

Sintesi e conclusioni

Prendere sul serio il rumore

Il rumore è la variabilità indesiderata dei giudizi, e la sua presenza è dilagante. In questo libro ci siamo prefissi di spiegarne il motivo e di trovare delle possibili soluzioni. Abbiamo messo in campo molte idee e, a mo' di conclusione, riportiamo qui una breve sintesi dei punti principali per poi inserirli in una prospettiva più ampia.

Giudizi

Nell'accezione che qui abbiamo dato al termine, il *giudizio* non va confuso con l'attività del "pensiero". È un concetto molto più ristretto: il giudizio è una forma di misurazione il cui strumento è la mente umana. Come altre misure, assegna a un oggetto un punteggio, che non deve essere necessariamente un numero. «Il tumore della signora Johnson è con ogni probabilità benigno» è un giudizio, come lo sono affermazioni quali: «L'economia nazionale è molto instabile», «Fred Williams sarebbe la persona migliore da assumere come nuovo manager» e «Il premio per assicurare questo rischio dovrebbe essere di dodicimila dollari». I giudizi integrano in maniera informale diverse informazioni in una valutazione complessiva. Non si tratta di calcoli o dell'applicazione di regole precise: un insegnante usa il giudizio per valutare un tema, ma non per assegnare un voto a un quiz a risposta multipla.

Molte persone esprimono giudizi professionali per lavoro, con forti ripercussioni sulla vita di ognuno di noi. Tra i giudici di professione, come qui li abbiamo chiamati, si annoverano allenatori di calcio e cardiologi, avvocati e ingegneri, produttori cinematografici, sottoscrittori assicurativi e molti altri ancora. I giudizi professionali rappresentano il focus di questo libro, sia perché già ampiamente studiati, sia perché il modo in cui vengono formulati ha un grande impatto su tutti noi. Riteniamo inoltre che le nostre acquisizioni possano applicarsi anche a giudizi che vengono richiesti in altri ambiti della vita.

Esistono giudizi che abbiamo definito *predittivi*, e alcuni di questi sono verificabili: prima o poi scopriremo se erano corretti. Di solito questo è il caso delle previsioni a breve termine su risultati quali gli effetti di un trattamento medico, l'andamento di una pandemia o gli esiti di un'elezione. Ma una gran parte di questi, per esempio le previsioni a lungo termine e le risposte a domande fittizie, non sono verificabili, e la loro qualità si può valutare solo sulla base di quella del processo di pensiero da cui sono emersi. Peraltro, molti giudizi non sono predittivi ma *valutativi*: la condanna emessa da un giudice o la posizione che si aggiudica un dipinto nella classifica di un premio non sono confrontabili con un valore reale oggettivo.

Curiosamente, però, chi esprime un giudizio si comporta sempre come se questo valore reale esistesse, come se vi fosse un bersaglio invisibile a cui puntare, impossibile da mancare di molto. Con l'espressione *giudizio opinabile* si indica sia la possibilità di un disaccordo sia l'aspettativa che tale disaccordo sarà limitato. Le questioni di giudizio sono dunque caratterizzate dall'aspettativa di un *disaccordo limitato* e si collocano a metà tra le questioni di calcolo, in cui non è permesso alcun disaccordo, e quelle di gusto, in cui non ci si aspetta un grande accordo, eccetto in casi estremi.

Errori: bias e rumore

Diciamo che esiste un *bias* quando in un insieme di giudizi la maggior parte degli errori va nella stessa direzione. Il bias è l'*errore medio*, come, per esempio, quello che emerge quando al tiro a segno una squadra colpisce sistematicamente l'area in basso a sinistra del bersaglio; quando i dirigenti sono troppo ottimistici sulle vendite, anno dopo anno; quando una società continua a reinvestire in progetti fallimentari che dovrebbe invece abbandonare.

Eliminando il bias da un insieme di giudizi non si eliminerà del tutto l'errore. Gli errori che rimangono una volta rimosso il bias non sono condivisi: sono la divergenza non voluta nei giudizi, l'inaffidabilità dello strumento di misurazione che applichiamo alla realtà; sono il *rumore*, ovvero la variabilità di giudizi che dovrebbero essere identici. Il *rumore sistemico* è quello osservabile in organizzazioni che si avvalgono di professionisti intercambiabili per prendere decisioni, come i medici del pronto soccorso, i giudici che infliggono sanzioni penali e i sottoscrittori di una compagnia assicurativa. A questa tipologia di rumore è dedicata gran parte di questo libro.

Misurare il bias e il rumore

L'*errore quadratico medio* (o *MSE*) è da due secoli lo standard di accuratezza nelle misurazioni scientifiche. Le sue principali caratteristiche sono le seguenti: restituisce la media semplice come stima priva di bias della media della popolazione; tratta allo stesso modo errori positivi e negativi; penalizza gli errori gravi in maniera sproporzionata. L'*MSE* non riflette i costi reali degli errori di giudizio, che sono spesso asimmetrici. Detto ciò, le decisioni professionali richiedono sempre previsioni accurate: per una

città che si prepara ad affrontare un uragano i costi di una sottostima o di una sovrastima della minaccia chiaramente non sono gli stessi, ma sarebbe opportuno che tali costi non influenzassero le previsioni meteorologiche sulla velocità e la traiettoria del temporale. L'MSE è lo standard appropriato per esprimere giudizi predittivi di questo tipo, quando si punta all'accuratezza oggettiva.

Nella misurazione dell'MSE, bias e rumore vengono considerati due fonti indipendenti e cumulative di errore. Ovviamente il bias è sempre dannoso, e la sua riduzione aumenterà sempre l'accuratezza; meno intuitivo è il fatto che il rumore sia altrettanto dannoso, e la sua riduzione porti sempre a un miglioramento. Il miglior grado di dispersione è zero, anche quando i giudizi sono chiaramente affetti da bias. L'obiettivo, naturalmente, è ridurre al minimo sia il bias sia il rumore.

In un insieme di giudizi verificabili, il bias consiste nella differenza tra il giudizio medio su un caso e il corrispondente valore reale. Questo confronto è impossibile per i giudizi non verificabili: per esempio, il valore reale di un premio stabilito da un sottoscrittore per un particolare rischio non si potrà mai conoscere, così come quello di una condanna giusta per un particolare reato. Per questo motivo, spesso e volentieri (per quanto non sempre sia corretto) si parte dal presupposto che i giudizi siano privi di bias e che la media di molti giudizi sia la stima migliore del valore reale.

Il rumore di un sistema può essere valutato attraverso un *controllo del rumore*, un esperimento in cui diversi professionisti esprimono giudizi indipendenti sugli stessi casi (reali o fittizi). È possibile misurare il rumore senza conoscere un valore reale, proprio come è possibile vedere, sul retro di un bersaglio, la dispersione di una serie di colpi. I controlli del rumore sono in grado di misurare la variabilità dei giudizi in molti sistemi, come in un reparto di radiologia o nel sistema della giustizia penale; talvolta possono richiamare l'attenzione su qualche lacuna nelle competenze o

nella formazione, e sono in grado di quantificare il rumore sistemico, come quello che si presenta quando più sottoscrittori all'interno di uno stesso gruppo differiscono nella propria valutazione dei rischi.

Qual è il problema più grave, il bias o il rumore? Dipende dalla situazione, ma potrebbe benissimo essere quest'ultimo. Entrambi contribuiscono all'errore complessivo (MSE) nella stessa misura quando la media degli errori (il bias) è uguale alle deviazioni standard degli stessi (il rumore). Quando la distribuzione dei giudizi è normale (la classica curva a campana), gli effetti di bias e rumore sono uguali se l'84% dei giudizi si colloca al di sopra (o al di sotto) del valore reale: un bias considerevole, che spesso in un contesto professionale sarà facilmente identificabile. Quando il bias è inferiore a una deviazione standard, è il rumore la fonte principale dell'errore complessivo.

Il rumore è un problema

In alcuni giudizi la variabilità non è di per sé problematica, anzi può anche essere gradita. La diversità di opinioni è essenziale per formulare idee e opzioni, e i pareri contrastanti sono cruciali per l'innovazione: la pluralità di vedute tra critici cinematografici è un tratto caratteristico, non un intoppo; il disaccordo tra gli operatori finanziari è alla base dei mercati; le differenze strategiche tra startup rivali permettono agli investitori di scegliere quella che ritengono migliore. Nelle cosiddette questioni di giudizio, tuttavia, il rumore sistemico è sempre un problema: se due medici effettuano due diagnosi diverse, almeno uno avrà sbagliato.

Questo libro è nato dalla nostra sorpresa di fronte all'ubiquità del rumore sistemico e all'entità dei danni che esso produce, entrambe di gran lunga più consistenti di quanto ci si potrebbe aspettare. Abbiamo presentato esempi da molti campi, come il mondo degli affari, la medicina,

la giustizia penale, l'analisi delle impronte digitali, le previsioni finanziarie, le valutazioni del personale e la politica, giungendo alla seguente conclusione: dove c'è giudizio, c'è rumore, e più di quanto non si pensi.

Il grande peso del rumore nell'errore contraddice la credenza condivisa che gli errori casuali non contino, perché "si compensano a vicenda". Non è affatto così: se più tiri si sparpagliano in ogni punto del bersaglio, non aiuta dire che, in media, colpiscono il centro; se un candidato riceve una valutazione più alta di quanto meriti e un altro una più bassa, è possibile che venga assunta la persona sbagliata; se per una polizza assicurativa viene fissato un prezzo troppo alto e per un'altra un prezzo troppo basso, entrambi gli errori costeranno cari alla compagnia, in quanto uno le farà perdere un cliente, l'altro le farà perdere soldi.

Insomma, possiamo essere certi che siamo in presenza di un errore di fronte a giudizi che variano senza un valido motivo. Il rumore è dannoso anche quando i giudizi non sono verificabili e l'errore non è misurabile: non è giusto che persone nella stessa situazione ricevano un trattamento diverso, e un sistema i cui giudizi professionali siano ritenuti incoerenti perde di credibilità.

Tipi di rumore

Il rumore sistemico si può suddividere in *rumore di livello* e *rumore strutturale*. Alcuni giudici sono generalmente più severi, altri più clementi; alcuni analisti finanziari generalmente scommettono al rialzo e altri al ribasso nelle prospettive di mercato; alcuni medici prescrivono antibiotici più spesso di altri: sono tutti esempi di *rumore di livello*, cioè della variabilità dei giudizi medi formulati da individui diversi. L'ambiguità delle scale di giudizio è una delle fonti di rumore di livello. Parole come

“probabile” o anche numeri (per esempio “4 su una scala da 0 a 6”) non hanno lo stesso significato per persone differenti. Il rumore di livello è un’importante fonte di errore nei sistemi di giudizio, nonché uno dei principali bersagli a cui devono mirare gli interventi volti alla riduzione del rumore.

Il rumore sistemico comprende un’altra componente, generalmente più ampia. A prescindere dal livello medio dei loro giudizi, due giudici possono avere idee diverse su quali siano i crimini che meritano le condanne più severe, e le loro decisioni produrranno una diversa *classificazione* dei casi. Chiamiamo questa variabilità *rumore strutturale* (il termine tecnico è *interazione giudice × caso*).

La principale fonte di rumore strutturale è stabile, e consiste nella differenza tra le reazioni personali e idiosincratice dei giudici di fronte a uno stesso caso. Alcune di queste differenze riflettono principi o valori a cui gli individui si ispirano, in maniera non sempre conscia. Per esempio, un giudice potrebbe essere particolarmente severo con i taccheggiatori e insolitamente clemente con chi infrange il codice della strada, mentre un altro potrebbe mostrare una tendenza strutturale opposta. Alcuni dei principi o dei valori soggiacenti potrebbero essere piuttosto complessi, e il giudice potrebbe non esserne consapevole: per esempio, senza volerlo un giudice potrebbe essere più clemente verso i taccheggiatori anziani. Infine, anche una reazione altamente personale a un caso particolare potrebbe essere stabile: un’imputata che assomigli alla figlia del giudice potrebbe suscitare in lui la stessa compassione, e quindi la stessa clemenza, anche se l’udienza cadesse in un giorno diverso.

Questo *rumore strutturale stabile* riflette l’unicità dei giudici: la loro reazione ai casi è individuale come la loro personalità. Spesso le sottili differenze tra le persone sono piacevoli e interessanti, ma diventano problematiche quando dei professionisti operano in un sistema che

presuppone coerenza. Negli studi che abbiamo esaminato, il rumore strutturale stabile prodotto da queste differenze individuali è in genere la maggior fonte di rumore sistemico.

Detto ciò, gli atteggiamenti distintivi dei giudici in casi particolari non sono perfettamente stabili: il rumore strutturale ha anche una componente temporanea, che abbiamo chiamato *rumore occasionale*. Identifichiamo questo tipo di rumore quando un radiologo effettua diagnosi diverse a partire dalla stessa radiografia in giorni diversi, o se un esaminatore di impronte digitali identifica una corrispondenza tra due impronte in un'occasione ma non in un'altra. Come illustrano questi esempi, il rumore occasionale è più facile da misurare quando i giudici non ricordano di aver già esaminato in precedenza lo stesso caso. Un altro modo per dimostrare la presenza del rumore occasionale è evidenziare l'effetto che può avere sui giudizi un elemento contestuale irrilevante, come quando un giudice è più clemente dopo la vittoria della sua squadra del cuore o un medico prescrive più oppioidi di pomeriggio che di mattina.

La psicologia del giudizio e il rumore

I difetti cognitivi dei giudici non sono l'unica causa di errore nei giudizi predittivi: l'*ignoranza oggettiva* spesso riveste un ruolo ancora più rilevante. Certi fatti sono effettivamente inconoscibili – quanti nipoti avrà tra settant'anni un bambino che è nato ieri o quale sarà il numero del biglietto vincente in un'estrazione che verrà effettuata l'anno prossimo – mentre altri sono forse conoscibili, ma ignoti ai giudici. L'eccesso di fiducia nel proprio giudizio predittivo porta qualunque persona a sottostimare la sua ignoranza oggettiva oltre che i suoi bias.

Vi è un limite nell'accuratezza delle previsioni, e spesso è piuttosto basso. Ciò nonostante, in genere siamo sereni sui nostri giudizi; a darci questo fiducioso appagamento è un *segnale interno*, un riconoscimento che ci autoattribuiamo per avere inserito fatti e giudizi all'interno di una storia coerente. La fiducia soggettiva nei propri giudizi non è necessariamente legata a un'accuratezza oggettiva.

Molte persone restano sorprese dalla scoperta che l'accuratezza dei propri giudizi predittivi non solo è bassa, ma è anche inferiore a quella delle formule. Perfino dei modelli lineari semplici costruiti su dati limitati o regole elementari e improvvisate arrivano a risultati sistematicamente superiori ai giudizi umani. Il vantaggio cruciale di regole e modelli è che sono esenti da rumore. Nella nostra esperienza soggettiva il giudizio è un processo sofisticato e complesso, e non ci rendiamo conto che quella complessità potrebbe non essere altro che rumore. Ci è difficile immaginare che l'aderenza meccanica a delle semplici regole spesso porterà a un'accuratezza maggiore della nostra, anche se questo è ormai un fatto assodato.

I *bias psicologici*, naturalmente, sono una fonte di errore sistematico, o bias statistico, ma – in maniera meno scontata – sono anche una fonte di rumore. Quando non sono condivisi da tutti i giudici, quando sono presenti in misura diversa e quando i loro effetti dipendono da circostanze esterne, i bias psicologici generano rumore. Se, per esempio, metà dei responsabili delle assunzioni ha un pregiudizio contro le donne e metà ha un pregiudizio in loro favore, nel complesso non vi sarà un bias, ma il rumore sistemico causerà molti errori nelle assunzioni. Un altro esempio è l'effetto spropositato della prima impressione: pur essendo un bias psicologico, produrrà rumore occasionale se l'ordine in cui vengono presentate le caratteristiche dei candidati varia in maniera casuale.

Abbiamo descritto il processo di giudizio come l'integrazione informale di una serie di segnali per produrre un giudizio all'interno di una scala. Per eliminare il rumore sistemico, quindi, i giudici dovrebbero mantenere un'uniformità nell'impiego di questi segnali, nel peso da assegnare a ciascuno e nell'uso della scala. Anche lasciando da parte gli effetti casuali del rumore occasionale, queste condizioni vengono raramente soddisfatte.

L'accordo spesso è piuttosto elevato nei giudizi espressi su singole dimensioni: non di rado, per esempio, reclutatori diversi converranno su chi è più carismatico o diligente tra due candidati. Il processo intuitivo condiviso del *matching* nelle dimensioni di intensità, come l'associazione tra una media dei voti alta e un'età di lettura precoce, in genere produrrà giudizi simili. Lo stesso vale per i giudizi basati su un numero ridotto di segnali che puntano nella stessa direzione.

Le grandi differenze individuali emergono quando un giudizio richiede la *ponderazione di più segnali contrastanti*. Nel valutare lo stesso candidato, alcuni reclutatori daranno più peso alle caratteristiche dell'intelligenza o del carisma, mentre altri saranno più influenzati da aspetti come la diligenza o la calma in situazioni di stress. Quando i segnali sono contrastanti e non si inseriscono in una narrazione coerente, è inevitabile che persone diverse diano più peso ad alcuni e ne trascurino altri. Ciò produrrà rumore strutturale.

L'oscurità del rumore

Il rumore è un problema a cui non viene dato grande risalto. Se ne parla raramente, e spicca meno del bias; probabilmente voi stessi non ci avevate mai pensato più di tanto. Vista la sua importanza, l'oscurità del rumore è un fenomeno di per sé interessante.

Per spiegare i giudizi inadeguati, vengono spesso chiamati in causa i bias cognitivi e altre distorsioni del pensiero emotive o motivate. Gli analisti citano l'eccesso di fiducia, l'ancoraggio, l'avversione alla perdita, il bias di disponibilità e altri ancora per spiegare decisioni che si sono rivelate fallimentari. Tali spiegazioni risultano soddisfacenti, perché la mente umana non desidera altro che motivazioni causali: ogni volta che qualcosa va male, cerchiamo una causa, e spesso la troviamo. In molti casi, la identificheremo in un bias.

Il bias ha una sorta di fascino esplicativo che manca al rumore: se cerchiamo di spiegare, col senno di poi, perché una certa decisione era sbagliata, troveremo facilmente l'uno e mai l'altro. Solo una *visione statistica* del mondo ci permette di vedere il rumore, ma questa visione non ci viene naturale, perché preferiamo le narrazioni causali. L'assenza del pensiero statistico nelle nostre intuizioni è uno dei motivi per cui il rumore riceve molta meno attenzione rispetto al bias.

Un altro motivo è che raramente i professionisti sentono il bisogno di affrontare la presenza del rumore nei propri giudizi e in quelli dei loro colleghi. È raro che gli esperti di impronte digitali, i sottoscrittori navigati e i funzionari più anziani degli uffici brevetti si soffermino a immaginare il possibile disaccordo dei propri pari, e ancor più raro che contemplino la possibilità di non trovarsi d'accordo perfino con se stessi.

I professionisti hanno quasi sempre fiducia nei propri giudizi: si aspettano che i colleghi convengano con loro, e non cercano mai di capire se è davvero così. In molti contesti è possibile che un giudizio non venga mai confrontato con un valore reale e che, al massimo, venga sottoposto alla verifica di un altro professionista considerato un *esperto di rispetto*. Solo in rare occasioni ci si trova di fronte a un disaccordo inatteso e, quando ciò accade, in genere si riesce a trovare dei motivi per considerarlo un caso isolato. Di norma anche le organizzazioni tendono a

trascurare o eliminare le prove di eventuali divergenze tra i propri esperti, il che è del tutto comprensibile: dal punto di vista dell'organizzazione, il rumore è una fonte di imbarazzo.

Come ridurre il rumore (e anche il bias)

Vi è motivo di credere che i giudizi di alcuni siano migliori di quelli di altri: le competenze specifiche, l'intelligenza e uno stile cognitivo definito come *apertura mentale attiva* caratterizzano i giudizi più validi, e non sorprende che un buon decisore commetterà pochi errori madornali. Considerando le varie origini delle differenze individuali, tuttavia, non dobbiamo aspettarci che i migliori siano in perfetto accordo su problemi di giudizio complessi. L'infinita varietà delle circostanze individuali, delle personalità e delle esperienze che ci rendono unici è anche alla base dell'inevitabilità del rumore.

Una possibile strategia per la riduzione del rumore consiste nell'eliminazione dei bias. Di solito le persone cercano di rimuoverli dai propri giudizi o correggendo questi ultimi a fatto compiuto o contenendo i bias prima che incidano sul giudizio. Noi proponiamo una terza opzione, applicabile soprattutto nelle decisioni che vengono prese in contesti di gruppo: identificare i bias in tempo reale designando un *osservatore decisionale* per individuarne i primi segni (vedi appendice B).

Per ridurre il rumore nei giudizi consigliamo innanzitutto di praticare l'*igiene decisionale*. Abbiamo scelto questo termine perché la riduzione del rumore, come l'igiene personale, è una forma di prevenzione contro un nemico sconosciuto: come lavarsi le mani, per esempio, impedisce a patogeni ignoti di entrare nel nostro corpo, così l'igiene decisionale ci aiuterà a prevenire gli errori senza neanche sapere quali siano. Così come il suo nome, l'igiene decisionale non ha il fascino di una lotta vittoriosa

contro i bias prevedibili, ed è di certo meno emozionante. Forse prevenire un pericolo non identificato non vi coprirà di gloria, ma ne vale sicuramente la pena.

Un'azione di riduzione del rumore all'interno di un'organizzazione dovrebbe sempre partire da un controllo del rumore (vedi appendice A). Questo passaggio ha l'importante funzione di ottenere un impegno da parte dell'organizzazione a prendere sul serio il problema. Un vantaggio essenziale sta nella valutazione di diversi tipi di rumore.

Abbiamo descritto i successi e i limiti delle azioni di riduzione del rumore in vari campi. Ricapiteremo ora i sei principi di base dell'igiene decisionale, descriveremo come essi tengono conto dei meccanismi psicologici che provocano il rumore e mostreremo il loro legame con le specifiche tecniche di igiene decisionale già discusse.

L'obiettivo del giudizio è l'accuratezza, non l'espressione individuale. Questa affermazione è, a nostro avviso, il primo principio di igiene decisionale nei giudizi, e riflette l'accezione ristretta e specifica con cui abbiamo definito il giudizio in questo libro. Abbiamo mostrato come il rumore strutturale stabile sia un'ampia componente del rumore sistemico e una diretta conseguenza delle differenze individuali, delle personalità di giudizio che portano persone diverse a formarsi opinioni diverse su uno stesso problema. Questa osservazione ci porta a una conclusione forse impopolare ma incontrovertibile: il giudizio non è la sede appropriata per esprimere la propria individualità.

Sia chiaro, i valori personali, la singolarità e la creatività sono necessari, anzi essenziali, in molte fasi della riflessione e del processo decisionale come la scelta degli obiettivi, la formulazione di nuove modalità di approccio a un problema e l'elaborazione delle opzioni, ma quando si tratta di esprimere un giudizio su tali opzioni, le espressioni di individualità sono una fonte di rumore. Quando l'obiettivo è l'accuratezza

e vi aspettate che gli altri siano d'accordo con voi, dovrete anche chiedervi cosa penserebbero altri giudici competenti se fossero al posto vostro.

Un'applicazione radicale di questo principio consiste nella sostituzione del giudizio con regole o algoritmi, in quanto la valutazione algoritmica garantisce l'eliminazione del rumore, anzi è l'unico approccio in grado di rimuoverlo del tutto. Gli algoritmi sono già in uso in molti ambiti importanti, e il loro impiego è in crescita, ma è improbabile che sostituiscano il giudizio umano nella fase finale di una decisione importante, e questa a noi sembra una buona notizia. In ogni caso, i giudizi possono essere migliorati, sia con un utilizzo appropriato degli algoritmi, sia con l'adozione di approcci che rendano le decisioni meno dipendenti dalle idiosincrasie di un unico professionista. Abbiamo visto, per esempio, come semplici linee guida decisionali possano contribuire a limitare la discrezionalità dei giudici o favorire l'omogeneità nelle diagnosi mediche, riducendo il rumore e migliorando le decisioni.

Pensare in termini statistici e assumere la visione esterna del caso. Diciamo che chi giudica assume la visione esterna di un caso quando lo considera come un elemento di una classe di riferimento di casi simili, piuttosto che come un problema singolo. Questo approccio diverge dalla modalità di pensiero automatica, che si concentra unicamente sulla situazione in esame e la inserisce all'interno di una narrazione causale. Quando le persone applicano le proprie esperienze peculiari per formarsi un'opinione peculiare su un caso, producono rumore strutturale. La visione esterna è un rimedio a questo problema: i professionisti che si basano su una stessa classe di riferimento saranno meno inclini al rumore. Inoltre, porta spesso ad avere preziose intuizioni.

Il principio della visione esterna favorisce l'ancoraggio delle previsioni alle statistiche di casi simili, ricordando inoltre l'importanza di effettuare

previsioni moderate (il termine tecnico è *regressive*; vedi appendice c). L'attenzione a un'ampia gamma di risultati pregressi e alla loro limitata prevedibilità dovrebbe aiutare i decisori a calibrare la fiducia nei propri giudizi: non si può biasimare qualcuno per non aver previsto l'imprevedibile, ma lo si può criticare per la sua mancanza di umiltà nelle previsioni.

Strutturare i giudizi in diversi compiti indipendenti. Questo principio del *divide et impera* è reso necessario dal meccanismo psicologico che abbiamo definito *eccesso di coerenza*, che porta a distorcere o escludere informazioni che non si inseriscono in una narrazione preesistente o emergente. Quando le impressioni su diversi aspetti di un caso si contaminano a vicenda, l'accuratezza generale ne risente: per trovare un'analogia, pensate a come cambia il valore probatorio di un gruppo di testimoni se viene permesso loro di comunicare.

È possibile ridurre questo eccesso di coerenza scorporando il problema di giudizio in una serie di compiti più piccoli. Questa tecnica è analoga alla pratica dell'intervista strutturata, in cui i selezionatori valutano un tratto alla volta e gli attribuiscono un punteggio prima di passare al successivo. Il principio della strutturazione è alla base di linee guida diagnostiche come il punteggio di Apgar, ed è inoltre il fondamento di un approccio che abbiamo chiamato *protocollo a valutazioni intermedie*, in cui un giudizio complesso viene scomposto in più valutazioni basate sui fatti, per garantire che ciascuna di esse venga giudicata in maniera indipendente dalle altre. Ove possibile, per garantire tale indipendenza, ciascuna valutazione verrà assegnata a gruppi diversi, riducendo al minimo la comunicazione tra loro.

Resistere alle intuizioni premature. Abbiamo parlato del segnale interno, che dà ai decisori grande fiducia nei propri giudizi. La loro riluttanza ad abbandonare questo segnale gratificante è un motivo

essenziale della resistenza all'uso di linee guida, algoritmi e altre regole vincolanti. È chiaro che i decisori hanno bisogno di sentirsi tranquilli sulla propria scelta finale e di arrivare a questa gratificante sensazione di sicurezza intuitiva. Ma non dovrebbero correre troppo: una scelta intuitiva fondata su una considerazione equilibrata e attenta dei dati empirici sarà molto più valida di un giudizio avventato. L'intuizione non va bandita, ma dovrebbe essere informata, disciplinata e differita.

Sulla base di questo principio, raccomandiamo di *sequenziare le informazioni*: i professionisti che esprimono dei giudizi non dovrebbero ricevere dati in più che non servono e che, anche se corretti, potrebbero farli incorrere in un bias. Nelle scienze forensi, per esempio, è buona norma tenere gli esaminatori all'oscuro di eventuali informazioni aggiuntive su un sospetto. In questo discorso rientra anche il controllo dei punti all'ordine del giorno, un elemento chiave del protocollo a valutazioni intermedie: un ordine del giorno efficace porterà a considerare diversi aspetti del problema in momenti separati e a rimandare la formazione di un giudizio olistico finché il profilo della valutazione non sarà completo.

Ottenere giudizi indipendenti da più valutatori, per poi eventualmente aggregarli. Il requisito dell'indipendenza viene normalmente violato nelle procedure attuate dalle organizzazioni, soprattutto nelle riunioni in cui le opinioni dei partecipanti si formano sulla base di quelle altrui. A causa dell'effetto cascata e della polarizzazione di gruppo, spesso le discussioni non fanno che aumentare il rumore; la semplice procedura di raccogliere i giudizi dei partecipanti *prima* della discussione servirà a rivelare il grado di rumore e a facilitare un'armonizzazione costruttiva delle differenze.

Fare la media dei giudizi indipendenti garantisce la riduzione del rumore sistemico (ma non del bias): un giudizio singolo è un campione

della popolazione di tutti i giudizi possibili, e aumentare la dimensione del campione migliorerà la precisione delle stime. Il vantaggio della media sarà ancora maggiore quando i decisori hanno competenze diverse e strutture di giudizio complementari. La media di un gruppo affetto da rumore potrebbe rivelarsi più accurata di un giudizio unanime.

Preferire giudizi e scale relativi. I giudizi relativi sono meno soggetti al rumore di quelli assoluti, perché la nostra capacità di effettuare confronti tra coppie di elementi è molto più alta della nostra abilità nel classificare degli oggetti su una scala. Le scale di giudizio che richiedono dei confronti saranno molto meno affette da rumore di quelle che richiedono dei giudizi assoluti: per esempio, una *scala di casi* richiede che i valutatori collochino un singolo caso su una scala precedentemente definita attraverso esempi noti a tutti.

I principi di igiene decisionale qui elencati si possono applicare non solo ai giudizi ricorrenti, ma anche alle grandi decisioni irripetibili, che abbiamo chiamato *decisioni singole*. L'esistenza del rumore nelle decisioni singole potrebbe sembrare controintuitiva: per definizione, non vi è alcuna variabilità da misurare se si decide una volta sola. Eppure il rumore c'è, e produce errori: in una squadra di tiratori, il rumore è invisibile se vediamo solo il primo in azione, ma la dispersione diverrebbe chiara se vedessimo anche gli altri. Analogamente, il modo migliore per pensare ai giudizi singoli è trattarli come *giudizi ricorrenti formulati una volta sola*. Pertanto, l'igiene decisionale porterà un miglioramento anche in questi ultimi.

Applicare l'igiene decisionale può essere un compito ingrato: il rumore è un nemico invisibile, e una vittoria contro un nemico invisibile non potrà che essere una vittoria invisibile. Ma, al pari dell'igiene personale, anche quella decisionale ha un'enorme importanza. Dopo un intervento riuscito, tendiamo a pensare che siano state le competenze del chirurgo a salvarci la vita – e ovviamente è così – ma se il chirurgo e il personale

presente in sala operatoria non si fossero lavati le mani, forse saremmo morti. Forse l'igiene non sarà mai un motivo di gloria, ma darà sempre dei risultati.

Quanto rumore?

Naturalmente, la lotta contro il rumore non è l'unica preoccupazione di decisori e organizzazioni. Ridurre il rumore potrebbe essere troppo costoso: un istituto scolastico potrebbe arrivare a eliminarlo del tutto nelle valutazioni chiedendo a cinque insegnanti di leggere ogni elaborato, ma sarebbe un onere ingiustificato. Un certo livello di rumore può essere di fatto inevitabile, un effetto collaterale necessario all'interno di un sistema basato sul giusto processo, che concede a ogni caso un esame individuale, che non tratta le persone come rotelle di un ingranaggio e assegna ai decisori un senso di responsabilità. Un certo livello di rumore può perfino essere auspicabile, se la variazione così creata permette al sistema di adattarsi ai tempi, come quando il rumore riflette un cambiamento di valori e obiettivi, e accende un dibattito che porta a modifiche nella prassi o nella legge.

È importante ricordare che le strategie di riduzione del rumore potrebbero avere svantaggi inaccettabili. Molti timori sugli algoritmi sono esagerati, ma alcuni sono legittimi: gli algoritmi potrebbero produrre errori stupidi che un essere umano non commetterebbe mai, e quindi perdere credibilità, anche se allo stesso tempo riescono a prevenire diversi altri errori che gli esseri umani, invece, commettono eccome. Potrebbero essere affetti da bias per un difetto di progettazione o per essere stati allenati su dati inadeguati, e la loro impersonalità potrebbe non ispirare fiducia. Anche le pratiche di igiene decisionale hanno i loro lati negativi: se mal gestite, rischiano di burocratizzare le decisioni e

demoralizzare i professionisti che le percepiscono come una minaccia alla propria autonomia.

Tutti questi rischi e limitazioni meritano piena considerazione. Tuttavia, un'obiezione alla riduzione del rumore può avere senso solo se riferita a una particolare strategia attuata per ridurlo. Per esempio, un'obiezione all'aggregazione dei giudizi, motivata magari dai suoi costi eccessivi, potrebbe non riguardare invece l'impiego di linee guida. Sicuramente, qualora i costi della riduzione del rumore superassero i suoi vantaggi, sarebbe meglio rinunciare; una volta effettuato il calcolo di costi e benefici, si potrebbe anche scoprire che il livello di rumore ottimale non è pari a zero. Il problema è che, senza un apposito controllo, non è possibile sapere quanto rumore c'è nei nostri giudizi. Stando così le cose, appellarsi alla difficoltà di ridurre il rumore non è altro che una scusa per non misurarlo.

I bias conducono a errori e ingiustizie, ma lo stesso vale per il rumore, eppure ci sforziamo molto meno per ridurlo. L'errore di giudizio può sembrare più tollerabile quando è casuale che non quando lo attribuiamo a una causa, ma non per questo è meno dannoso. Se vogliamo arrivare a decisioni migliori su ciò che conta, dovremmo prendere sul serio la riduzione del rumore.

Epilogo

Un mondo con meno rumore

Immaginate come sarebbero le organizzazioni se venissero ripensate in modo da ridurre il rumore: ospedali, commissioni addette alle assunzioni, società di consulenza finanziaria, agenzie governative, compagnie assicurative, autorità sanitarie, sistemi di giustizia penale, studi legali e università sarebbero attentissimi al problema e farebbero di tutto per ridurlo. I controlli del rumore sarebbero di routine; forse verrebbero effettuati addirittura una volta all'anno.

I direttori delle organizzazioni impiegherebbero gli algoritmi in sostituzione o a integrazione del giudizio umano in molte più aree di quelle oggi interessate. Le persone scomporrebbero i giudizi complessi in valutazioni intermedie più semplici, sarebbero a conoscenza dell'igiene decisionale e ne seguirebbero le prescrizioni. Verrebbero richiesti giudizi indipendenti, successivamente aggregati. Le riunioni sarebbero molto diverse, le discussioni più strutturate. Nel processo decisionale verrebbe sistematicamente integrata una visione esterna, gli aperti contrasti sarebbero molto più frequenti e verrebbero risolti in maniera più costruttiva.

Ne risulterebbe un mondo con meno rumore, in cui si risparmierebbero molte risorse, migliorerebbero la sicurezza e la salute pubblica, aumenterebbe l'equità e si prevenirebbero molti errori evitabili. Il nostro obiettivo in questo libro era di portare l'attenzione su questa opportunità. Ci auguriamo che voi e altri riuscirete a coglierla.