiegano più della somma di ciò che spiegano da soli, che in quella letteratura si chiamano *enhancement*: è l'oltre Pitagora, visto dalla regressione.

Nessuna coppia è esattamente così. Resta sempre una parte di errore che i due non hanno in controfase, e quella parte fissa fin dove si arriva. La tabella 1 dà la misura per due menti di pari sensibilità, per cui la formula si riduce a $1/\cos(\varphi/2)$. L'ultima colonna usa una lettura dell'articolo precedente: il quadrato del turing dice quante copie della persona, con errori indipendenti e indizi messi in comune, servirebbero per fare quanto la coppia.

Tabella 1. Quanto promette una coppia, secondo l'angolo
due menti di pari sensibilità ($k = 1$), dentro il modello; l'ultima colonna è il quadrato dei turing

Angolo fra gli errori	Che cosa vuol dire	Turing apparenti	Copie indipendenti equivalenti
0 gradi	Gli stessi errori	1,00	1
60 gradi	Dove stanno, più o meno, le coppie misurate	1,15	1,3
90 gradi	Errori su casi diversi: Pitagora	1,41	2
120 gradi	Controfase parziale	2,00	4
150 gradi	Controfase forte	3,86	15
170 gradi	Quasi opposti	11,5	132

Lo stesso conto si può fare con i numeri veri dell'articolo precedente. Su domande d'esame GPT-4o valeva 2,16 volte un gruppo di persone, e i loro errori stavano a 75 gradi: la coppia ideale prometteva 2,21 turing, appena il 2 per cento più della macchina da sola. Con le stesse persone e la stessa macchina, ma a 120 gradi, ne prometterebbe **3,23**: una volta e mezza la macchina da sola. Non servirebbero una macchina o delle persone più brave. Servirebbe sbagliare in un altro modo.

Un turing alto non dice che la coppia è brava. Dice quanto vale rispetto alla persona da sola: cresce anche se la coppia resta ferma, quando è la persona da sola a valere poco. E la controfase, come il circuito della radio, non aggiunge niente: toglie di mezzo l'errore che copriva ciò che i due sapevano già. Anche la sensibilità va letta con cura, perché non è una percentuale: fra due risposte, ogni punto di sensibilità in più rende meno del precedente (tabella 2).

Tabella 2. Sensibilità e risposte giuste
fra due risposte ugualmente probabili, per chi non ne preferisce una; dentro il modello

Sensibilità	0	0,5	1	2	3	4
Risposte giuste	50%	60%	69%	84%	93%	98%

La manopola

Fin qui la coppia era ideale: pesava gli indizi nel modo migliore. Una coppia vera li pesa come può, e conviene chiedersi che cosa succede quando il peso cambia. Sia w il peso che la persona dà all'indizio della macchina rispetto al proprio: zero per chi la ignora, molto grande per chi ne ricopia le risposte. Nel modello il turing è

$$\mathrm{Tu}(w)^2 = \frac{(1 + wk)^2}{1 + w^2 + 2\rho w}$$

e ha un massimo, che sono i turing apparenti. Per due menti di pari sensibilità la formula si lascia riscrivere in un modo che un elettrotecnico riconosce al primo sguardo. Con $\Omega = \sqrt{w}$,

$$\mathrm{Tu}^2 - 1 = \frac{\tan^2(\varphi/2)}{1 + Q^2\left(\Omega - \dfrac{1}{\Omega}\right)^2} \qquad \text{con} \qquad Q = \frac{1}{2\cos(\varphi/2)}$$

È la **curva di risonanza** di un circuito con una resistenza, una bobina e un condensatore in serie: la stessa funzione, con la radice del peso al posto della frequenza e con un *fattore di merito* Q che dipende soltanto dall'angolo. A sinistra c'è $\mathrm{Tu}^2 - 1$, che nell'articolo precedente era il peso dell'evidenza che la macchina aggiunge, in unità della persona. Al picco il turing vale $2Q$ (figura 1).

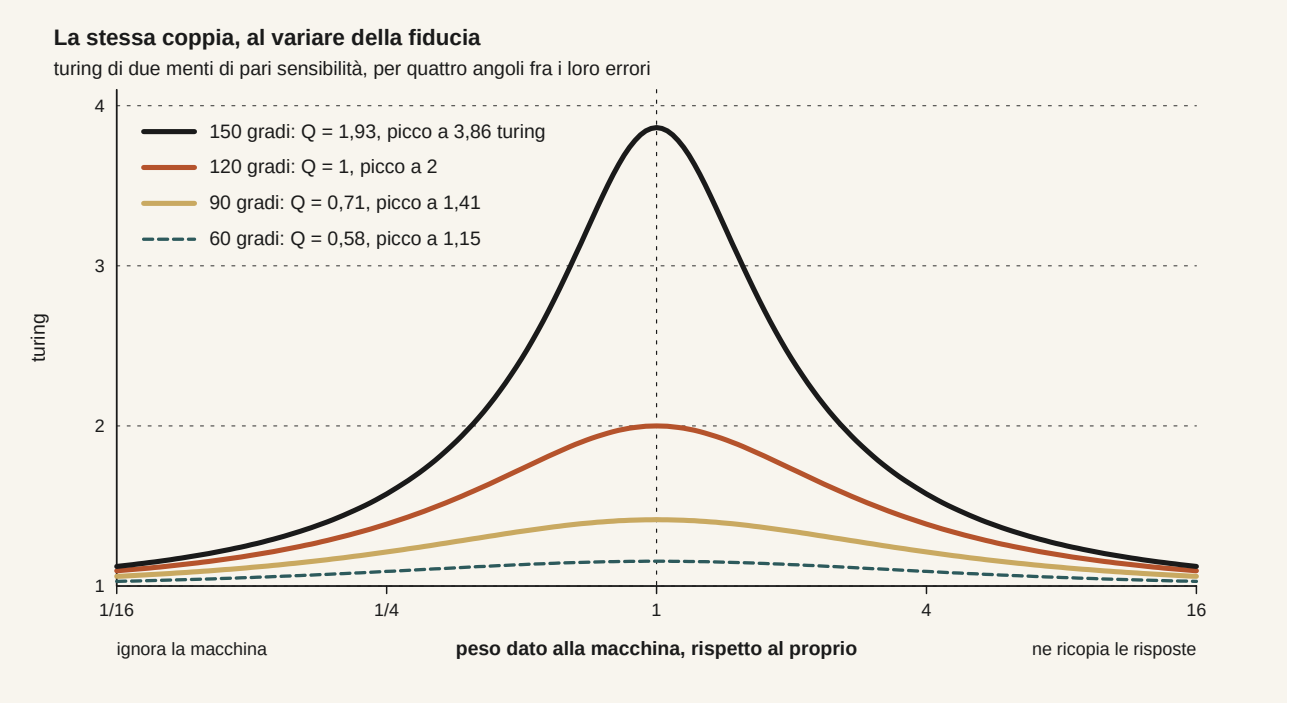


Figura 1. Il turing della coppia al variare del peso dato alla macchina, per quattro angoli: la curva di risonanza.

Tabella 3. La corrispondenza
che cosa diventa ogni grandezza del circuito quando al posto di bobina e condensatore ci sono due menti; vale esattamente per due menti di pari sensibilità, dentro il modello

Nel circuito	Nella coppia
La frequenza, rispetto a quella di risonanza	La radice del peso dato alla macchina, rispetto al proprio
Il fattore di merito Q	$1/(2\cos(\varphi/2))$: lo fissa l'angolo fra gli errori
Lo smorzamento critico, $Q = 1/2$	Angolo zero: stessi errori, nessun guadagno
L'altezza del picco	$2Q$ turing
La larghezza di banda	Di quanto si può sbagliare il peso prima di perdere metà del guadagno
La resistenza, che limita il picco	La parte di errore che i due non hanno in controfase

Dalla corrispondenza (tabella 3) vengono tre conseguenze, e nessuna delle tre è ovvia senza la curva.

La fiducia è la sintonia. Chi ignora la macchina e chi ne ricopia le risposte stanno alle due code della stessa curva: nel primo caso la coppia vale un turing, nel secondo quanto la macchina. Il guadagno sta in mezzo, e si prende solo dando alla macchina il peso giusto. Nell'articolo precedente i gruppi di persone realizzavano una parte della promessa, e ora c'è un modo di dire perché: la manopola c'era, ed era girata male.

Più alto il picco, più stretta la banda. A 90 gradi il guadagno resta sopra la metà del massimo per pesi che vanno da 0,27 a 3,73: ci si può fidare della macchina quasi quattro volte troppo, o quasi quattro volte troppo poco, e si perde al più metà di ciò che si poteva avere. A 150 gradi la banda va da 0,60 a 1,67. A 170 gradi da 0,84 a 1,19: basta fidarsi un quinto di troppo. Una coppia in risonanza vale molto, e chiede una fiducia tarata con una precisione che cresce insieme al premio.

Niente si crea. La somma di due indizi con errori opposti non contiene più segnale di quanto ne avesse ciascuno: contiene meno rumore. Il turing sale perché il denominatore, la persona da sola, è coperto dal suo errore, e la coppia no. Lo avevamo scritto anche nell'articolo precedente, a proposito di un turing che saliva quando chi faceva da campione non vedeva più niente: un turing si legge accanto a ciò che il campione fa da solo.

> **Una cosa a due.** Tre menti non possono essere tutte in controfase fra loro: se la prima sbaglia al contrario della seconda, e la seconda al contrario della terza, la prima e la terza sbagliano insieme. Per n menti che si somigliano tutte allo stesso modo la correlazione non può scendere sotto $-1/(n-1)$. Una folla si avvicina al più all'indipendenza; la risonanza è un fatto di coppia. Va un po' meglio a stella: una mente al centro e m macchine intorno, ciascuna in controfase con lei su un pregiudizio diverso e indipendente dalle altre. Lì la correlazione di ciascuna con il centro può scendere fino a $-1/\sqrt{m}$: tutte insieme cancellano al più l'intero errore del centro, una volta sola.

Tutto questo, va ripetuto, vale dentro un modello: indizi gaussiani, messi insieme in modo lineare. L'identità è un teorema del modello. Se due menti vere, o due reti vere, si comportino così è una domanda per un esperimento.

Il nullo

Una curva di sintonia non ha soltanto un massimo. Chi ha cercato una stazione in onde medie ruotando una radiolina lo ha sentito: c'è un orientamento in cui la voce sparisce, ed è molto più netto di quello in cui è più forte. I radiogoniometri, per trovare la direzione di una trasmittente, cercavano quel silenzio. Fin qui il peso era positivo: della macchina ci si fidava, poco o tanto. Lasciamo che la manopola giri anche dall'altra parte. Un peso negativo vuol dire usare l'indizio della macchina al contrario: credere a una risposta tanto più quanto più la macchina la sconsiglia.

Non è lo stesso che non fidarsi. Chi non si fida la ignora: il peso è zero, e resta ciò che la persona valeva da sola, un turing. Chi la usa al contrario sottrae. E c'è una misura in cui ciò che sa la macchina cancella esattamente ciò che sa la persona: se la macchina è due volte più sensibile, basta sottrarne la metà. Della somma resta soltanto il rumore. Due menti che da sole riconoscono la risposta giusta, **messe insieme con quel peso, tirano a indovinare**.

Con pesi di ogni segno la curva intera ha una forma che i fisici conoscono bene, e questa volta vale per qualunque k, non solo a pari sensibilità:

$$\mathrm{Tu}^2 = k^2\,\frac{(q+\varepsilon)^2}{1+\varepsilon^2} \qquad \text{con} \qquad \varepsilon = \frac{w+\cos\varphi}{\sin\varphi}, \quad q = \frac{1-k\cos\varphi}{k\sin\varphi}$$

Fano. Ugo Fano, torinese, la scrisse una prima volta nel 1935 sul *Nuovo Cimento*, a ventidue anni, mentre lavorava a Roma nel gruppo di Enrico Fermi, per spiegare certe righe asimmetriche nello spettro dei gas nobili; la forma che si usa oggi è del 1961. Descrive che cosa succede quando una stessa cosa può avvenire per due vie che interferiscono: da un lato della riga le due vie si sommano, dall'altro si cancellano. Lo zero sta a $\varepsilon = -q$; il massimo, che vale $1+q^2$, a $\varepsilon = 1/q$.

È il *profilo di Fano*: un picco e uno zero nella stessa curva. Il picco è la risonanza, e vale i turing apparenti. Lo zero è **il nullo**, e sta al peso $w = -1/k$. Oltre il nullo il turing cambia segno. Un turing negativo non è un artificio del conto: la coppia sa ancora qualcosa, ma lo usa a rovescio, e sbaglia più spesso del caso. Basterebbe capovolgerne le risposte per rimetterla dalla parte giusta (figura 2).

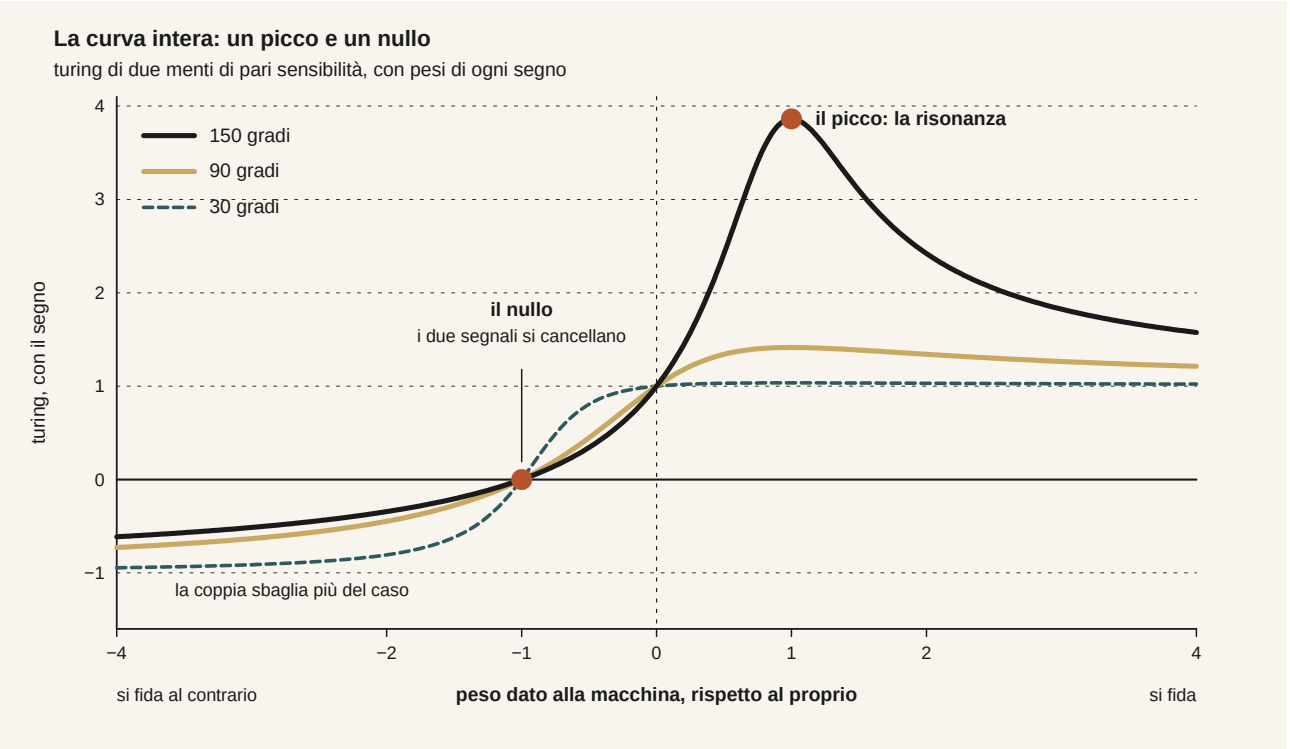


Figura 2. Picco e nullo nella stessa curva: il turing con il segno, anche per pesi negativi.

Del nullo conviene sapere quattro cose.

C'è sempre. Il picco dipende dall'angolo, e può essere alto o quasi invisibile. Il nullo c'è per qualunque angolo e per qualunque macchina che del compito sappia qualcosa, ed è sempre uno zero esatto. Le eccezioni sono due casi limite: due menti identiche in tutto, e una macchina che del compito non sa niente, per cui il peso $-1/k$ non esiste.

Ci si arriva per due strade, e conviene tenerle distinte: nella prima a essere capovolta è la fiducia, nella seconda la macchina.

Nella prima la macchina funziona, e la persona fa il bastian contrario: più la macchina dice «cane», più la persona propende per «gatto». Con la giusta misura la coppia arriva a zero, e con una macchina forte la misura è piccola. Accanto a una macchina dieci volte più sensibile di chi la

usa il nullo sta a un peso di meno un decimo: basta contraddirla poco, ma sempre, per cancellare insieme ciò che sa la macchina e ciò che sa la persona.

Nella seconda la persona si fida nel verso normale, ed è la macchina a essere capovolta: su una famiglia di casi sbaglia per sistema, più spesso di chi tira a sorte. In simboli k è negativo, e il nullo passa fra i pesi positivi, a $1/|k|$. Chi si fida nella misura sbagliata di una macchina che su certi casi ha un difetto, su quei casi vale zero, e chi si fida di più sbaglia insieme a lei (tabella 4).

Tabella 4. Le due strade verso il nullo
in tutte e due il peso che annulla la coppia è $-1/k$; cambia chi dei due è capovolto

Chi è capovolto	La macchina, da sola	La fiducia della persona	Dove sta il nullo
La fiducia	Fa meglio del caso: k positivo	Al contrario: peso negativo	Al peso $-1/k$
La macchina	Fa peggio del caso: k negativo	Nel verso normale: peso positivo	Al peso $1/\|k\|$

La sua larghezza va al contrario di quella del picco. Con errori simili il nullo è una fessura: a 30 gradi, per due menti pari, il turing resta sotto il mezzo solo fra i pesi −1,35 e −0,74. Con errori opposti è una conca: a 150 gradi, fra −2,90 e −0,35. La stessa controfase che alza il picco e lo stringe, dall'altra parte allarga la buca.

L'angolo utile è il punto in cui resta solo il nullo. Il numero q della formula dice quanto la curva somiglia a un picco. Quando è grande il nullo è lontano, e si vede soltanto la risonanza. Quando vale zero il picco non c'è più: la curva sale fino a k senza superarlo, e l'unica cosa notevole è lo zero. Succede quando $\cos\varphi = 1/k$, che nell'articolo precedente avevamo chiamato angolo utile senza sapere che fosse questo.

Sotto quell'angolo q diventa negativo e il picco passa dalla parte dei pesi negativi. È il servizio che allora avevamo escluso: usare i propri errori per togliere alla macchina il rumore che si ha in comune con lei.

> **Quanto vale q nelle coppie già misurate.** In unità della macchina i turing apparenti sono $\sqrt{1+q^2}$. Con GPT-4o, su domande d'esame, q valeva 0,21, e il margine della coppia sulla macchina era il 2 per cento; con Llama 3.1 8B valeva 0,61, e il margine il 17. Per due menti di pari sensibilità q è la tangente di metà dell'angolo.

Dove stanno le coppie che esistono

Prima di costruire qualcosa conviene guardare ciò che c'è. Gli angoli misurati nell'articolo precedente si possono rileggere sulla scala nuova: il fattore di merito dipende soltanto dall'angolo, e l'altezza del picco dice quanto varrebbe la coppia se le due menti avessero la stessa sensibilità e pesassero gli indizi nel modo migliore (tabella 5).

Tabella 5. Le coppie già misurate, sulla scala della risonanza
angoli mediani da «Intelligenza Apparente»; il fattore di merito è $1/(2\cos(\varphi/2))$, il picco è il turing che la coppia prometterebbe a pari sensibilità. È una rilettura, senza dati nuovi

Coppia	Angolo	Fattore di merito	Picco, in turing
Due reti per la visione fra loro	22-43	0,51-0,54	1,02-1,07
Due persone fra loro, storie di senso comune	54	0,56	1,12
Persone e reti, riconoscere oggetti	59	0,57	1,15
Persone e macchine, storie di senso comune	64	0,59	1,18
Persone e macchine, sarcasmo e censimento	65	0,59	1,19
Due persone fra loro, domande d'esame	66	0,60	1,19
Persone e macchine, arte, città, pelle	78	0,64	1,29
Persone e macchine, domande d'esame	81	0,66	1,32
Errori indipendenti, per confronto	90	0,71	1,41
Dove due menti pari valgono il doppio	120	1,00	2,00

Tutte le mediane stanno **fra 0,51 e 0,66**: poco sopra lo smorzamento critico, e sotto il valore dei 90 gradi. Un circuito con un fattore di merito così un elettrotecnico non lo chiamerebbe risonante. Per arrivare a uno, dove due menti di pari sensibilità valgono due turing, servirebbero 120 gradi, e nessuna delle famiglie di dati dell'articolo precedente ci si avvicina. Qualche confronto isolato passa l'angolo retto, e uno arriva a 103 gradi, ma con intervalli che lo comprendono. La risonanza non si trova in natura fra le coppie che abbiamo misurato: se la si vuole, va costruita.

Il banco: due pregiudizi opposti

Per mettere alla prova la formula servono due menti di cui si possa regolare l'angolo a piacere. Con le persone non si può fare. Con le reti sì, a patto di accontentarsi di un banco di prova: reti piccole, fotografie vere, e un pregiudizio fabbricato apposta.

Il compito è distinguere due classi di fotografie che si confondono: gatto e cane, cervo e cavallo, automobile e camion, aereo e nave, uccello e rana. Sono le immagini di CIFAR-10, 32 pixel di lato. Ogni fotografia porta una cornice di due pixel, più chiara o più scura. Durante l'addestramento la cornice è un indizio: per una rete, cornice chiara vuol dire più spesso «cane»; per l'altra, più spesso «gatto». Sono due *pregiudizi* nel senso letterale, giudizi formati prima di guardare la fotografia, e hanno verso opposto. Alla prova la cornice non ha più niente a che fare con la classe, e chi le dà retta sbaglia. Un numero regola la forza del pregiudizio, cioè quanto la cornice era affidabile in addestramento: a zero non diceva niente, a 2 bastava da sola a indovinare 98 volte su cento (figura 3).

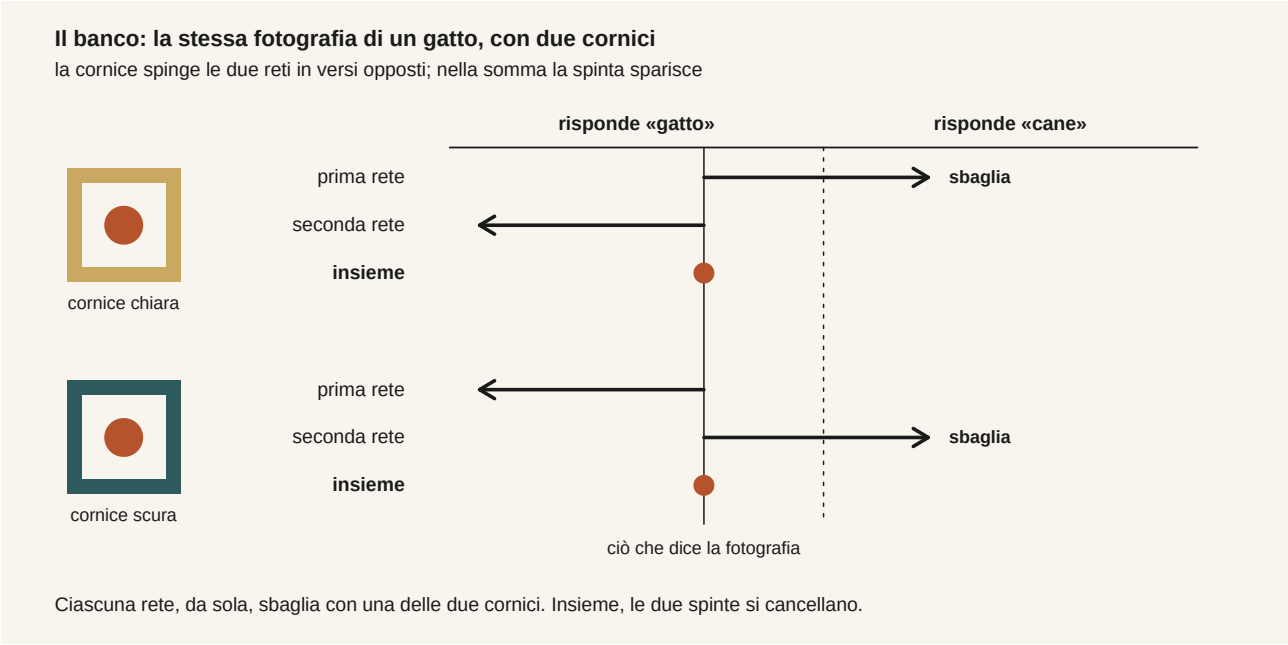


Figura 3. Il banco: la stessa fotografia con due cornici. Ciascuna rete sbaglia con una sola delle due cornici; nella somma le spinte opposte si cancellano.

Chiamiamo persona la prima rete e macchina la seconda, ma sono due reti uguali, addestrate sulle stesse fotografie: **il banco prova la formula, non dice niente delle persone**. Accanto a ogni coppia di pregiudizi opposti ce n'è una di controllo, con due reti che hanno lo stesso pregiudizio. Cinque compiti, sette forze, due addestramenti indipendenti: settanta coppie opposte e settanta uguali, ciascuna misurata su 10.000 presentazioni (tabella 6).

Tabella 6. Che cosa succede quando il pregiudizio cresce
coppie con pregiudizi opposti; mediane su cinque compiti e due addestramenti. Il turing è un rapporto: sale anche mentre le risposte giuste insieme scendono, perché quelle della rete da sola scendono di più. Le risposte giuste insieme sono quelle della somma dei due indizi alla pari, portati prima alla stessa scala; i turing previsti vengono dall'angolo e dal rapporto fra le sensibilità misurati sulla stessa coppia

Forza del pregiudizio	Angolo fra gli errori	Giuste da sola	Giuste insieme	Turing misurati	Turing previsti
0	23 gradi	95,6%	95,7%	1,02	1,02
0,25	26	95,4%	95,8%	1,03	1,03
0,5	35	95,1%	95,7%	1,05	1,05
0,75	53	93,5%	95,3%	1,14	1,12
1	76	89,9%	94,7%	1,31	1,28
1,5	122	74,1%	92,5%	2,18	2,14
2	149	59,7%	86,6%	3,94	3,88

Le due colonne delle risposte giuste vanno lette insieme. La coppia scende dal 95,7 all'86,6 per cento, ma ciascuna rete, da sola, precipita dal 95,6 al 59,7. **Il turing misura lo scarto fra le due discese, non la bravura della coppia**: per questo sale mentre tutte e due peggiorano.

Senza pregiudizio le due reti stanno a 23 gradi: sono quasi la stessa rete, e metterle insieme non serve. Alla prova la cornice inganna ciascuna, e l'inganno ha verso opposto nelle due, quindi

l'angolo si apre: 76 gradi, poi 122, poi **149**. Alle due forze più alte tutte e venti le coppie hanno l'angolo sopra i 90 gradi, intervallo di confidenza compreso.

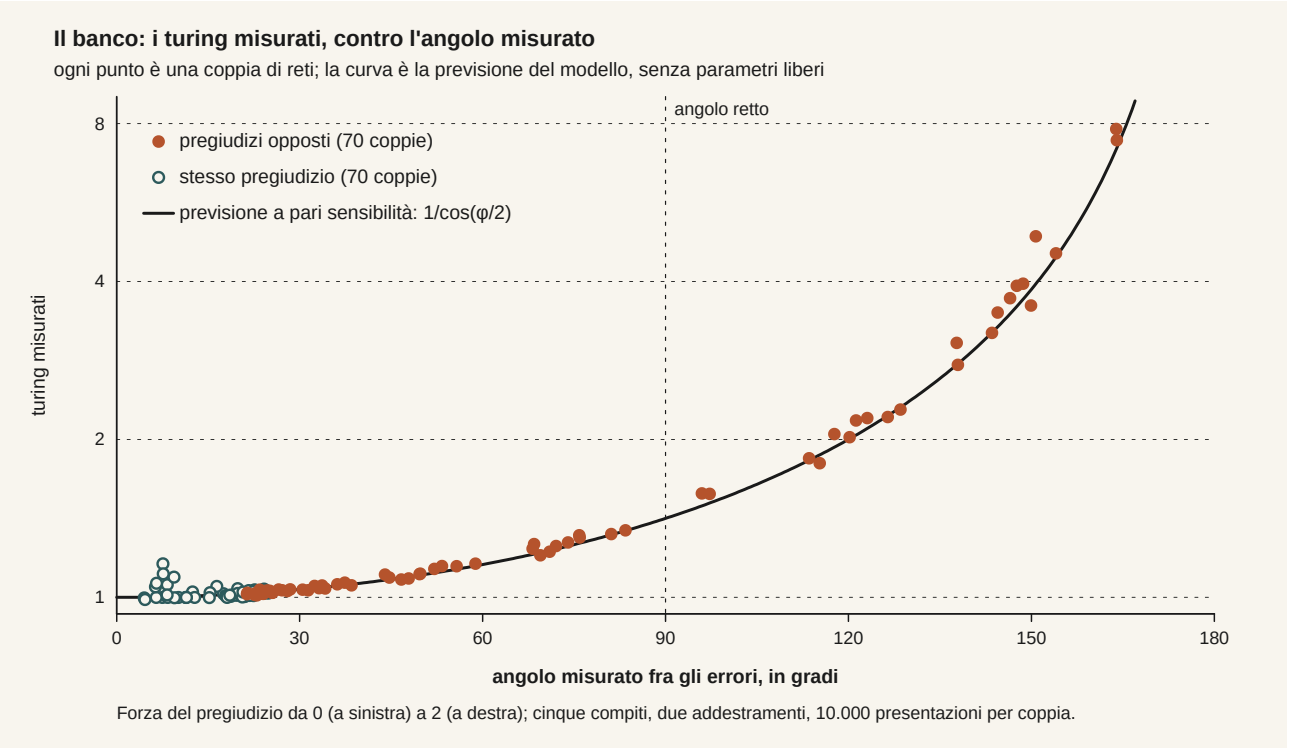


Figura 4. Il banco: turing misurati e angolo misurato, con la previsione per due menti di pari sensibilità.

I turing misurati stanno sulla curva prevista, da 23 gradi fino a 164 (figura 4). Nelle 22 coppie in cui l'angolo supera con certezza l'angolo retto, la formula sbaglia in media del **3 per cento**; il teorema di Pitagora, che a 90 gradi sarebbe esatto, sbaglia lì del 47. In tutte e 22 il turing sta sopra Pitagora con l'intero intervallo. Il caso più spinto è gatto contro cane con il pregiudizio al massimo. Ciascuna rete, da sola, indovina 53 volte su cento: quasi tira a sorte. Insieme indovinano 72 volte su cento. Sono **7,8 turing**, con un intervallo che va da 6,0 a 12,5, a 164 gradi; la formula ne prevedeva 7,3.

Le coppie di controllo fanno il contrario. Con lo stesso pregiudizio l'angolo si chiude, da 23 gradi a 7, e il turing resta a uno: due menti che sbagliano nello stesso verso non si correggono, sbagliano insieme. A fare la risonanza è l'opposizione dei due pregiudizi.

Anche la sintonia si comporta come previsto. Al pregiudizio di forza 1 la coppia conserva almeno metà del guadagno per pesi che vanno da 0,25 a 4,2; a forza 1,5 fra 0,42 e 2,5; a forza 2 fra 0,61 e 1,7. Le previsioni erano 0,24 e 4,1, 0,40 e 2,5, 0,59 e 1,8. Più sale il picco, più la fiducia va tarata con cura.

Resta un'ultima cosa, quella che ridimensiona il numero più vistoso. La sensibilità della coppia, in assoluto, non cresce mai con il pregiudizio: vale 3,44 quando le reti non ne hanno, 3,30 a forza 1, 2,92 a forza 1,5, **2,25** a forza 2. I turing salgono perché scende il denominatore: la rete da sola passa da 3,38 a 0,57. Fra le coppie con pregiudizi opposti, i turing più alti appartengono a quelle che in assoluto ne sanno meno. La risonanza ha restituito una parte di ciò che il pregiudizio copriva, e nemmeno tutto: una rete che in addestramento si è appoggiata alla cornice ha imparato meno dalla fotografia, e quel meno non torna.

Il nullo, misurato

Gli stessi file del banco bastano per cercare il nullo: si rifà il conto lasciando che il peso prenda anche valori negativi. Le coppie sono 140, le settanta opposte e le settanta uguali.

Il nullo c'è in tutte e 140, e sta dove la formula lo mette. Il prodotto fra il peso a cui il turing si annulla e il rapporto k dovrebbe valere -1: in mediana vale **−1,005**. Al peso previsto, calcolato su metà delle fotografie e letto sull'altra metà, restano in mediana 0,05 turing: la coppia, lì, tira a indovinare, pur essendo fatta di due reti che da sole distinguono le classi. Su tutta la curva, picco e nullo compresi, la distanza media fra i punti misurati e quelli previsti è lo 0,3 per cento dell'altezza del picco (figura 5).

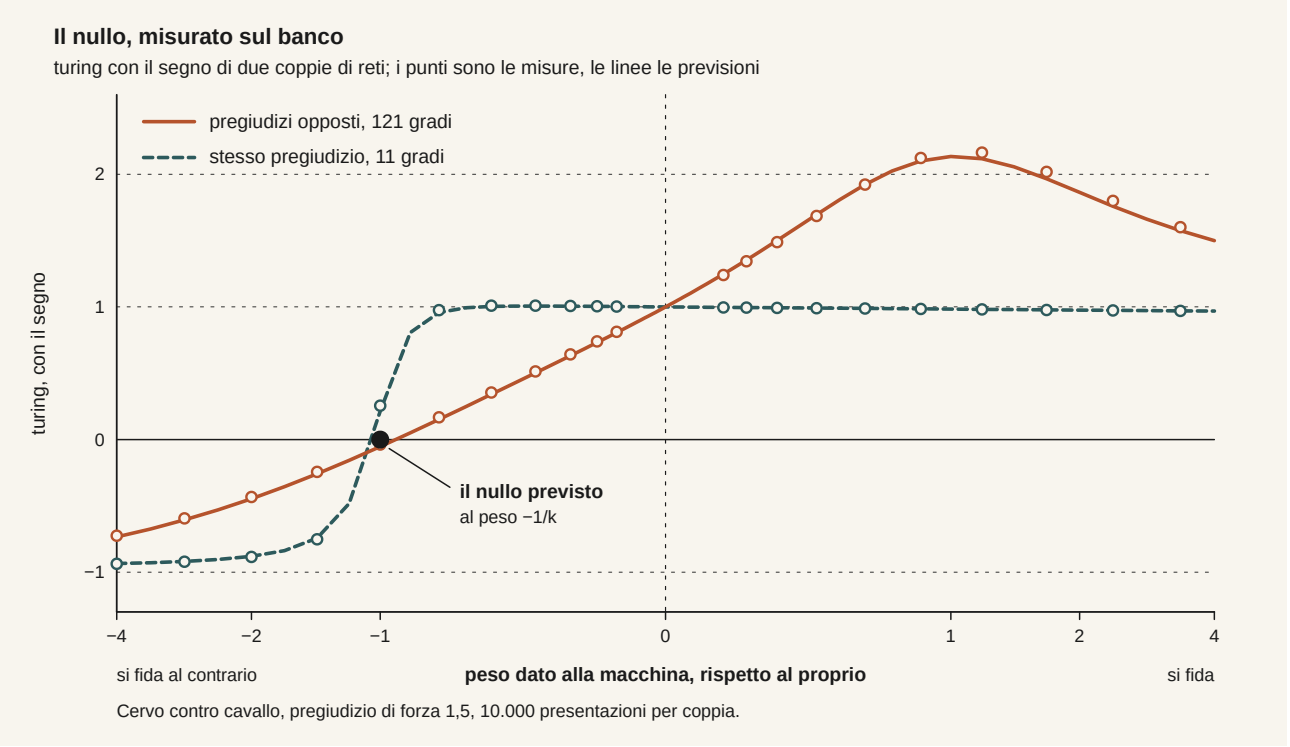


Figura 5. Il nullo misurato: i punti sono le misure, le linee le previsioni.

Anche la larghezza va come previsto, al contrario di quella del picco. Misurata come rapporto fra i due pesi ai bordi della zona in cui il turing resta sotto il mezzo, vale 1,6 a 23 gradi e **7,3** a 149 (le previsioni erano 1,6 e 7,9). Nelle coppie di controllo, dove l'angolo si chiude fino a 7 gradi, la zona si stringe a 1,2: una fessura.

La seconda condizione, il nullo a fiducia positiva, chiede una macchina che su una famiglia di casi stia sotto il caso. Sul banco quella famiglia si può isolare: sono le presentazioni in cui la cornice punta con forza contro la verità, nel verso del pregiudizio della macchina. Lì, in 22 coppie su 30, la macchina sbaglia in modo sistematico, e in tutte e 22 il turing si annulla a un peso positivo: fidarsi della macchina, in quella misura, porta la coppia a zero. Il criterio fissato prima passa, ma è il punto in cui previsione e misura sono più lontane di tutto il banco. Il peso misurato sta in mediana il 17 per cento sopra quello previsto, e in 19 di queste 22 selezioni la rete che fa da persona separa le classi senza un errore: lo strumento, lì, è al limite di ciò che può leggere.

Si può costruire?

Sul banco la controfase era fabbricata: i due pregiudizi li avevamo decisi noi, uguali e contrari. Fuori dal banco nessuno conosce il pregiudizio di una mente. Si conoscono i suoi giudizi. La domanda diventa allora se una macchina possa imparare a sbagliare al contrario di una mente guardando soltanto ciò che quella mente risponde.

La ricetta ingenua. La prima prova, scritta nel protocollo, era la più diretta. Una rete con il pregiudizio della cornice fa da persona. La macchina si addestra sul compito e in più viene punita, con una forza regolabile, ogni volta che il suo indizio va d'accordo con quello della persona a parità di risposta giusta. L'angolo si è aperto subito, fino a 158-166 gradi. E il risultato è stato inutile. Le macchine punite stanno **sotto il caso**: in mediana indovinano fra 21 e 40 volte su cento, dove la macchina non punita ne indovina 93 e la persona 85. Hanno trovato la via più corta per andare al contrario della persona: capovolgere tutto, il segnale insieme al rumore.

Una macchina così è la persona stessa, con il segno cambiato e più debole. Basta leggerla al rovescio per ritrovare una copia sbiadita della persona, fra 14 e 22 gradi, e una copia non aggiunge quasi niente: la coppia vale fra 1,09 e 1,17 turing. Nella formula si vede in una riga: cambiare segno insieme a k e a ρ lascia i turing apparenti come sono. La controfase che conta è quella degli errori quando i segnali vanno nello stesso verso, e l'angolo da solo non basta a riconoscerla: bisogna guardare anche da che parte sta k. Delle tre ipotesi scritte per questa prova due non passano, e la terza (il pregiudizio imparato dalla macchina ha il verso opposto a quello della persona) passa per la stessa scorciatoia.

Le macchine capovolte hanno però un merito involontario: sono la seconda condizione del nullo allo stato puro. Una macchina che sta sotto il caso ha k negativo, e la formula mette il nullo fra i pesi positivi, a $1/|k|$. In 38 coppie su 38 il turing si annulla proprio lì, a un peso positivo che dista in mediana l'1,5 per cento da quello previsto, e in quel punto restano 0,04 turing. Chi desse a una di queste macchine un po' più peso che a sé stesso non saprebbe più niente. Questo controllo non era nel protocollo: l'abbiamo aggiunto dopo aver visto i dati.

La ricetta che non ammette la scorciatoia. Visto il risultato abbiamo aggiunto una prova, dichiarata esplorativa nel protocollo, con le attese scritte prima di eseguirla. La macchina non viene più punita perché somiglia alla persona: impara dall'errore della *somma*. Guarda la stessa fotografia, e viene corretta solo su quanto sbaglia la somma del suo indizio e di quello della persona. Capovolgere tutto, così, non conviene più. Una seconda macchina è addestrata allo stesso modo, ma bendata: della fotografia vede soltanto la cornice, e del compito non può sapere niente (tabella 7).

Tabella 7. Tre compagni per la stessa persona
la persona è una rete con il pregiudizio della cornice, che da sola indovina l'85,2% delle volte; mediane su cinque compiti e due addestramenti. Le risposte giuste insieme sono quelle della somma dei due indizi così come escono dalle reti, la stessa su cui le due macchine in controfase sono state addestrate

Il compagno	Giuste, da solo	Angolo con la persona	Turing della coppia	Giuste, insieme
Macchina addestrata per conto suo	93,4%	52 gradi	1,46	92,1%
Macchina addestrata sulla somma	51,8%	132	1,37	91,9%
Macchina bendata, addestrata sulla somma	50,0%	131	1,34	91,9%

La macchina bendata, da sola, indovina 50 volte su cento: tira a sorte, com'è inevitabile. Accanto alla persona la porta dall'85 al **92 per cento**. Sono 1,34 turing, e la formula, per una macchina che non sa niente, ne prevedeva 1,33. Sul gatto e sul cane non le dice niente. Le toglie il pregiudizio: ha imparato a sbagliare al contrario di lei, a 131 gradi, questa volta senza capovolgere nessun segnale. È il principio delle cuffie che cancellano il rumore, dove un microfono che della musica non sente niente la fa sentire meglio, ed è la variabile soppressore degli statistici nella sua forma più pura. Il caso da manuale è dello stesso Horst, e viene dalla selezione dei piloti nella seconda guerra mondiale: aggiungere alla previsione una prova di abilità verbale, che con l'abilità da prevedere non aveva a che fare, la migliorava, e la prova entrava nel conto con il segno meno. Del volo non sapeva niente. Sapeva qualcosa dell'errore delle altre prove.

Fuori dal banco una macchina bendata sarebbe uno strumento strano: non conosce la risposta, conosce le nostre distorsioni. Perché funzioni servono due cose. La distorsione deve lasciare una traccia in qualcosa che la macchina può vedere senza guardare il caso, come la cornice. E servono molti giudizi passati di quella persona, con la risposta giusta accanto, su cui addestrarla. Una traccia del genere esiste. Nel 2016 Daniel Chen, Tobias Moskowitz e Kelly Shue hanno mostrato che giudici d'asilo, funzionari che valutano prestiti e arbitri di baseball tendono a decidere al contrario della decisione appena presa, a prescindere dal merito: dopo un sì diventa un po' più probabile un no. L'ordine dei casi non dice niente su chi ha ragione, e dice qualcosa sull'errore di chi giudica. Una macchina che vedesse soltanto l'ordine potrebbe, in linea di principio, restituire ciò che quell'abitudine toglie. Non l'abbiamo provato, e il banco dice anche quanto aspettarsi: **si recupera ciò che il pregiudizio copriva, non di più**.

La sorpresa è la macchina che la fotografia la vedeva. Poteva imparare tutte e due le cose, il compito e la correzione, e ha imparato quasi soltanto la seconda: da sola indovina 52 volte su cento, e in tre casi su dieci non si distingue da una che tira a sorte. Alla coppia dà 1,37 turing, poco più della bendata. Togliere il pregiudizio alla persona era la via più breve per abbassare l'errore della somma, e l'addestramento si è fermato lì.

> **Altrove succede.** Che la macchina più accurata non sia sempre il compagno migliore lo hanno mostrato Bansal e colleghi nel 2021: addestrando per la squadra un classificatore meno accurato, l'utilità attesa della squadra saliva da 0,524 a 0,606. Lì però la persona decide caso per caso se accettare il consiglio o fare da sé. Wilder, Horvitz e Kamar (2020) addestrano la macchina sui casi che per le persone sono difficili. Sul nostro banco la persona non sceglie caso per caso, e la macchina vede tutto ciò che vede la persona.

Il conto finale ridimensiona tutte e due. La macchina addestrata per conto suo, senza pensare alla persona, dà alla coppia **1,46 turing**: più delle due in controfase. Da sola vale già una volta e mezza la persona, e la coppia con lei vale quanto lei. Avevamo scritto, come attesa, che il miglior compagno non sarebbe stato il miglior solista. Non passa: su questo banco il miglior solista resta il miglior compagno, in tutti e cinque i compiti. La controfase si può costruire e rende quanto promette la formula, ma una macchina che sa di più rende ancora di più. Ciò che la risonanza offre di suo è un'altra cosa: un terzo di sensibilità in più da una macchina che del compito non sa niente.

Restava una strada: non partire da zero. La macchina addestrata sulla somma aveva preso la via più corta. Partendo dal miglior solista, e chiedendole di restare brava da sola mentre impara la correzione, poteva forse tenere tutte e due le cose. È la prova che abbiamo fatto per ultima, con le attese scritte prima di eseguirla, in due modi: ritoccando il solista con le due richieste insieme, e lasciandolo intatto con accanto un ramo piccolo che vede solo la cornice (tabella 8).

Tabella 8. Il solista che impara la persona
mediane su cinque compiti e due addestramenti. Il solista di controllo è ritoccato per lo stesso numero di epoche, ma senza pensare alla persona

La macchina	Da sola, fatta 1 la persona	Angolo fra gli errori	Turing della coppia
Il solista, di controllo	1,46	53 gradi	1,47
Il solista ritoccato anche sulla somma	1,39	66	1,47
Il solista con il ramo che vede la cornice	1,18	83	1,47

La macchina fa ciò che le si chiede. Ritoccata, tiene il 95 per cento della sua sensibilità e impara la correzione: la cornice la tira nel verso opposto a quello della persona in dieci casi su dieci. L'angolo si apre in tutti e cinque i compiti, da 53 a 66 gradi, e a 83 con il ramo. Ma **la coppia vale quanto prima, 1,47 turing**. Ciò che la macchina guadagna in angolo lo paga in ciò che sa da sola, e i due effetti si pareggiano: con gli angoli e i rapporti misurati la formula prevede 1,47, 1,46 e 1,46, e le misure le stanno entro l'1 per cento. Avevamo scritto che questo compagno avrebbe superato il solista, di poco. Non passa, in nessuno dei cinque compiti.

Col senno di poi il motivo sta nel banco, prima che nella ricetta. Qui la persona non sa niente che la macchina non possa vedere: guardano la stessa fotografia. Due reti senza pregiudizio, messe insieme, valgono 1,02 turing: la seconda aggiunge il 2 per cento a ciò che vede la prima. Tolto alla persona il suo pregiudizio, quel 2 per cento è più o meno ciò che le resta da dare a una macchina che vede quanto lei, e **sopra il miglior solista c'era poco da prendere**. Perché un compagno costruito lo superi, la persona deve avere qualcosa di suo. È il caso del radiologo con la cartella: dalla cartella prende un'informazione che nella lastra non c'è, e insieme una spinta. Lì togliere la spinta libera l'informazione. Quel banco non l'abbiamo costruito.

Con le persone vere. Qui i giudizi sono umani. CIFAR-10H raccoglie una cinquantina di giudizi di persone per ognuna delle 10.000 fotografie di prova di CIFAR-10, questa volta con tutte e dieci le classi. Il singolo giudizio umano è giusto il 95 per cento delle volte; la macchina, una ResNet-18, il 94. L'angolo fra i loro errori è 68 gradi, in linea con quelli dell'articolo precedente. Questa prova era nel protocollo dall'inizio, e usa quindi la ricetta della punizione. La macchina è stata ritoccata

con più cautela che sul banco: poche epoche, a partire dalla rete già addestrata, su metà delle fotografie, e misure sull'altra metà (tabella 9).

Tabella 9. Rifasare una macchina sulle persone vere

ResNet-18 e giudizi umani di CIFAR-10H; medie su due addestramenti e due metà, misure sulla metà non usata per il ritocco. I turing apparenti sono quelli della coppia ideale persona più macchina

Forza della punizione	Angolo fra gli errori	Giuste, la macchina da sola	Turing apparenti
0	68 gradi	94,1%	1,19
1	71	92,7%	1,18
3	75	88,8%	1,17
10	94	18,6%	1,02

L'angolo si apre davvero, e l'apertura è fuori dal rumore già con la punizione più leggera. Ma ogni grado costa alla macchina più sensibilità di quanta ne renda alla coppia: **i turing apparenti scendono**, da 1,19 a 1,17. Il motivo si vede guardando dove la macchina perde. Sulle fotografie in cui quasi tutte le persone rispondono giusto, con la punizione di forza 3 scende dal 97 al 92 per cento; su quelle in cui le persone si dividono non sale, e passa dal 74 al 70. La punizione le toglie ciò che sapeva dove le persone hanno ragione e non le dà niente dove sbagliano. Con la punizione più forte la rete crolla, e i 94 gradi che si leggono lì sono quelli di una macchina che non riconosce più niente. Avevamo previsto che con le persone vere l'angolo sarebbe rimasto sotto i 90 gradi a ogni forza: detta così la previsione non passa, per una rete rotta.

La ricetta della somma è nata dopo, dal fallimento della punizione, e con le persone l'abbiamo provata per ultima, anche questa con le attese scritte prima. Al posto dell'indizio di una persona c'è quello della folla: per ogni fotografia, la quota di giudizi andata a ciascuna classe. La stessa macchina è stata ritoccata in tre modi: con la sua sola perdita, come controllo; sull'errore della somma con la folla; con le due cose insieme (tabella 10).

Tabella 10. La ricetta della somma con le persone vere

ResNet-18 e giudizi umani di CIFAR-10H; medie su due addestramenti e due metà, misure sulla metà non usata per il ritocco. Il controllo ripete la prova a forza zero con una rete riaddestrata: per questo non coincide al centesimo

Il ritocco	Angolo fra gli errori	Giuste, la macchina da sola	Turing apparenti
La sua sola perdita (controllo)	67 gradi	94,0%	1,18
L'errore della somma	68	93,8%	1,18
Le due insieme	68	93,8%	1,18

Questa volta non succede niente, né in male né in bene. La macchina non si guasta, ma **l'angolo non si apre**: 67 gradi con il controllo, 68 con la somma, e la differenza sta dentro il rumore. I turing apparenti restano a 1,18. Delle cinque attese passano le due di prudenza, che la macchina non si sarebbe guastata e che l'angolo sarebbe rimasto sotto i 90 gradi; le tre che promettevano un guadagno, no. Una lettura possibile, che il protocollo non prevedeva: sul banco l'errore della persona stava in una cornice, e la macchina poteva vederlo. Qui la folla sbaglia di rado (la risposta più votata è giusta 99 volte su cento) e non per un motivo che si legga nella fotografia. E una folla non è una mente: nella quota dei voti i pregiudizi dei singoli si compensano già fra loro, prima che

la macchina arrivi. Dove le persone si dividono, la macchina resta al 72 per cento, con la somma o senza: sull'errore della somma c'era poco da imparare.

Che cosa abbiamo trovato

Tabella 11. Il bilancio
per ogni affermazione, da dove viene e come è andata. I criteri «fissati prima» stanno nel protocollo registrato prima degli esperimenti, quelli del nullo sono stati aggiunti prima dell'analisi dei dati; le prove sulla somma e quella sul solista sono state aggiunte dopo, e come tali vanno lette

L'affermazione	Da dove viene	Com'è andata
La curva della coppia è una curva di risonanza, con un nullo	Teorema del modello, controllato al calcolatore	Esatta, dentro il modello
Lo strumento ritrova i valori veri	Simulazione, due criteri fissati prima	Passano; lo strumento satura quando la coppia separa quasi tutto
Oltre i 90 gradi i turing seguono la formula	Banco con pregiudizi opposti, sei criteri fissati prima	Passano tutti; fino a 164 gradi e 7,8 turing
Il nullo esiste e sta al peso previsto	Stesso banco, cinque criteri fissati prima dell'analisi	Passano tutti; quello a fiducia positiva con lo strumento al limite
Punire la somiglianza costruisce la controfase	Banco, tre criteri fissati prima	Due non passano: la macchina capovolge il segnale
Addestrare sulla somma costruisce la controfase	Banco, prova aggiunta dopo; cinque attese scritte prima di eseguirla	Tre su cinque: 132 gradi e 1,37 turing, come previsto; ma il miglior solista resta il miglior compagno
Partendo dal miglior solista, il compagno costruito lo supera	Banco, prova aggiunta per ultima; cinque attese scritte prima di eseguirla	Tre su cinque: l'angolo si apre fino a 83 gradi, ma la coppia resta a 1,47 turing
Una macchina si può rifasare sulle persone vere	Giudizi umani di CIFAR-10H, due criteri fissati prima	L'angolo si apre da 68 a 75 gradi, i turing apparenti scendono
La ricetta della somma funziona con le persone vere	Giudizi umani di CIFAR-10H, prova aggiunta per ultima; cinque attese scritte prima di eseguirla	Due su cinque: l'angolo resta a 68 gradi, i turing apparenti a 1,18
Le coppie che esistono sono in risonanza	Rilettura dei dati dell'articolo precedente	No: fattore di merito fra 0,51 e 0,66

La formula ha mantenuto la promessa dove era più esposta a essere smentita: su reti vere, con angoli e turing misurati caso per caso, senza parametri da aggiustare. La promessa che l'idea portava con sé, una macchina costruita apposta per sbagliare al contrario di chi la usa e per questo più utile di una macchina più brava, sul banco non ha retto.

Che cosa non dice

Un turing alto non è una coppia brava. I 7,8 turing del banco sono di due reti che insieme indovinano 72 volte su cento. La risonanza si vede bene dove i singoli sono stati rovinati da un pregiudizio, e lì restituisce una parte del perduto. Accanto a ogni turing va letta la sensibilità di chi lo riceve.

Il banco è un banco. La «persona» è una rete, il pregiudizio ha una sola dimensione, e l'abbiamo scelto noi. I pregiudizi di una persona vera sono molti, non si conoscono in anticipo e cambiano con l'esperienza. Con le persone vere abbiamo due prove sugli stessi giudizi, in un compito in cui sbagliano una volta su venti: in una l'angolo si è mosso di pochi gradi, nell'altra per niente.

Qualcuno deve girare la manopola. Nel modello il peso giusto si sceglie; sul banco l'abbiamo scelto su metà delle fotografie. Una persona davanti a una macchina non ha quella metà, e la sua fiducia cambia da un caso all'altro. Più la coppia è vicina alla risonanza, più stretta è la banda in cui la fiducia rende, e dall'altra parte c'è il nullo: la stessa controfase che alza il premio alza il costo di una fiducia sbagliata. Il modello dice anche di quanto. Per due menti di pari sensibilità, chi dà alla macchina un peso troppo grande o troppo piccolo di un fattore f non supera mai $(f+1)/(f-1)$ turing, qualunque sia l'angolo. Con una fiducia doppia del dovuto, o dimezzata, **il tetto è 3 turing**, anche con errori perfettamente opposti; sbagliandola di un quarto è 9. I picchi più alti della curva sono fuori dalla portata di chi non sa tarare la fiducia quasi alla perfezione. Una fiducia decisa caso per caso, secondo quanto ciascuno dei due è sicuro, è un altro meccanismo, con un'altra curva: il modello, con il suo peso unico, non la descrive.

È una macchina su misura, e da sola non vale niente. La macchina bendata sbaglia al contrario di una mente precisa. Accanto a un'altra mente, con un altro pregiudizio, aggiunge soltanto rumore, e usata da sola tira a sorte. Una macchina in controfase è la metà di una coppia: fuori dalla coppia per cui è stata costruita è peggio di una macchina qualsiasi.

I numeri sono pochi. Cinque compiti, due addestramenti per ciascuno, una sola famiglia di immagini. Dove lo strumento era al limite lo abbiamo scritto. Le due prove aggiunte per ultime, il solista che impara la persona e la somma con i giudizi umani, non hanno cambiato il quadro. Manca ancora la prova che conta: persone di cui si misurano gli errori uno per uno, non una folla d'archivio, una macchina addestrata sulla somma con ciascuna, e un gruppo di controllo. La si può cercare fra le decisioni in serie di una stessa persona, dove l'errore segue una regola che una macchina bendata può imparare.

La manopola della vecchia radio non creava la voce: toglieva ciò che la copriva. Per le menti vale lo stesso, e i dati lo dicono con una precisione che non ci aspettavamo. Due difetti opposti, regolati l'uno sull'altro, lasciano passare ciò che ciascuno copriva. Ma una radio più sensibile sente più lontano di una radio ben sintonizzata, e le macchine di oggi, addestrate sui nostri testi e sui nostri giudizi, sbagliano in gran parte dove sbagliamo noi. Per ora la via seguita è una sola, farle più brave. Farle anche diverse da noi nel modo giusto è una via che esiste, e che si può misurare.

Il nome

L'idea di guardare a una persona e a una macchina come ai due rami di un circuito viene da vecchie conversazioni a voce con un chatbot. Alla domanda su come chiamare quella risonanza fra le due intelligenze, il chatbot propose l'aggettivo, «risonanza cognitiva», e, interrogato, rispose che in quel senso il termine non era diffuso. In un'altra conversazione il chatbot lo scambiò per la dissonanza cognitiva, che è tutt'altra cosa. Il nome è rimasto perché la curva gli ha dato ragione. Ma l'espressione ha già altri usi, e chi la incontra altrove non deve confonderla con questa.

Cognitive Resonance è il nome dell'organizzazione di Benjamin Riley e della sua newsletter, critica verso l'entusiasmo facile sull'intelligenza artificiale. In ecologia del comportamento Ferrari, Horn e

Chivers lo hanno coniato nel 2019 per dire quanto ciò che un animale impara da piccolo pesa su ciò che saprà imparare dopo (lo studio è sui girini). Negli studi sul *framing* Giorgi distingue dal 2017 una risonanza cognitiva e una emotiva dei messaggi. Nosta lo ha usato nel 2024, su *Psychology Today*, per l'intesa fra una persona e un modello linguistico. La risonanza adattiva di Stephen Grossberg, degli anni Settanta, e la risonanza stocastica di Benzi, Sutera e Vulpiani, del 1981, sono altri due meccanismi ancora. Qui il termine ha un solo significato, ed è una condizione che si misura: **l'angolo fra gli errori di due menti oltre i 90 gradi, con i segnali nello stesso verso**.

I conti

L'identità. Nel modello dell'articolo precedente (indizi gaussiani con la stessa varianza nelle due classi, messi insieme in modo lineare) la coppia che dà peso w all'indizio della macchina ha $\mathrm{Tu}(w)^2 = (1 + wk)^2/(1 + w^2 + 2\rho w)$. Per $k = 1$ si ha $\mathrm{Tu}^2 - 1 = 2(1-\rho)/(w + 1/w + 2\rho)$; con $w + 1/w = (\Omega - 1/\Omega)^2 + 2$ e $(1-\rho)/(1+\rho) = \tan^2(\varphi/2)$ si arriva alla forma risonante, con $Q^2 = 1/(2(1+\rho))$. Per k qualunque, la sostituzione $\varepsilon = (w + \cos\varphi)/\sin\varphi$ porta al profilo di Fano. Le due forme sono state controllate al calcolatore contro la formula di partenza, con scarti al livello dell'arrotondamento del calcolatore. A peso fissato e $k = 1$ la formula di partenza cresce al diminuire di ρ, e per $\rho \to -1$ tende a $(1+w)^2/(1-w)^2$: poiché il picco sta a $w = 1$, un peso f volte quello giusto dà al più $(f+1)/(f-1)$ turing. Per la stella, la matrice di correlazione di un centro legato con ρ a m giudici fra loro indipendenti ha autovalori $1 \pm \rho\sqrt{m}$, ed è ammissibile solo se $|\rho| \le 1/\sqrt{m}$.

Gli stimatori. La sensibilità è $d_a = \sqrt{2}\, z(\mathrm{AUC})$, con il segno; ρ è la correlazione, a parità di classe, fra gli indizi standardizzati dei due giudici, e la versione tetracorica dell'articolo precedente fa da ponte (sul banco le due distano in mediana da meno di 1 a circa 5 gradi, di più quando il pregiudizio è forte). Il peso varia su 49 valori da 1/64 a 64, con il segno meno per il nullo. Il turing sintonizzato sceglie il peso su metà delle immagini e si legge sull'altra metà. Gli intervalli vengono da 1.000 ricampionamenti per immagine, con il peso scelto di nuovo a ogni ricampionamento.

Lo strumento sul modello. Su dati simulati dal modello il turing sintonizzato sbaglia al più dello 0,8 per cento fino a $\rho = -0{,}9$, e l'intervallo al 95 per cento contiene il valore vero 92-93 volte su cento. Lo strumento satura quando la coppia separa quasi tutti i casi: con persone già brave da sole e una coppia che dovrebbe superare una sensibilità di 6-7, il turing esce per difetto, dal 18 a oltre il 50 per cento secondo i casi.

Fuori dal modello. La curva usa soltanto le medie degli indizi e le loro correlazioni. Come rapporto fra segnale e rumore della somma pesata resta quindi vera anche per indizi che gaussiani non sono, purché abbiano la stessa dispersione nelle due classi. Fuori dalla gaussiana si perdono due cose: la somma pesata non è più per forza il modo migliore di mettere insieme gli indizi, e una sensibilità ricavata dall'area sotto la curva ROC non coincide più esattamente con quel rapporto. Il banco è la prova pratica: le uscite di reti vere non hanno motivo di essere gaussiane, e sulle 70 coppie con pregiudizi opposti lo scarto fra i turing misurati e quelli previsti è in mediana lo 0,7 per cento, al più il 7.

Il banco. Una rete convoluzionale piccola (cinque strati, da 48 a 192 canali), 30 epoche, 9.000 fotografie di addestramento per compito e 2.000 di prova, ognuna presentata cinque volte con cornici diverse. La cornice è il bordo di due pixel, impostato a un valore proporzionale al numero estratto. Gli addestramenti indipendenti sono due e non tre, come voleva il protocollo, per restare nel tempo di calcolo fissato: è la sola deviazione nell'esecuzione degli esperimenti, e sta scritta con le altre scelte nella sezione «Deviazioni» del protocollo.

La prova sulla somma. All'uscita della macchina si somma, nella perdita, il logaritmo delle probabilità a favore calcolato dalla persona, congelata, sulla stessa presentazione: la macchina si addestra sull'entropia incrociata della somma. La macchina bendata vede la stessa immagine con l'interno azzerato. Delle cinque attese ne passano tre: l'angolo si apre in tutti e cinque i compiti, e i turing stanno in mediana entro il 2 per cento dalla previsione, sia con la macchina che vede sia con quella bendata. Non passano quella sul miglior compagno e quella che voleva la macchina che vede sopra il caso in ogni unità (lo è in 7 su 10).

Le due prove aggiunte per ultime. Nella prima la persona e il solista sono quelli della prova sulla somma. Il controllo è il solista ritoccato per altre 10 epoche, a un quinto del passo di apprendimento, con la sua sola perdita; il solista ritoccato sulla somma aggiunge a quella perdita l'entropia incrociata della somma con la persona, congelata. Il ramo è una rete nuova, della stessa forma delle altre, che vede solo la cornice, addestrata da zero sulla somma di persona, solista e ramo, con i primi due congelati. Delle cinque attese ne passano tre: la macchina non dimentica, l'angolo si apre, i turing stanno entro il 10 per cento dalla previsione (di fatto entro l'1). Non passano le due che volevano la coppia sopra quella con il solista di controllo, e sopra lo stesso solista da solo. Nella seconda l'indizio della folla è il logaritmo della quota di voti di ogni classe, lisciata; la ResNet-18, le due metà e il ritocco di 8 epoche sono quelli della prova con la punizione. Passano due attese su cinque.

Dati, codice, tempi. Le immagini sono quelle di CIFAR-10; i giudizi umani sono CIFAR-10H, con licenza CC BY-NC-SA 4.0, che non ridistribuiamo. Gli addestramenti sono stati eseguiti su una sola scheda NVIDIA L4, per meno di due ore in tutto. Il protocollo con le sue versioni, il codice, i risultati e la pagina che dice che cosa passa e che cosa no stanno nel repository cognitive-resonance; gli angoli fra persone e macchine si calcolano con turing-meter.

Per approfondire

- endjneer, *Intelligenza Apparente*, paithon, 2026: il turing, l'angolo fra gli errori, i turing apparenti.
- İpek Oruç, Laurence T. Maloney e Michael S. Landy, *Weighted linear cue combination with possibly correlated error*, Vision Research, 2003.
- Robert D. Sorkin e Huanping Dai, *Signal Detection Analysis of the Ideal Group*, Organizational Behavior and Human Decision Processes, 1994.
- Yu Hu, Joel Zylberberg e Eric Shea-Brown, *The Sign Rule and Beyond: Boundary Effects, Flexibility, and Noise Correlations in Neural Population Codes*, PLoS Computational Biology, 2014.
- Richard P. Larrick e Jack B. Soll, *Intuitions About Combining Opinions: Misappreciation of the Averaging Principle*, Management Science, 2006.
- Y. Liu e X. Yao, *Ensemble learning via negative correlation*, Neural Networks, 1999.

- Paul Horst, *The Prediction of Personal Adjustment*, Social Science Research Council, 1941; Anthony J. Conger, *A Revised Definition for Suppressor Variables: A Guide to Their Identification and Interpretation*, Educational and Psychological Measurement, 1974.
- Lynn Friedman e Melanie Wall, *Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression*, The American Statistician, 2005.
- Ugo Fano, *Sullo spettro di assorbimento dei gas nobili presso il limite dello spettro d'arco*, Il Nuovo Cimento, 1935, e *Effects of Configuration Interaction on Intensities and Phase Shifts*, Physical Review, 1961.
- Andrey E. Miroshnichenko, Sergej Flach e Yuri S. Kivshar, *Fano resonances in nanoscale structures*, Reviews of Modern Physics, 2010.
- Bryan Wilder, Eric Horvitz e Ece Kamar, *Learning to Complement Humans*, IJCAI 2020.
- Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz e Daniel S. Weld, *Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork*, AAAI 2021.
- Kate Donahue, Alexandra Chouldechova e Krishnaram Kenthapadi, *Human-Algorithm Collaboration: Achieving Complementarity and Avoiding Unfairness*, FAccT 2022.
- Joshua C. Peterson, Ruairidh M. Battleday, Thomas L. Griffiths e Olga Russakovsky, *Human Uncertainty Makes Classification More Robust* (CIFAR-10H), ICCV 2019.
- Robert Geirhos e colleghi, *Shortcut learning in deep neural networks*, Nature Machine Intelligence, 2020.
- Bernard Widrow e colleghi, *Adaptive noise cancelling: Principles and applications*, Proceedings of the IEEE, 1975.
- Daniel L. Chen, Tobias J. Moskowitz e Kelly Shue, *Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires*, The Quarterly Journal of Economics, 2016.
- Gli altri usi del nome: Maud C. O. Ferrari, Marianna E. Horn e Douglas P. Chivers, *Cognitive resonance: When information carry-over constrains cognitive plasticity*, Functional Ecology, 2019; S. Giorgi, *The Mind and Heart of Resonance*, Journal of Management Studies, 2017; J. Nosta, *"Cognitive Resonance" and the Power of Large Language Models*, Psychology Today, 2024; Cognitive Resonance, l'organizzazione di Benjamin Riley.

Come citare. endjneer (2026). *Risonanza cognitiva.* paithon.it, versione 1, 11 novembre 2026. https://doi.org/10.5281/zenodo.24140357

Codice e dati: https://github.com/paithon-it/cognitive-resonance.