

Cognizione, decisioni ed errori in medicina: *calibration, overconfidence e hindsight bias*

Matteo Motterlini

matteo.motterlini@hsr.it
Università Vita e Salute, San Raffaele

Abstract. In order to be useful (for example in pointing at an exam or a treatment) a medical diagnosis must be reliable: i.e. how probable is the diagnosis to be correct. In this paper I shall examine some specific sources of errors regarding the cognitive processes underlining our reasoning under risk and uncertainty. I shall show how such processes may systematically undermined the trustworthiness of probability judgments. I then raise the problem of how anyone can actually improve his own judgments and decisions by learning from experience.

Sommario. Perché una data diagnosi medica sia utile (per esempio nel suggerire un esame, o un trattamento) è necessario stimarne il grado di affidabilità. Un modo comune per farlo è formulare un giudizio probabilistico: quanto è probabile che la diagnosi sia corretta. In questo articolo vedremo alcune specifiche fonti di errore determinate dai processi cognitivi che presidono al nostro modo di ragionare in condizione di incertezza. Mostreremo quindi come tali processi influiscano sistematicamente sull'attendibilità dei nostri giudizi sollevando il problema di come si possa imparare dall'esperienza per poter giudicare e decidere meglio.

1. Il problema della calibrazione

Se un vostro amico sostenesse che c'è il 70% di probabilità che la squadra di casa vinca la partita di domenica prossima, come potreste mostrare che si sbaglia? Se anche la squadra di casa perdesse, egli potrebbe dire che si è verificato un evento che non aveva comunque escluso, assegnandogli semmai una probabilità del 30%. Tuttavia, se i risultati che il vostro amico giudica poco probabili si verificassero uno dopo l'altro per tutto il campionato, potreste a buon diritto sostenere che le sue stime di probabilità in fatto di sport sono inaffidabili – e che farebbe meglio a tenersi alla larga dalle scommesse.

Questo semplice esempio contiene un principio generale di grande importanza: per misurare l'affidabilità di una persona nel valutare la probabilità di un'ipotesi è spesso necessario considerare *interclassi* di eventi probabilistici. Applicando questo principio si può addirittura fissare un preciso livello di qualità *ideale*: i giudizi di un individuo si dicono *perfettamente calibrati* se esattamente l' $X\%$ (per esempio, il 75%) delle ipotesi cui egli attribuisce una probabilità del $X\%$ si rivelano effettivamente corrette. I giudizi di un soggetto perfettamente calibrato possono

<http://www.dif.unige.it/epi/networks>

Networks 5: 116-127, 2005

© SWIF - ISSN 1126-4780 <http://www.swif.uniba.it/lei/ai/networks/>

essere caratterizzati attraverso una precisa rappresentazione grafica, che riportiamo in figura 1: si tratta di una semplice retta con un'inclinazione di quarantacinque gradi, che rappresenta la perfetta corrispondenza fra stime probabilistiche e livello di accuratezza.

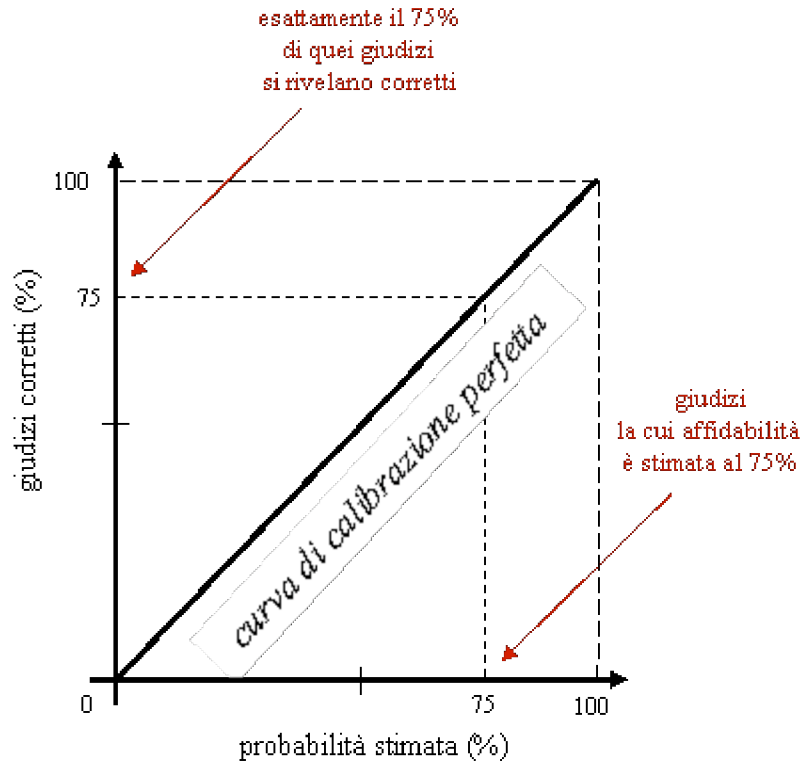


Figura 1. La curva di calibrazione perfetta nelle stime probabilistiche.

Chi giudica in modo perfettamente calibrato evita due errori simmetrici e opposti: *sovrastimare* (*overconfidence*) oppure *sottostimare* (*underconfidence*) la probabilità che le proprie ipotesi siano corrette. Le ricerche che studiano se le persone esprimono giudizi calibrati prendono in esame gli ambiti più disparati (psicologia clinica, analisi finanziaria, perizie legali, giochi d'azzardo, meteorologia ecc.). In generale, i giudizi ben calibrati sono l'eccezione, e non la regola. Molti esperimenti documentano una *sistematica sovrastima dell'affidabilità delle proprie ipotesi*: è normale osservare che persone comuni e esperti si dichiarino certi al 100% di giudizi che si rivelano corretti in non più dell'80% dei casi. (Fischhoff, Slovic, Lichtenstein, 1977).

Lo stesso fenomeno è riscontrabile in riferimento ai giudizi clinici – anch'essi non calibrati nel senso della sovrastima. Una ricerca sul campo ormai classica riguardava nove medici e un totale di 1531 pazienti con una forma acuta di tosse. (Christensen-

Matteo Motterlini

Szalanski, Bushyhead, 1981). I medici dovevano stimare, per ogni paziente, la probabilità che soffrisse di polmonite sulla base dell'esame fisico e dei dati raccolti riguardo alla storia clinica (ogni paziente veniva visitato da un solo medico). Per determinare la validità delle stime, tutti i pazienti venivano poi sottoposti a una radiografia al torace (comunemente considerata un test "definitivo" per la diagnosi di polmonite). Anche nei casi in cui i medici stimavano come superiore o molto superiore al 50% la probabilità che la diagnosi di polmonite fosse corretta, la percentuale di casi di polmonite, indicativa dell'effettiva accuratezza, restava ampiamente al di sotto del 30%. I risultati complessivi di questa indagine sono illustrati nella figura 2. (Si noti che l'*overconfidence* è rappresentata graficamente da curve che si collocano al di sotto di quella della calibrazione perfetta; le curve che si collocano al di sopra di essa rappresentano al contrario il fenomeno dell'*underconfidence*.)

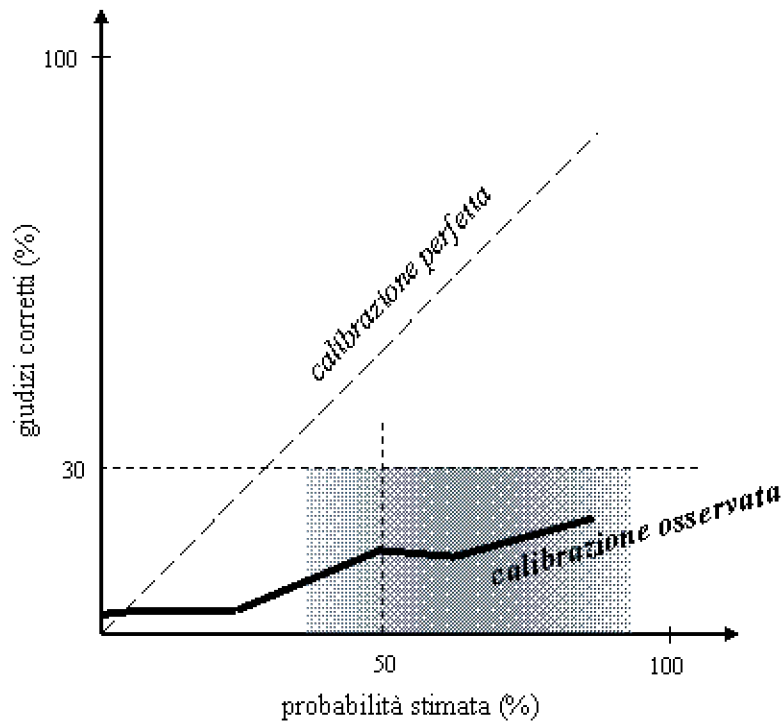


Figura 2. *Adattata da Christensen-Szalanski, Bushyhead (1981). Quando i medici considerano la diagnosi di polmonite probabile al 50% o più, meno del 30% dei corrispondenti pazienti ne sono realmente affetti – come mostra la curva relativa alla calibrazione osservata*

A distanza di qualche tempo, ai medici che avevano partecipato a questo studio è stato chiesto di valutare su una scala da -50 (valore associato a "la cosa peggiore che

Cognizioni, decisioni ed errori in medicina

io possa fare”) a +50 (valore associato a “la cosa migliore che io possa fare”) ognuno dei quattro scenari seguenti:

- | | | |
|--|---|------------------|
| (1) assegnare la diagnosi di polmonite a un paziente che in realtà ne è affetto | } | giudizi corretti |
| (2) escludere la diagnosi di polmonite in un paziente che in realtà non ne è affetto | | |
| (3) assegnare la diagnosi di polmonite a un paziente che in realtà non ne è affetto | } | giudizi erronei |
| (4) escludere la diagnosi di polmonite in un paziente che in realtà ne è affetto | | |

Le risposte a queste domande permettono di escludere che nell'indagine sulla diagnosi di polmonite i medici possano dare stime di probabilità eccessivamente alte per la preoccupazione di non lasciarsi sfuggire alcuno dei pazienti realmente malati, anche a costo di incrementare i falsi allarmi. Se questa spiegazione fosse corretta, infatti, i medici avrebbero dovuto valutare lo scenario (1) più desiderabile del (2) e lo scenario (4) più grave del (3). Al contrario, non si è osservata alcuna differenza significativa né nel valore che i medici hanno assegnato ai due tipi di giudizi corretti, né nella valutazione della gravità che essi hanno associato ai due diversi tipi di errore. In breve, “la distorsione dei giudizi verso la sovrastima è di tipo cognitivo e non motivazionale”. Christensen-Szalanski, Bushyhead (1981, p. 364).

In un lavoro recente sono stati indagati casi di sospetta batteriemia (la presenza di batteri nel sangue). I medici partecipanti avevano ordinato ad alcuni loro pazienti esami per la coltura di organismi batterici in campioni di sangue (emocultura). Poco dopo che gli esami erano stati prescritti si chiedeva ai medici di stimare la probabilità che i risultati rivelassero la presenza di batteriemia. I medici coinvolti in questa ricerca erano in totale 34, i pazienti 227. Per i 71 pazienti per cui le stime dei medici variavano fra il 16% e il 40% (media: 26,8%), l'*effettiva* proporzione di casi di batteriemia documentata dagli esami era del 4,2%. Per il gruppo di 74 pazienti per cui le stime variavano fra 41% e 99% (media: 62%), l'*effettiva* proporzione di casi di batteriemia era del 12,2%; solo due dei cinque pazienti (dunque il 40%) per i quali i medici si erano detti *certi* della diagnosi di batteriemia (stima di probabilità = 100%) risultarono *effettivamente* positivi ai test.

Come si può constatare, l'*overconfidence* rappresenta un problema serio per il giudizio degli esperti. Persiste, e talvolta risulta persino aggravata, in condizioni che intuitivamente dovrebbero contenerne gli effetti. Per esempio, si è potuto osservare che l'esperienza clinica tende ad accrescere l'*overconfidence*, anziché ridurla. Nei più esperti, cioè, la sovrastima dell'accuratezza delle proprie opinioni professionali è spesso *maggiore* che nei meno esperti (come documentato da uno studio svolto a Cleveland con 198 medici alle prese con 864 pazienti in condizioni cliniche critiche - Dawson et al., 1993).

A prima vista può apparire naturale interpretare la *sicurezza* soggettiva che un medico ripone nel suo giudizio come un'indice dell'*accuratezza* di quel particolare

Matteo Motterlini

giudizio. Questo presupposto svolge un ruolo decisivo nella comunicazione fra medico e paziente. Gli studi cognitivi sull'*overconfidence* mettono seriamente in discussione questa assunzione di senso comune: la sicurezza soggettiva del medico (misurata dalle stime probabilistiche) non sempre è un indice dell'affidabilità delle sue ipotesi.

2 Overconfidence

Una plausibile ipotesi di lavoro individua nella tendenza a sottovalutare le probabilità iniziali una delle cause principali dell'*overconfidence* (Koehler, Brenner, Griffin 2002). Vediamo qual è la possibile connessione fra i due fenomeni.

A scopo illustrativo, immaginiamo che un medico debba stimare la probabilità di un'ipotesi h (una diagnosi, una prognosi, o altro) alla luce della presenza di un indicatore clinico mediamente rilevante e , per esempio presente in 6 su 10 casi se h è vera e solo in 2 su 10 se h è falsa. Le ricerche mostrano che è probabile che egli basi la sua stima prevalentemente sui dati clinici del caso individuale e trascuri $p(h)$, comportandosi come se la probabilità iniziale di h e di non- h si aggirassero intorno al 50%. (Wallsten, 1981). Facendo un'assunzione di questo tipo, il medico potrebbe concludere che la probabilità di h alla luce di e è di 6 contro 2, cioè del 75% (6 diviso per 6+2). Ma se la corretta probabilità iniziale fosse molto bassa, diciamo del 10% (su un totale di 10 casi, 1 contro 9), il medico avrebbe *drammaticamente sopravvalutato* la probabilità di h alla luce di e . Infatti, per il teorema di Bayes in forma di rapporti (R):

$$\frac{p(h|e)}{p(\text{non-}h|e)} = \frac{6}{2} \cdot \frac{1}{9} = \frac{6}{18}$$

Il che significa che la probabilità di h alla luce di e è in realtà del 25% (6 diviso per 6+18), cioè un terzo di quella stimata.

Viceversa, se la probabilità iniziale corretta fosse piuttosto alta, diciamo dell'80%, il medico avrebbe *significativamente sottovalutato* la probabilità che l'ipotesi sia corretta. Infatti, in questo caso:

$$\frac{p(h|e)}{p(\text{non-}h|e)} = \frac{6}{2} \cdot \frac{8}{2} = \frac{48}{4}$$

Il che implica che stavolta $p(h|e)$ va a più del 92% (48 diviso per 48+4).

Per illustrare come giudizi più o meno calibrati possano dipendere dal trascurare le probabilità iniziali è utile considerare i risultati ottenuti in uno studio condotto nel 1997 in tre diversi ospedali americani. Poses et al. (1997). I medici coinvolti hanno fornito stime di probabilità relative a più di mille pazienti colpiti da insufficienza cardiaca acuta riguardo a due esiti prognostici: *sopravvivenza a tre mesi* (90 giorni) e *sopravvivenza a un anno*. I giudizi dei medici sono stati meglio calibrati nel secondo

Cognizioni, decisioni ed errori in medicina

compito che non nel primo. Per esempio, fra i pazienti cui i medici assegnavano una probabilità di sopravvivenza del 60% a tre mesi, quasi l'80% sopravviveva realmente (*underconfidence*), mentre le loro stime di una probabilità di sopravvivenza del 60% a un anno rispecchiavano fedelmente il tasso di sopravvivenza realmente osservato nel gruppo di pazienti corrispondente.

Come si può spiegare questa strana forma di “presbiopia predittiva” – previsioni più accurate per un futuro più lontano? La chiave dell'enigma potrebbe risiedere proprio nelle probabilità iniziali. In questo campione di pazienti colpiti da insufficienza cardiaca acuta, il tasso generale di sopravvivenza a tre mesi era piuttosto alto (81%). Ammettendo che i medici lo abbiano trascurato, ci si dovrebbe aspettare precisamente che essi forniscano stime di probabilità di sopravvivenza sistematicamente pessimistiche (cioè non calibrate *per difetto*), come in effetti è avvenuto. Per contro, per quegli stessi pazienti il tasso di sopravvivenza a un anno era – come si può immaginare – sensibilmente inferiore (64%). Ci si dovrebbe perciò aspettare che gli effetti del trascurare le probabilità iniziali siano in questo caso più modesti, e le stime dei medici migliori. E, come abbiamo visto, questo è proprio quanto si è osservato.

Vari studi condotti in ambito medico illustrano una preoccupante e sistematica *overconfidence* in problemi con probabilità iniziali molto basse (soprattutto eventi patologici rari nelle popolazioni di riferimento), (Bobbio et al. (1992), DeSmet, Fryback, Thornbury (1979) e Poses et al. (1985)). Una significativa *underconfidence* in presenza di probabilità iniziali alte (per esempio, per previsioni di mortalità in reparti di terapia intensiva), (Arkes et al. (1995), Detsky et al. (1981), McClish, Powell (1989) e Poses et al. (1990)).

e giudizi relativamente ben calibrati quando le probabilità iniziali sono moderate (non troppo lontane da 50%) (MacKillop, Quirt (1997), Tierney et al. (1986) e Weeks et al. (1998)).

È improbabile che questi risultati riflettano la *mancata conoscenza* della probabilità iniziale pertinente (cioè della prevalenza della condizione in questione nella popolazione di riferimento) da parte dei medici, perché negli studi in cui viene chiesto loro espressamente di stimare questo tipo di dati i medici sono normalmente piuttosto accurati, con valutazioni spesso vicine a quelle fornite dalla letteratura scientifica. È pertanto più plausibile pensare che l'errore che distorce i giudizi risieda non nella disponibilità dell'informazione, quanto nell'*elaborazione* dell'informazione disponibile.

La diffusione dell'*overconfidence* negli studi sul giudizio medico potrebbe quindi almeno in parte rispecchiare il fatto che le condizioni da diagnosticare o prevedere sono spesso relativamente rare nelle popolazioni di riferimento. Poiché la bassa probabilità iniziale, anche se nota, viene trascurata, le conseguenze della presenza di dati clinici specifici risultano sopravvalutate. Nei casi in cui le probabilità iniziali sono alte tenderebbe invece a verificarsi il fenomeno opposto.

Si noti che entrambi i tipi di giudizi scarsamente calibrati (per eccesso oppure per difetto) rappresentano un potenziale problema in relazione all'impiego efficace di strumenti di sostegno per la diagnosi e la decisione medica, che si tratti di tecniche formali, linee-guida cartacee, sistemi informatici, o altro ancora. Il punto è che

Matteo Motterlini

l'utilità effettiva di questi strumenti non sempre è percepita dall'utente: in particolare, se il medico *sottostima* la propria accuratezza, ricorrerà a strumenti di sostegno anche quando potrebbero essere superflui (disperdendo risorse); se la *sovrastima*, non vi farà ricorso, e le sue prestazioni saranno soggette a errori che si sarebbero potuti evitare.

3 Imparare dall'esperienza

Chiediamoci ora come sia possibile che una sistematica sopravvalutazione delle proprie abilità di giudicare possa *protrarsi nel tempo*. Sul lungo periodo, infatti, è ragionevole immaginare che essa possa essere corretta ed eliminata sulla base dell'informazione che l'ambiente fornisce circa la frequenza dei nostri errori di valutazione. Come ora vedremo, uno dei motivi per cui è possibile che sopravvalutiamo sistematicamente ciò che sappiamo è il fatto che sopravvalutiamo sistematicamente ciò che *sapevamo* o che *avremmo potuto sapere*.

L'esperienza di tutti i giorni ci insegna che c'è una grossa differenza fra predire gli sviluppi futuri di una situazione e spiegare il corso di eventi già accaduti. Col *senno di poi*, infatti, siamo tutti più bravi.

Gli esseri umani sembrano avere una speciale attitudine a dare senso agli eventi passati, descrivendoli come conseguenze inevitabili (o quasi) di condizioni che erano presenti fin dall'inizio. Talvolta, questo nostro peculiare talento ci induce a credere fallacemente che fatti già avvenuti si sarebbero potuti (o dovuti) prevedere, data la situazione nota in anticipo. Quello che chiamiamo "senno di poi" si fonda spesso su un fatalismo riguardo al passato che è il frutto inconsapevole e fuorviante di uno sguardo retrospettivo. Questo fenomeno è noto agli psicologi cognitivi come *hindsight bias*, vale a dire una sorta di "distorsione retrospettiva del giudizio" messa in luce in una serie di ingegnosi lavori sperimentali da Baruch Fischhoff (1975 e 1980).

L'*hindsight bias* proietta sul piano della pratica medica le considerazioni di senso comune esposte in precedenza, e mette in evidenza come ragionare con il senno di poi possa incidere negativamente sull'affidabilità dei giudizi, e ridurre seriamente le possibilità di *imparare dall'esperienza clinica*.

In una importante indagine sull'argomento, condotta presso il Medical College del Wisconsin, si presentava a quindici medici il seguente problema (Arkes et al. 1981):

Questa è la descrizione di un caso clinico. Vorremmo che la leggessi e decidessi quale probabilità avresti assegnato a ciascuna delle quattro possibili patologie, se ti fossi trovato a fare una diagnosi. Assegna una probabilità a ogni diagnosi, accertandoti che le probabilità sommino al 100%.

Un barista di trentasette anni è stato in buona salute finché recentemente ha sviluppato un dolore di crescente intensità al ginocchio sinistro, che è diventato caldo e gonfio. Qualche giorno dopo si sono manifestati dolore, gonfiore e calore al polso sinistro e al ginocchio destro. L'esame fisico ha in effetti rivelato gonfiore, riscaldamento e versamento a entrambe le ginocchia e al polso sinistro. Non ci sono deformità. Il fegato risulta ingrandito di due

Cognizioni, decisioni ed errori in medicina

centimetri al di sotto del livello del costato. I risultati dell'esame emocromocitometrico sono normali. L'esito del test della velocità di eritrosedimentazione (VES) è 30 mm. Un Latex test è risultato negativo. Non sono ancora disponibili i risultati dei test CHEM-7 sul siero sanguigno e degli antigeni Hb_sAg. Il liquido sinoviale risulta contenere 20.000 globuli bianchi per mm cubo, con 80% di neutrofili e bassa viscosità. Si è trovato pus in eccesso nelle urine. Febbre a 38°.

Ora ti chiediamo di assegnare a ognuna delle quattro possibili diagnosi la probabilità che ritieni avresti assegnato. Accertati che le probabilità sommino a 100%.

- sindrome di Reiter (incompleta)
- artrite batterica da streptococchi
- gotta
- epatite da siero in fase anitterica

In media, i partecipanti assegnarono alla diagnosi di epatite da siero una probabilità del 16% circa. Le risposte ottenute da questo gruppo rispecchiano giudizi clinici in una condizione di autentica incertezza. Per contro, un secondo gruppo di quindici medici affrontava un questionario del tutto simile, tranne che per una importante differenza: in questo caso, la frase di apertura era *“questa è la descrizione di un caso di epatite da siero in fase anitterica”*. Stavolta, tredici medici su quindici assegnarono alla diagnosi di epatite una probabilità significativamente maggiore del 16%, con una media del 38%. Si è quindi potuto osservare un risultato simile con un'altra della quattro diagnosi (artrite batterica da streptococchi) in un ulteriore campione di quindici medici (media delle stime: 11% nel gruppo di controllo originario, 31% nel gruppo di studio con il questionario modificato).

Il solo fatto di sapere *a posteriori* qual era la diagnosi corretta, ci induce a ritenere sistematicamente più probabile quella diagnosi alla luce dei dati clinici che erano disponibili fin dal principio. In altre parole: le stesse informazioni inizialmente a disposizione per una previsione sembrano indicare con più forza verso un certo esito *dopo* aver scoperto che quel particolare esito si è verificato.

Dati sperimentali come quelli descritti hanno importanti implicazioni riguardo alle formazioni dei medici. Per esempio, le “conferenze clinicopatologiche” sono state considerate un utile strumento educativo, e in molte istituzioni sono parte integrante della formazione medica. In una conferenza clinicopatologica, si chiede a un medico di formulare una diagnosi riguardo a un certo paziente reale (anonimo) sulla base della storia clinica e dei risultati di alcuni esami preliminari. In seguito, si rivela all'uditorio la diagnosi corretta. L'idea è che l'esposizione a questo tipo di esercizio “per casi concreti” possa incoraggiare una valutazione critica delle proprie capacità diagnostiche. Il fenomeno dell'*hindsight bias* mette però in discussione questa possibilità. Infatti, quando il caso viene presentato, un partecipante potrebbe stimare la probabilità della diagnosi che si rivelerà corretta in una certa misura. Ma, se interrogato dopo che la soluzione è stata svelata, spesso egli sarà convinto che, dovendo affrontare un caso come quello discusso, avrebbe certamente associato a quella diagnosi una stima di probabilità più alta. In questo modo egli *si inganna* sulla reale qualità del suo giudizio clinico. E l'inganno può essere sistematico.

Matteo Motterlini

La rilevanza di queste osservazioni è stata mostrata in uno studio che ha coinvolto centosessanta partecipanti di conferenze clinicopatologiche tenutesi presso il Cleveland Metropolitan General Hospital (Dawson et al. 1988). Una parte di loro valutava la probabilità delle principali alternative di diagnosi *prima* che la diagnosi corretta venisse comunicata: solo nel 30% dei casi la diagnosi corretta era classificata come la più probabile. Un secondo gruppo di soggetti svolgeva lo stesso compito *dopo* che la diagnosi corretta era stata comunicata: questa volta il 50% era pronto a dichiarare che quella era la più probabile alla luce del quadro clinico inizialmente a disposizione.

Gli studi hanno inoltre documentato che uno sguardo retrospettivo può avere un'influenza fuorviante sul modo in cui valutiamo non solo i giudizi, ma anche le *decisioni*.

Consideriamo due casi di una stessa operazione chirurgica (come un taglio cesareo) nei quali viene effettuato un certo intervento anestesiológico (per esempio, l'anestesia spinale). Assumiamo che i due casi siano fra loro identici (condizioni cliniche della paziente, tipo di intervento praticato, dosi impiegate ecc.), tranne che per un aspetto: in un caso si verifica una complicazione che produce un esito clinico avverso *temporaneo* (per esempio, un arresto cardiaco durante l'operazione, a seguito del quale la madre e il neonato si riprendono pienamente); nell'altro caso, invece, lo stesso tipo di complicazione produce un esito clinico avverso *permanente* (per esempio, l'arresto cardiaco causa la morte della partoriente e danni cerebrali al neonato). In un sofisticato esperimento, un gruppo di studiosi del Virginia Mason Medical Center e della Facoltà di Medicina dell'Università di Washington ha costruito, a partire da episodi reali, ventuno coppie di casi simili a quella appena descritta. (Per la prima osservazione sistematica e la definizione del fenomeno – noto come *outcome bias* – si veda Baron, Hershey, 1988). Centododici anestesisti sono quindi stati suddivisi in due gruppi. Ai partecipanti del primo gruppo venivano presentati ventuno casi, ognuno dei quali selezionato casualmente all'interno di ciascuna delle due possibilità (esito avverso temporaneo oppure permanente). Ai partecipanti del secondo gruppo venivano presentati i ventuno casi restanti. Il compito dei medici era quello di esprimere una valutazione sul trattamento anestesiológico praticato in ognuno dei casi che si trovavano di fronte, classificandolo come “appropriato”, “non appropriato”, o “impossibile da giudicare”.

I risultati di questa indagine mostrano che il senno di poi distorce significativamente la valutazione dell'appropriatezza di una specifica decisione clinica. Si è infatti potuto rilevare che, in ben quindici delle ventuno coppie di casi considerati, la valutazione era prevalentemente positiva (“intervento appropriato”) in presenza di un danno temporaneo, e prevalentemente negativa (“intervento inappropriato”) in presenza di un danno permanente – a dispetto del fatto che, come abbiamo detto, i due casi erano identici tranne che per l'estensione del danno procurato. Un danno permanente rende pertanto il giudizio a posteriori molto più severo del giudizio basato sugli *stessi elementi* qualora l'esito dell'intervento si sia rivelato meno grave.

4 Se dall'esperienza non si impara

Considerare i successi e fallimenti dei nostri giudizi, e delle decisioni basati su di essi è un meccanismo fondamentale dell'apprendimento e della crescita della conoscenza, e in molti casi è un esercizio perfettamente razionale che ci riesce con successo. Gli esperimenti appena descritti mostrano però che in determinati contesti, ben documentati e prevedibili, questa naturale attitudine può essere sistematicamente distorta dal senno di poi.

Con il senno di poi, è relativamente facile concludere (spesso erroneamente) che noi avremmo saputo giudicare e scegliere in modo da ottenere risultati altrettanto buoni o anche migliori di quelli osservati. Si pensi a un giovane praticante che lavora a contatto con colleghi più esperti e ufficialmente responsabili. Egli avrà in molti casi accesso ai dati clinici, alle diagnosi e alle decisioni dei suoi colleghi e, soprattutto, agli esiti dei singoli casi. In queste condizioni, sarà probabilmente indotto a sovrastimare l'accuratezza del proprio giudizio diagnostico e le sue capacità decisionali, sottovalutando sensibilmente la differenza fra elaborare correttamente una diagnosi o prendere una decisione appropriata e rifletterci su una volta che la storia ha fatto il suo corso.

Gli studi qui considerati, e molte altre indagini simili, suggeriscono i diversi modi in cui gli errori di valutazione possono indebolire gli effetti dell'apprendimento attraverso l'esperienza. I medici devono far fronte a una realtà complessa e incerta, e scenari come quelli appena descritti possono essere facilmente riprodotti nel contesto clinico. Per esempio, un medico, di fronte a un paziente con un certo quadro clinico, può elaborare la diagnosi in maniera scorretta o prescrivere un trattamento inadeguato. L'osservazione di un esito clinico positivo, che è comunque possibile per diverse ragioni (come un recupero spontaneo), può incoraggiarlo a procedere nello stesso modo in un successivo caso simile. Viceversa, un esito sfortunato può indurre un medico a mettere erroneamente in discussione un'inferenza diagnostica o una scelta terapeutica corretta, basate su conoscenze e ragionamenti solidi, e a modificare in peggio la sua pratica successiva. Ragionare con il senno di poi può quindi corroborare pratiche inadeguate, e suggerire l'abbandono ingiustificato di pratiche appropriate. Imparare dal passato può sembrare semplice e intuitivo; in realtà è un percorso pieno di insidie.

Ringraziamenti. Desidero esprimere la mia riconoscenza a Baruch Fischhoff, Vincenzo Crupi, Massimo Egidi, Vittorio Girotto Gian Franco Gensini, Massimo Piattelli Palmarini e Paolo Vineis, con i quali ho discusso le idee di questo contributo. Ringrazio inoltre il Department of Social and Decision Sciences della Carnegie Mellon University per l'ospitalità; e il Laboratorio di Scienze Cognitive dell'Università di Trento, l'Istituto per la Ricerca Scientifica e Tecnologica dell'Istituto Trentino di Cultura - nell'ambito delle attività dell'Unità di Informatica Medica e Telemedicina, l'Istituto per lo Studio e la Cura dei Tumori di Milano che hanno sostenuto questa ricerca.

Matteo Motterlini

Riferimenti bibliografici

- ARKES, H.R., DAWSON, N.V., SPEROFF, T., HARRELL, F.E. JR., ALZOLA, C., PHILLIPS, R., DESBIENS, N., OYE, R.K., KNAUS, W., CONNORS, A.F. JR, THE SUPPORT INVESTIGATORS (1995), "The covariance decomposition of the probability score and its use in evaluating diagnostic estimates". In *Medical Decision Making*, 15, pp. 120-131.
- ARKES, H.R., WORTMANN, R.L., SAVILLE, P.D., HARKNESS, A.R. (1981), "Hindsight bias among physicians weighing the likelihood of diagnoses". In *Journal of Applied Psychology*, 66, pp. 252-254.
- BARON, J., HERSHEY, J.C. (1988), "Outcome bias in decision evaluation". In *Journal of Personality and Social Psychology*, 54, pp. 569-579.
- BOBBIO, M., DEORSOLA, A., PISTIS, G., BRUSCA, A., DIAMOND, G.A. (1988), "Physician perception of exercise electrocardiography as a prognostic test after acute myocardial infarction". In *American Journal of Cardiology*, 62, pp. 675-678.
- CAPLAN, R.A., POSNER, K.L., CHENEY, F.W. (1991), "Effect of outcome on physicians' judgments of appropriateness of care". In *Journal of the American Medical Association*, 265, pp. 1957-1960.
- CHRISTENSEN-SZALANSKI, J.J.J., BUSHYHEAD, J.B. (1981), "Physicians' use of probabilistic information in a real clinical setting". In *Journal of Experimental Psychology: Human Perception and Performance*, 7, pp. 928-935.
- DAWSON, N. V., CONNORS, A. F., JR., SPEROFF, T., KEMKA, A., SHAW, P., ARKES, H. R. (1993), "Hemodynamic assessment in managing the critically ill: is physician confidence warranted?". In *Medical Decision Making*, 13, pp. 258-266.
- DAWSON, N.V., ARKES, H.R., SICILIANO, C., BLINKHORN, R., LAKSHMANAN, M., PETRELLI, M. (1988), "Hindsight bias: an impediment to accurate probability estimation in clinicopathologic conferences". In *Medical Decision Making*, 8, pp. 259-264.
- DESMET, A.A., FRYBACK, D.G., THORNBURY, J.R. (1979), "A second look at the utility of radiographic skull examination for trauma". In *American Journal of Roentgenology*, 132, pp. 95-99.
- DETSKY, A.S., STRICKER, S.C., MULLEY, A.G., THIBAUT, G.E. (1981), "Prognosis, survival, and the expenditure of hospital resources for patients in an intensive-care unit". In *New England Journal of Medicine*, 305, pp. 667-672.
- EINHORN, H.J. (1982), "Learning from experience and suboptimal rules in decision making". In KAHNEMAN, D., SLOVIC, P., TVERSKY, A. (a cura di), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, pp. 269-283.
- FISCHHOFF, B. (1975), "Hindsight \neq foresight: the effect of outcome knowledge on judgment under uncertainty". In *Journal of Experimental Psychology: Human Perception and Performance*, 1, pp. 288-299.
- FISCHHOFF, B. (1980), "For those condemned to study the past". In SCHWEDER, R.A., FISKE, D.W. (a cura di), *New Directions for Methodology of Social and Behavioral Science*. Jossey-Bass, San Francisco (CA), pp. 79-93.
- FISCHHOFF, B., SLOVIC, P., LICHTENSTEIN, S. (1977), "Knowing with certainty: the appropriateness of extreme confidence". In *Journal of Experimental Psychology: Human Perception and Performance*, 3, pp. 552-564.

Cognizioni, decisioni ed errori in medicina

- FRIEDMAN, C., GATTI, G., ELSTEIN, A., FRANZ, T., MURPHY, G., WOLF, F. (2001), "Are clinicians correct when they believe they are correct? Implications for medical decision support". In *Medinfo*, 10, pp. 454-458.
- KOEHLER, D.J., BRENNER, L., GRIFFIN, D. (2002), "The calibration of expert judgment". In GILOVICH, T., GRIFFIN, D., KAHNEMAN, D. (a cura di), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, New York, pp. 686-715.
- MACKILLOP, W.J., QUIRT, C.F. (1997), "Measuring the accuracy of prognostic judgments in oncology". In *Journal of Clinical Epidemiology*, 50, pp. 21-29.
- MCCLISH, D.K., POWELL, S.H. (1989), "How well can physicians estimate mortality in a medical intensive care unit?". In *Medical Decision Making*, 9, pp. 125-132.
- POSES, R. M., CEBUL, R. D., COLLINS, M., FAGER, S.S. (1985), "The accuracy of experienced physicians' probability estimates for patients with sore throats". In *Journal of the American Medical Association*, 254, pp. 925-929.
- POSES, R.M. ANTHONY, M. (1991), "Availability, wishful thinking, and physicians' diagnostic judgment for patients with suspected bacteremia". In *Medical Decision Making*, 11, pp. 159-168.
- POSES, R.M., BEKES, C., COPARE, F.J., SCOTT, W.E. (1990), "What difference do two days make? The inertia of physicians' sequential prognostic judgments for critically ill patients". In *Medical Decision Making*, 10, 6-14.
- POSES, R.M., SMITH, W.R., MCCLISH, D.K., HUBER, E.C., CLEMO, F.L., SCHMITT, B.P., ALEXANDER-FORTI, D., RACHT, E.M., COLENDIA, C.C. III, CENTOR, R.M. (1997), "Physicians' survival predictions for patients with acute congestive heart failure". In *Archives of Internal Medicine*, 157, pp. 1001-1007.
- TIERNEY, W.M., FITZGERALD, J., MCHENRY, R., ROTH, B.J., PSATY, B., STUMP, D.L., ANDERSON, F.K. (1986), "Physicians' estimates of the probability of myocardial infarction in emergency room patients with chest pain". In *Medical Decision Making*, 6, pp. 12-17.
- WALLSTEN, T.S. (1981), "Physician and medical student bias in evaluating diagnostic information". In *Medical Decision Making*, 1, pp. 145-164.
- WEEKS, J.C., COOK, E.F., O'DAY, S.J., PETERSON, L.M., WENGER, N., REDING, D., HARRELL, F.E., KUSSIN, P., DAWSON, N.V., CONNORS, A.F. JR, LYNN, J., PHILLIPS, R.S. (1998), "Relationship between cancer patients' predictions of prognosis and their treatment preferences". In *Journal of the American Medical Association*, 279, pp. 1709-1714.