

Bruno Cancellieri, ChatGPT e Claude.ai

Il giudizio presunto

Una prospettiva unificante sulla previsione del giudizio sociale

24 marzo 2026

NOTA: Questo articolo nasce da un dialogo con ChatGPT e Claude.ai, secondo un metodo che ho adottato anche per altri lavori: propongo un tema, discuto con questi strumenti l'impostazione e i contenuti, e faccio scrivere e poi correggere i testi attraverso un confronto articolato, in cui le modifiche vengono proposte, argomentate e discusse insieme.

Il quadro teorico specifico si sviluppa in questi dialoghi, ma si inserisce in una formazione psicologica e filosofica che ho maturato in autonomia ben prima di ricorrere a questi strumenti.

La responsabilità delle tesi sostenute, e delle loro eventuali debolezze, resta comunque mia.

Bruno Cancellieri

Introduzione

Siamo portati a credere che il nostro comportamento sia influenzato dal giudizio che gli altri esprimono realmente nei nostri confronti. In realtà, nella maggior parte dei casi, quel giudizio non viene mai pronunciato. Eppure esercita ugualmente una profonda influenza sulle nostre decisioni. Prima ancora di parlare, di agire o di scegliere, la mente sembra domandarsi, per lo più in modo inconsapevole: *che cosa penseranno gli altri di me se farò questo?*

Non occorre che qualcuno ci osservi, né che qualcuno ci critichi. È sufficiente che la nostra mente ne preveda la possibilità.

Conviene definire subito l'oggetto di questo saggio. Per *giudizio presunto* intendo la rappresentazione anticipata, generalmente implicita e inconscia, del giudizio che altri potrebbero formulare su un nostro comportamento, attuale o futuro. "Giudizio" va inteso qui in senso ampio: non l'opinione verbalizzata, ma l'intera risposta sociale prevista — approvazione o disapprovazione, ammirazione o delusione, accoglienza o rifiuto, ostilità o benevolenza, disponibilità alla cooperazione o perdita di fiducia, ecc. Tutto ciò che segue è un'esplorazione di questa definizione: dove il meccanismo si manifesta, come funziona e dove smette di spiegare.

L'idea non è nuova. Charles Cooley la formulò già nel 1902 con la metafora del *Looking-glass self*: il sé come riflesso di ciò che immaginiamo gli altri vedano in noi. George Herbert Mead la sviluppò con il concetto di *Altro generalizzato*, la voce interiorizzata della comunità. Erving Goffman descrisse la vita sociale come continua gestione dell'impressione. E dagli anni Novanta la *Sociometer theory* di Mark Leary e Roy Baumeister ha dato a questa intuizione una base empirica: l'autostima funzionerebbe come un indicatore interno che monitora il nostro grado di accettazione sociale percepita. Ciò che questo saggio propone non è dunque una scoperta, ma una prospettiva. La sua tesi può essere riassunta in una frase: numerosi fenomeni psicologici tradizionalmente studiati separatamente — autostima, vergogna, imbarazzo, conformismo, autocensura, parte del senso morale — sarebbero espressioni di un unico sistema di simulazione sociale che anticipa le conseguenze relazionali dei comportamenti e contribuisce alla loro selezione. Il resto del testo è lo sviluppo di questa frase.

Il supervisore del giudizio sociale

Si può ipotizzare che la mente disponga di un sistema di monitoraggio incaricato di stimare le conseguenze sociali delle nostre azioni. Ogni comportamento produce infatti almeno due tipi di effetti. Il primo riguarda il mondo esterno: ottenere un risultato, evitare un pericolo, soddisfare un bisogno. Il secondo riguarda la nostra immagine sociale, in tutte le sue dimensioni: simpatia, competenza, affidabilità, prestigio, status, desiderabilità, credibilità. La reputazione è uno degli esiti di questo monitoraggio, non il suo unico oggetto: come cambierà, in generale, l'opinione che gli altri hanno di noi?

Questa seconda valutazione avviene per lo più automaticamente, sotto la soglia della coscienza. La mente non aspetta che il giudizio venga espresso: lo anticipa. Questa interpretazione è coerente con il paradigma del *predictive processing*, secondo cui il cervello costruisce continuamente previsioni sul mondo e le corregge sulla base dell'errore. Qui si ipotizza che tra gli oggetti di tali previsioni rientrino anche le conseguenze reputazionali delle nostre azioni: un'estensione teorica specifica, non un'implicazione diretta del paradigma generale. Il ponte tra i due livelli è la teoria della mente: per prevedere un giudizio occorre disporre di un modello, per quanto rozzo e implicito, della mente di chi giudica — di ciò che nota, di ciò che valuta, dei criteri che applica. Il supervisore del giudizio sociale può quindi essere descritto come un caso particolare di elaborazione predittiva applicata ai modelli delle menti altrui: il punto in cui predictive processing e cognizione sociale si incontrano.

C'è però un aspetto che questa descrizione non cattura ancora. Se, come si vedrà, il sistema calcola un costo relazionale atteso per ogni comportamento possibile, allora non è soltanto un meccanismo di previsione: è un sistema di *selezione delle azioni*. Il flusso può essere schematizzato così: comportamenti possibili → previsione delle conseguenze fisiche → previsione delle conseguenze sociali → integrazione delle due stime → scelta. Il supervisore non serve a sapere che cosa penseranno gli altri; serve a decidere quale comportamento conviene adottare. “Supervisore” va dunque inteso come abbreviazione: funzionalmente, ciò che viene descritto è un sistema di valutazione sociale che predice, attribuisce valore, inibisce e facilita.

Il giudizio immaginato

Non reagiamo soltanto ai giudizi reali: reagiamo soprattutto ai giudizi che prevediamo. Ogni possibile comportamento è accompagnato da una simulazione implicita delle sue conseguenze relazionali. *Se dico questa cosa, penseranno che sono ingenuo. Se faccio quella domanda, sembrerò incompetente.* Queste simulazioni possono durare una frazione di secondo e non raggiungere mai la consapevolezza, ma influenzano il comportamento.

Un dato sperimentale rilevante: tendiamo sistematicamente a sovrastimare quanto gli altri ci osservino e ci giudichino. È lo *spotlight effect* documentato da Gilovich e Savitsky: il simulatore non è neutro, è tarato verso l'eccesso di prudenza. Da un punto di vista evolutivo un simile bias è comprensibile — il costo di sottovalutare un rischio reputazionale era probabilmente maggiore del costo di sopravvalutarlo — ma occorre cautela: che un tratto sia oggi utile a raccontarsi non dimostra che sia stato selezionato per quello scopo.

Il costo relazionale atteso

Descrivere il giudizio presunto come “una previsione” è ancora troppo generico. Ogni previsione di questo tipo ha una struttura: *chi* mi giudicherà, *su quale aspetto*, e *con quali conseguenze* per la mia posizione sociale. E queste componenti non pesano allo stesso modo. Essere giudicati negativamente da un estraneo, dal proprio partner, dal proprio capo o da un figlio non produce la stessa simulazione, né la stessa inibizione.

C'è però un passo ulteriore da fare. A ben vedere, il giudizio non è la variabile finale: è un passaggio intermedio. Ciò che il sistema sta realmente cercando di prevedere sono le conseguenze di quel giudizio — perdita di status, di opportunità, di cooperazione, di attrattività, fino al rischio di esclusione; oppure, sul versante opposto, guadagno di prestigio e influenza. Il supervisore sembra quindi valutare simultaneamente la probabilità che il giudizio venga formulato, la sua valenza, il peso sociale di chi giudica e l'impatto previsto sulle relazioni che contano. Il risultato è qualcosa di analogo a un valore atteso: non tanto un costo reputazionale, quanto un *costo relazionale atteso* che la mente calcola, per lo più senza accorgersene, prima di ogni comportamento socialmente visibile. Questo spiega perché giudizi identici abbiano pesi diversi: la stessa valutazione di incompetenza costa quasi nulla se viene da qualcuno che non rivedremo mai, moltissimo se viene dal nostro datore di lavoro.

Quasi nulla — ma non nulla, e l'anomalia è istruttiva. Arrossiamo anche davanti a estranei che non incontreremo mai più, e ci curiamo dell'opinione di perfetti sconosciuti. Se il sistema calcolasse davvero solo le conseguenze, questi giudizi dovrebbero lasciarci indifferenti. Una possibile spiegazione è che il sistema utilizzi il giudizio come *proxy* delle sue conseguenze relazionali: la valuta ultima del calcolo sarebbero le conseguenze, ma la valuta prossima, quella effettivamente registrata, il giudizio stesso. Se così fosse, il disagio provocato dal giudizio di estranei irrilevanti potrebbe riflettere il fatto che il meccanismo si è sviluppato in contesti sociali nei quali osservatori e futuri partner di interazione coincidevano molto più spesso di quanto accada oggi — e non sarebbe un controesempio, ma la firma di un sistema tarato su quel mondo. Va detto con chiarezza: è un'ipotesi evolutiva plausibile, non un fatto verificabile direttamente.

Va detto che l'idea di un calcolo di questo tipo non è inedita: l'economia comportamentale ha formalizzato l'utilità reputazionale nelle scelte prosociali (Bénabou e Tirole, 2006). Ciò che qui si propone è di estenderla oltre le decisioni economiche, a meccanismo psicologico generale, continuo e prevalentemente inconsapevole, e di spostarne l'oggetto dal giudizio alle sue conseguenze. Il vantaggio non è solo descrittivo: una formulazione di questo tipo rende il modello più falsificabile. Se il costo atteso dipende dalle conseguenze relazionali previste e non dal giudizio in sé, allora manipolare sperimentalmente l'identità dell'osservatore e la probabilità di interazioni future — estraneo di passaggio, pari, figura di riferimento — dovrebbe modulare in modo prevedibile autocensura, conformismo e arousal da valutazione, con un residuo irriducibile anche nella condizione anonima, riflesso del proxy. Se non lo fa, il modello è sbagliato.

L'autocoscienza pubblica

Il caso in cui il simulatore lavora più visibilmente è quello dell'attenzione rivolta a come appariamo: arrossire dopo essere inciampati, sentirsi osservati entrando in una sala, modificare il proprio comportamento appena qualcuno entra nella stanza. Nessuno ha espresso un giudizio; è bastato che diventasse possibile.

Anche questi fenomeni hanno una tradizione di ricerca precisa. Duval e Wicklund, con la *Objective Self-Awareness Theory* (1972), mostrarono che quando l'attenzione si sposta su di sé come oggetto — davanti a uno specchio, a una telecamera, a un pubblico — il confronto tra come si è e come si dovrebbe apparire diventa saliente e il comportamento cambia. Fenigstein e colleghi distinsero poi la *public self-consciousness* come tratto stabile: la disposizione individuale a rappresentarsi come oggetto dello sguardo altrui. Nel linguaggio di questo saggio: condizioni situazionali e differenze individuali che aumentano l'attività del simulatore del giudizio. È forse il caso più diretto in cui il meccanismo qui ipotizzato è stato, di fatto, già osservato in laboratorio.

L'osservatore virtuale

C'è una distinzione sottile che merita di essere resa esplicita. Il sistema non prevede soltanto *che cosa penseranno gli altri di me*: genera emozioni a partire dalla simulazione stessa. La vergogna, l'imbarazzo, l'orgoglio non richiedono che il giudizio arrivi davvero; nascono dal confronto con un osservatore virtuale che il supervisore costruisce e mantiene attivo. In un certo senso, la domanda operativa non è "che cosa penseranno di me?" ma "che cosa proverò io, immaginando il loro giudizio?". Il giudizio reale può anche non arrivare mai: l'osservatore interno ha già emesso il suo verdetto.

Anche questa idea ha antecedenti illustri, ma vanno presi per ciò che sono: antecedenti parziali, non equivalenti. Lo *Spettatore imparziale* di Adam Smith ("Teoria dei sentimenti morali", 1759) è soprattutto un dispositivo normativo — un osservatore ideale con cui valutiamo moralmente noi stessi. L'*Altro generalizzato* di Mead è un costrutto sociologico — l'interiorizzazione delle aspettative della comunità. L'osservatore virtuale qui descritto è invece un meccanismo cognitivo predittivo, e per questo più generale: può simulare il partner, il capo, un estraneo, un gruppo specifico o la comunità intera, e cambiare bersaglio da un momento all'altro. Ciò che i tre costrutti condividono è l'esito: con il tempo il giudizio presunto si stacca dagli osservatori reali e diventa una struttura permanente della mente, una voce che sembra appartenere esclusivamente alla nostra coscienza — e che invece è la comunità, interiorizzata.

L'autostima come previsione

Secondo la *Sociometer theory*, l'autostima non misura ciò che valiamo in assoluto, ma il valore che stimiamo di avere agli occhi degli altri. Questo spiega perché una persona possa sentirsi inadeguata senza che nessuno l'abbia criticata.

Ma qui emerge un problema che va affrontato, non aggirato: se l'autostima è un predittore, perché in certe persone non si aggiorna? Perché qualcuno continua a sentirsi incapace pur ricevendo riconoscimenti continui? Un buon sistema predittivo dovrebbe correggersi con l'evidenza. La risposta più plausibile è che il modello interno non si costruisce sui dati recenti ma su quelli formativi: esperienze precoci di rifiuto o svalutazione possono fissare parametri che i feedback successivi faticano a modificare — un predittore mal calibrato, non assente. È esattamente ciò che la terapia cognitiva chiama schemi disfunzionali, e la loro resistenza all'evidenza è un fatto che qualunque teoria dell'autostima deve spiegare, non una conferma automatica di questa.

Lo status e l'autocensura

Ma perché il cervello dovrebbe dedicare tante risorse a prevedere le reazioni altrui? Per una specie profondamente cooperativa come *Homo sapiens*, anticipare la risposta degli altri non significava soltanto evitare emozioni spiacevoli: significava stimare la probabilità di mantenere relazioni di cooperazione, ricevere aiuto, accedere alle risorse del gruppo ed evitare l'esclusione. Essere stimati significava alleanze, risorse, partner; essere disprezzati significava spesso isolamento. Non sorprende che la mente attribuisca grande peso alle variazioni previste del proprio status.

Uno degli effetti più evidenti è l'autocensura. Molte idee non vengono espresse, molte domande non vengono formulate — non perché sbagliate, ma perché il sistema prevede che possano compromettere l'immagine di chi le pronuncia. Il silenzio, l'inazione possono essere decisioni prese prima che la coscienza inizi a riflettere.

Il senso morale: un caso più complicato

Verrebbe naturale estendere il meccanismo anche alla morale: evitiamo certi comportamenti perché anticipiamo la condanna della comunità. In parte è vero, ma l'evidenza impone distinzioni.

Primo: la ricerca di June Tangney ha mostrato che vergogna e colpa non sono la stessa cosa. La vergogna è centrata sull'immagine di sé davanti agli altri, e si accorda bene con il modello reputazionale; la colpa è centrata sul danno fatto alla vittima, si nutre di empatia e può essere intensa anche in perfetta segretezza. Secondo: negli esperimenti economici in condizioni di anonimato garantito — i *dictator game* — la generosità diminuisce, il che conferma il ruolo della reputazione, ma non scompare, il che dimostra che la reputazione non è tutto.

La conclusione onesta è che l'anticipazione del giudizio è *una* componente del senso morale, accanto a empatia, norme interiorizzate e ragionamento etico — non il suo fondamento nascosto. La reputazione non è tutto: e la teoria è più forte, non più debole, proprio perché rinuncia a spiegare tutto.

Ciò che il sistema rende possibile

Prima di parlare dei suoi errori, va detto che cosa il sistema produce quando funziona — perché normalmente funziona. La cooperazione tra non consanguinei, la fiducia negli sconosciuti, la cortesia, il rispetto spontaneo delle norme, l'apprendimento sociale, la stessa esistenza della reputazione come istituzione: tutto questo presuppone individui capaci di anticipare l'effetto delle proprie azioni sugli altri e di regolarsi di conseguenza. Senza questo sistema, la convivenza umana sarebbe probabilmente impossibile. L'autocensura che tratteremo come problema è la stessa capacità che ci impedisce di dire a un lutto ciò che pensiamo del defunto. I fenomeni clinici che seguono non sono il funzionamento del sistema: sono le sue tarature sbagliate.

Quando il sistema sbaglia

Come ogni sistema predittivo, anche questo produce errori: può sovrastimare la probabilità del rifiuto, generalizzare esperienze negative del passato, continuare a funzionare quando nessuno ci osserva. Da questi errori possono derivare ansia sociale, perfezionismo, bisogno compulsivo di approvazione. Non a caso la terapia cognitivo-comportamentale dell'ansia sociale lavora proprio sulla correzione delle previsioni: il problema non è il giudizio degli altri, ma il modello che la mente ne costruisce.

I limiti dell'ipotesi

Una prospettiva unificante è utile solo se riconosce ciò che non spiega.

Il meccanismo non è uniforme tra culture: il peso del giudizio anticipato varia sensibilmente tra società più collettiviste e più individualiste, e tra le cosiddette culture della vergogna e culture della colpa. Non è uniforme tra individui: le condizioni in cui la simulazione del giudizio altrui è atipica — come nell'autismo — o in cui il suo peso motivazionale è ridotto — come nella psicopatia — costituiscono casi di test naturali che una teoria seria dovrebbe affrontare. E non è chiaro quando e come il sistema si sviluppi nel bambino.

Soprattutto, un'ipotesi che spiega tutto rischia di non spiegare niente. Perché sia scienza e non solo narrazione, deve produrre predizioni verificabili. Alcune sono possibili: se l'autostima è un sociometro, manipolazioni sperimentali dell'accettazione percepita dovrebbero modificarla più di manipolazioni della competenza oggettiva (ed è ciò che Leary ha trovato); se l'autocensura dipende dal rischio reputazionale previsto, dovrebbe ridursi in condizioni di anonimato autentico; se la

vergogna è reputazionale e la colpa empatica, le due emozioni dovrebbero dissociarsi in condizioni di segretezza.

La predizione più caratteristica del modello, però, è un'altra: se il supervisore seleziona le azioni sulla base del costo relazionale atteso, allora manipolare la *credenza* di essere osservati dovrebbe modificare il comportamento anche quando nessun osservatore reale è presente. È questa a distinguere il modello da una spiegazione puramente comportamentista, perché attribuisce un ruolo causale alla simulazione interna e non alla presenza effettiva di un pubblico. Un paradigma sperimentale esiste già: il *Watching-eyes effect* (Bateson, Nettle e Roberts, 2006), in cui semplici immagini di occhi aumentano i comportamenti cooperativi e onesti. Le repliche sono però contrastanti — e questo, per il saggio, è un punto di forza più che di debolezza: la sua predizione centrale coincide con una questione empirica genuinamente aperta, il cui esito può corroborare o ridimensionare il modello. Dove le predizioni falliscono, l'ipotesi va ridimensionata.

Conclusione

Il giudizio altrui non è semplicemente un evento esterno che talvolta ci influenza: è uno scenario che la mente costruisce incessantemente, quasi sempre senza che ce ne accorgiamo. In questa prospettiva non viviamo soltanto tra gli altri, ma anche tra le rappresentazioni che la nostra mente costruisce del loro possibile giudizio.

Il contributo di questa prospettiva non consiste nell'introdurre un nuovo meccanismo psicologico, ma nel proporre che molte funzioni apparentemente distinte — dall'autostima all'autocensura, dalla vergogna a una parte della moralità — possano essere interpretate come espressioni di un unico sistema predittivo dedicato alla simulazione delle conseguenze sociali dei nostri comportamenti. È un filo che attraversa Cooley, Mead, Goffman e la psicologia sperimentale contemporanea: qui si è cercato di annodarlo. In psicologia, è talvolta proprio questo tipo di riorganizzazione concettuale — non un nuovo dato, ma un nuovo modo di collegare i dati — ad aprire linee di ricerca inedite.

Riferimenti essenziali: A. Smith, *"The Theory of Moral Sentiments"* (1759); C.H. Cooley, *"Human Nature and the Social Order"* (1902); G.H. Mead, *"Mind, Self, and Society"* (1934); E. Goffman, *"The Presentation of Self in Everyday Life"* (1956); S. Duval & R. Wicklund, *"A Theory of Objective Self-Awareness"* (1972); A. Fenigstein, M. Scheier & A. Buss, *"Public and private self-consciousness: Assessment and theory"* (1975); M. Leary & R. Baumeister, *"The nature and function of self-esteem: Sociometer theory"* (2000); T. Gilovich & K. Savitsky, *"The spotlight effect and the illusion of transparency"* (2000); J. Tangney & R. Dearing, *"Shame and Guilt"* (2002); U. Frith & C. Frith, *"Development and neurophysiology of mentalizing"* (2003); R. Bénabou & J. Tirole, *"Incentives and prosocial behavior"* (2006); M. Bateson, D. Nettle & G. Roberts, *"Cues of being watched enhance cooperation in a real-world setting"* (2006); A. Clark, *"Whatever next? Predictive brains, situated agents, and the future of cognitive science"* (2013).