

- Schanck R.L., Goodman C. (1959), *Reactions to Propaganda on Both Sides of a Controversial Issue*, «Public Opinion Quarterly», 3, pp. 107-112.
- Schwartz R.D. (1961), *Field Experimentation in Sociological Research*, «Journal of Legal Education», 13, pp. 401-410.
- Schwartz R.D., Orleans S. (1967), *On Legal Sanctions*, «University of Chicago Law Review», 34, 274-300.
- Schwartz R.D., Skolnick J.H. (1963), "Televised Communication and Income Tax Compliance", in Arons L., May M. (a cura di), *Television and Human Behavior*, Appleton-Century-Crofts, New York.
- Selvin H. (1957), *A Critique of Tests of Significance in Survey Research*, «American Sociological Review», 22, pp. 519-527.
- Simon J.L. (1966), *The Price Elasticity of Liquor in the U.S. and a Simple Method of Determination*, «Econometrica», 34, pp. 193-205.
- Solomon R.W. (1949), *An Extension of Control Group Design*, «Psychological Bulletin», 46, pp. 137-150.
- Stieber J.W. (1949), *Ten Years of the Minnesota Labor Relations Act*, Industrial Relation Center, Minneapolis.
- Stouffer S.A. (1949), "The Point System for Redeployment and Discharge", in Stouffer S.A. et al., *The American Soldier*, vol. II, *Combat and its Aftermath*, Princeton University Press, Princeton.
- Suchman E.A. (1967), *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*, Russell Sage, New York.
- Sween J., Campbell D.T. (1965), *A Study of the Effect of Proximally Auto-Correlation Error on Tests of Significance for the Interrupted Time-Series Quasi-Experimental Design*, mimeo.
- Thistlethwaite D.L., Campbell D.T. (1960), *Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment*, «Journal of Educational Psychology», 51, pp. 309-317.
- Walker H.M., Lev J. (1953), *Statistical Inference*, Holt, New York.
- Webb E.J., Campbell D.T., Schwartz R.D., Sechrest L.B. (1966), *Unobtrusive Measures: Nonreactive Research in the Social Sciences*, Rand McNally, Chicago.
- Wolf E., Lüke G., Hax H. (1959), *Scheidung und Scheidungsrecht: Grundfragen der Ehescheidung in Deutschland*, J.C.B. Mohr, Tübingen.

Logica della valutazione e pratica della valutazione

di Michael Scriven*

Questo capitolo definisce alcuni termini usati spesso quando si valutazione, e identifica una serie di problemi cruciali che riguardano della valutazione e ciò che la compone. In questa sede, mi occupo di questi problemi: il primo è quello, fondamentale, di come possono logicamente possibile una valutazione. Il secondo è quello delle razioni: come una valutazione possa fornire supporto alle raccomandazioni che vanno distinte da mere conclusioni valutative sull'oggetto dell'azione. Qui vengono chiarite alcune implicazioni di ciascuno di questi problemi ai fini della pratica valutativa.

Concetti di base

Uso qui l'espressione *logica della valutazione* per riferirmi ai specifici delle argomentazioni che, in tutti gli ambiti della valutazione, sottostanti ai processi di inferenza. La logica generale dell'inferenza, induttiva e statistica, non costituisce parte della logica della valutazione; non si tratta di una logica specifica della valutazione, applicazioni particolari di questi principi generali possono rivelarsi prioritari alla pratica della valutazione e ricadere nella sua logica.

Per comodità, faccio alcune distinzioni che esulano dall'uso della terminologia. Gli ambiti della valutazione, nei quali le assertive argomentative sono i principali risultati, si possono distinguere all'argomento della valutazione. Fra gli ambiti più comuni trovarsi la valutazione di programmi, di proposte, di personale, di prodotti, di procedure di politiche; oltre a questi vi è anche una decina di altri ambiti per la valutazione, che vanno dalla valutazione immobiliare all'eti-

* University of California, Berkeley. Fonte: "The Logic of Evaluation and Evaluation" in Fournier D.M., a cura di, *Reasoning in Evaluation: Inferential Links*, «New Developments in Evaluation», n. 68, Jossey-Bass, San Francisco, 1995. Tr.

cata, alla critica letteraria, alla metavalutazione (la valutazione di valutazioni). Gli ambiti vanno distinti dai settori: i settori si distinguono per la materia non valutativa che li caratterizza, la disciplina applicativa preesistente, a cui viene successivamente applicata la valutazione. Esempi tipici di settori sono l'istruzione, i servizi sanitari e le prove di automobili su strada. Alcuni settori possono richiedere l'applicazione di molti ambiti della valutazione: l'istruzione prevede la valutazione del personale e dei programmi (e parecchi altri ambiti), mentre le prove su strada riguardano solo la valutazione di prodotto. Le asserzioni valutative sono quelle che legano un predicato valutativo a un argomento del tipo di quelli elencati nella lista degli ambiti, comprese affermazioni che attribuiscono un valore ad una certa proprietà oppure che negano che un dato programma sia ancora considerato eccellente. Un predicato valutativo è un predicato estratto dal lessico usato per attribuire un valore, per stabilire una graduatoria qualitativa, per dare un voto e per collocare in una graduatoria qualitativa - oppure un predicato che non è definibile senza ricorrere a quel tipo di lessico. Infine, i tipi di valutazione sono gli approcci usati dal singolo valutatore o da diversi valutatori: sperimentale, di tipo antagonistico, orientata al raggiungimento degli obiettivi, indipendente dagli obiettivi, naturalistica, eccetera. Il tipo di valutazione è strettamente legato al metodo ma comprende approcci che sono più ideologici e definiti in maniera meno chiara di quanto normalmente avviene per gli approcci metodologici, come ad esempio la valutazione di empowerment.

È utile compiere, come avviene nella logica deduttiva, una distinzione approssimativa fra due livelli di problema nella logica della valutazione. Il primo livello, il livello delle fondamenta, comprende la definizione, la chiarificazione e la messa in relazione, attraverso definizioni e assiomi, dei concetti chiave. Mettere in chiaro gli aspetti fondamentali significa mettere in relazione questi concetti con altri concetti presenti in discipline simili. Il secondo livello comprende lo sviluppo di quelli che si potrebbero chiamare i principi di lavoro, approssimativamente l'equivalente di costruzioni e teoremi utili. Nel lavoro del topografo, ad esempio, questo tipo di discorso potrebbe riguardare la costruzione geometrica per trovare il punto medio di un segmento e il teorema di Pitagora, che sono procedure e principi operativi chiave nel suo lavoro. Questo è il livello applicativo, ma fa sempre parte della logica della valutazione, o della metodologia logica. Più oltre, c'è la pratica della valutazione e tutto il sapere operativo che le si costruisce attorno, vale a dire il campo della metodologia pratica e non logica. A questo livello, che corrisponde al sapere pratico che il topografo acquisisce con la pratica nell'uso dei suoi strumenti, possiamo inserire le generalizzazioni che si fanno riguardo alla giusta combinazione di approcci quantitativi e qualitativi per affrontare determinati tipi di problemi: i vantaggi e i limiti di una valutazione indipendente dagli obiettivi, il come fare analisi dei costi

che si possano rivelare utili, la relazione che intercorre fra credibilità e probabilità delle conclusioni, e così via¹. La differenza fra la logica applicata della valutazione e la metodologia della valutazione è, essenzialmente, la differenza tra le conclusioni che si possono trarre a partire da motivazioni logiche e le conclusioni basate sull'esperienza o sull'evidenza empirica. Queste ultime possono variare tra i vari campi della valutazione, mentre le prime devono essere applicabili a tutti i campi. La teoria della valutazione, al contrario della logica della valutazione, può avere a che fare con problemi sia riguardanti la logica che la pratica della valutazione. Nel primo caso, produrrà teorie normative, nel secondo le teorie potranno essere normative o descrittive a seconda dell'argomento trattato.

Nella valutazione, tra i concetti fondanti che richiedono una definizione c'è la valutazione stessa, definita qui come una ricerca sistematica sul merito intrinseco (*merit*), sul merito estrinseco (*worth*) e sull'importanza di un determinato oggetto (*significance*); la definizione di questi termini; e l'identificazione e la definizione di quattro distinte operazioni di base: attribuire un valore, stabilire una graduatoria qualitativa, dare un voto e collocare in una graduatoria qualitativa. Questi concetti si devono distinguere da concetti a loro imparentati e da concetti a loro opposti. La valutazione si deve distinguere da sinonimi quali da una parte la stima (*assessment*) (e una ventina di altri meno simili) e da concetti sostanzialmente diversi come il monitoraggio, l'analisi dei bisogni, la misurazione, la spiegazione e la descrizione. Il pregiudizio si deve distinguere dalla preferenza, la valutazione formativa dalla valutazione rieplögativa e così via. Ci sono state discussioni estese su questi argomenti, quindi non verranno affrontate in questa sede (Scriven 1991, 1995). Invece, mi concentrerò su una selezione dei più importanti problemi logici che la valutazione si trova ad affrontare, problemi sulle costruzioni logiche che essa richiede, e sui principi logici che dovrebbero governarla. Questi problemi sono tanto seri per la valutazione, e l'attenzione che gli è stata data finora così ridotta, che la disciplina corre il serio rischio di essere costruita sul nulla, o almeno di non riuscire a muoversi in quelle direzioni che potrebbero darne sviluppo e influenza³. Si spera

¹ Una lista di principi più lunga è proposta in Scriven (1993). Un buon esempio di un problema poco discusso in quest'area è la trasparenza, cioè il problema di quando, e fino a che punto, i valutatori esterni debbano mostrare le bozze o i rapporti di valutazione finale ai soggetti oggetto di valutazione. Cf. Scriven e Kramer (1995).

² C'è una vaga relazione fra questi concetti e i concetti fondamentali della logica della misurazione: le scale nominali, ordinali, di intervalli o di rapporti. Comunque, solo due di questi concetti sono strettamente confrontabili: stabilire una graduatoria qualitativa fornisce una scala ordinale sul merito intrinseco, il merito estrinseco e così via, e il dare un voto teoricamente, anche se raramente, fornisce una scala di rapporti.

³ Tre degli esempi forniti qui sono problemi di costruzione, e i problemi di costruzione a volte non hanno una soluzione. Un esempio, preso dalla geometria, è il problema della quadratura del cerchio, sul quale si è dovuto spendere molto tempo prima che ne fosse dimostrata

la differenza fra l'uso di un indicatore come un metodo euristico di ricerca (accettabile) e il suo uso come criterio di merito (inaccettabile, in linea generale). Certo, la variazione intersoggettiva nella valutazione del personale è molto alta. A volte, anche se non abbastanza spesso, riconosciamo la non validità di indicatori che sono anche molto correlati fra di loro, ad esempio nei *referee* anonimi di articoli di riviste, nella frase disprezzativa "dimmi con chi vai e ti dirò chi sei", e quando si impedisce alle giurie di avere accesso ai precedenti penali degli imputati. Ho discusso altrove questo argomento in maniera estensiva (Scriven 1987, 1990), quindi non ripeterò le mie tesi in questa sede.

Deborah Fournier ha sollevato un'interessante questione, se cioè questa posizione riguardo agli indicatori metta in pericolo la valutazione basata sulla teoria, che sembrerebbe usare teorie riguardo al modo di operare dell'evaluando per identificare indicatori a scopo valutativo. Sembra possibile che in questo modo si arrivi ad indicatori che potrebbero essere utilizzati solo come criteri di merito in assenza di criteri primari, ma la questione ha bisogno di studi ulteriori.

Altri problemi

A questa lista di problemi principali se ne potrebbero aggiungere altrettanti che meriterebbero di essere presi in seria considerazione, che elenchiamo per stimolare, da parte dei lettori, contributi per la loro risoluzione e ulteriori suggerimenti: il problema della spiegazione, che richiede di specificare la relazione fra valutazione e spiegazione, che sarebbe la questione per la quale la valutazione basata sulla teoria rimanda ad una risposta; il problema dei bisogni, che ne richiede la definizione e l'analisi; e il problema del parametro che riguarda l'identificazione dei parametri che definiscono un particolare problema di valutazione, fino al punto in cui si possa suggerire un disegno di valutazione appropriato (ne esistono almeno una decina⁸). A questi dovremmo aggiungere il problema 2B, il problema della sintesi dei valori, che è importante in certe valutazioni di policy.

Il problema delle raccomandazioni

Dei due problemi da affrontare in questa sede, sembra preferibile iniziare dal problema delle raccomandazioni, dato che ha conseguenze pratiche e immediate per tutti i valutatori e può essere affrontato senza dover sconfi-

nare nel vocabolario tecnico della logica. Lo sforzo è quello di stabilirne le credenziali in quanto problema e di fornire una bozza di trattamento.

Per comprendere il problema delle raccomandazioni e la sua importanza dobbiamo renderci conto di quanto questo problema sia il punto centrale di un tipo di disaccordo fondamentale - anche se spesso non riconosciuto - sulla natura della valutazione, specialmente per quanto riguarda la valutazione di programmi. Coloro che effettuano quest'ultimo tipo di valutazioni si dividono in tre gruppi, in base all'opinione riguardo il momento in cui termina il loro compito: li chiameremo *minimalisti*, *puristi* e *attivisti*.

I *minimalisti* pensano che la valutazione consista nel determinare quali siano i fatti rilevanti per orientare le scelte dei decisori, ad esempio determinare che cosa abbia imparato un certo gruppo di bambini in seguito all'intervento oggetto della valutazione o nel dare un giudizio sull'intera esperienza (processo). Credono che sia meglio lasciare al committente il compito di dare un valore a questi fatti o a queste percezioni. Un problema sociologico che emerge da questo punto di vista è che in questo modo non si crea un nuovo territorio per la valutazione, dato che la determinazione dei fatti rientra nel campo tradizionale della scienza sociale applicata, e il "fornire un punto di vista dall'interno" è una visione più recente (minoritaria) di quello che dovrebbe essere lo scopo vero e proprio delle scienze sociali. Un problema pragmatico è che, in una valutazione generale, il committente non si trova in una posizione favorevole per fare un asssemblaggio oggettivo di un set molto complesso di dati - che si riferiscono al processo e agli esiti - riunendoli in una valutazione complessiva.

Il punto di vista *minimalista*, del quale Marvin Alkin rappresenta un sostenitore molto in vista, è molto influente nel settore della valutazione di programmi, come si può vedere dal fatto che almeno due riviste con la parola *valutazione* nel titolo non pubblicano quasi altro che rapporti di ricerche empiriche incentrate sugli esiti. Come abbiamo visto, la nozione di valutatori che non giungono mai a conclusioni valutative non solo è paradossale, ma rende anche ridondante lo stesso concetto di valutatore come figura professionale autonoma. In realtà, non sono altro che scienziati sociali con un'altra etichetta, cosa che anche molti scienziati sociali ritengono giu-

sta. Il paradosso - che equivale a una vera e propria assurdità - non può essere evitato, e quindi i *minimalisti* non dovrebbero definire se stessi come valutatori; nonostante svolgano un lavoro preparatorio di grande valore per le valutazioni, non fanno valutazione. Questa ridondanza è sospetta, perché i valutatori possono esistere ed avere un ruolo utile e specifico in molte altre aree, come la valutazione dei prodotti o del personale. E i valutatori di processo, che esaminano i programmi in termini di equità o di uso legale dei fondi (ad esempio), lavorano nelle scienze sociali senza dover dipendere particolarmente dal filone tradizionale delle scienze sociali. Pensare che il

⁸ Un primo tentativo si trova in Scriven (1989), riassunto nella voce "Evaluation Parameters" dell'*Evaluation Thesaurus*.

ruolo del valutatore, definito come qualcosa di distinto dalla raccolta dei dati, sia impossibile o inappropriato nella valutazione dei programmi significa solo che c'è poca chiarezza sui concetti.

I puristi hanno un'idea della valutazione che va ben oltre quella dei minimalisti, ma non fino al punto cui vogliono arrivare gli attivisti. Credono che la valutazione non abbia termine fino al momento in cui è stato determinato il merito intrinseco, il merito estrinseco o il significato di un evahando. Determinare il merito, chiaramente, richiede più che la mera determinazione degli esiti o l'esame del processo; richiede di pesare l'importanza di ogni dimensione del merito e poi di combinare questi risultati con tutti i valori pertinenti specialmente i dati necessari all'analisi dei bisogni e gli standard in uso, siano essi etici, legali o scientifici. Il purista pensa che la conclusione di una valutazione, a seconda dei bisogni del committente, sia un'affermazione sul merito intrinseco - in senso assoluto o relativo - sul merito estrinseco o sul significato. In questo modo si rivendica un territorio che non appartiene alla scienza sociale empirica, un territorio nel quale bisogna stabilire credenziali particolari, in quanto, chiaramente, la pratica legittima della valutazione intesa in questo senso viene fatta risalire a molte migliaia di anni fa. Oltre questo punto, per quanto riguarda possibili sconfinamenti nel campo delle raccomandazioni o delle istruzioni, il valutatore non vede uno spazio per se stesso, a meno che non sia anche un decisore o uno specialista nel trovare soluzioni.

Gli attivisti ritengono che i puristi si arrendano troppo presto, e che facciano lo stesso errore di autolimitazione impropria che i puristi rimproverano ai minimalisti. Gli attivisti credono che la funzione principale - o almeno una funzione molto importante - delle valutazioni sia quella di generare raccomandazioni.

Il ragionamento è che la prima migliore approssimazione sia la posizione del purista, ma con un paio di concessioni significative nella direzione degli attivisti, senza renderla accettabile come prospettiva principale.

Problemi logici riguardo alla posizione degli attivisti

Il problema fondamentale della posizione degli attivisti è che si tratta di una posizione debole dal punto di vista logico. Dalle conclusioni sul merito intrinseco o sul merito estrinseco di un programma è generalmente impossibile trarre ulteriori conclusioni riguardo quello che il decisore dovrebbe fare. Supponiamo che il decisore sia un'agenzia finanziatrice che ha chiesto al valutatore di esaminare il programma X, che deve essere rifinanziato. L'agenzia chiarisce di sentirsi ancora impegnata nei confronti della popolazione beneficiaria di X. Sia il purista che l'attivista concordano che è quantomeno possibile concludere quanto segue:

C1. Tenendo conto delle risorse finanziarie dell'agenzia, il programma X rappresenta il modo più etico e più efficiente di soddisfare i bisogni del gruppo bersaglio cui si riferisce.

Quali raccomandazioni ne seguono? Probabilmente la seguente:

R1. Continuare a finanziare X

Ma è chiaro che non è questa la conseguenza. Molte delle premesse su cui si basa questa inferenza mancano. Il gruppo bersaglio ha altri bisogni oltre quelli cui si riferisce X. L'agenzia può credere che sia il momento di occuparsi di questi ultimi. Può essersene andato via il responsabile del programma, una persona molto competente in materia, che giocava un ruolo essenziale per ogni supporto ulteriore da dare a X. I fondi possono essere diminuiti da quando la valutazione è stata commissionata. E così via. Il valutatore sa poco o niente di tutte queste circostanze, e non ha compiuto nessuna determinazione sistematica di importanza al riguardo. Proporre R1 come esito di una valutazione è pretestuoso e pressapochista. L'esito della valutazione è C1, e rappresenta un input utile per i decisori, che si trovano ora nella posizione di poter trarre una conclusione sull'azione più appropriata da intraprendere, potendosi basare su molte considerazioni, fra le quali ci sono i risultati della valutazione.

Ecco un esempio più estremo (non che il primo esempio sia qualcosa che si incontra spesso). Supponiamo che le decisioni sul merito intrinseco e sul merito estrinseco vadano in direzioni opposte:

C2. Il programma X sembra essere molto inferiore a parecchie alternative, in realtà sembra essere quasi completamente inefficace.

Possiamo sicuramente dedurre questa raccomandazione:

R2. Non fornire a X ulteriori fondi.

No. L'inferenza contenuta in R2 non è, in linea generale, valida. Ci sono molti fattori di contesto che potrebbero rendere appropriato, e forse imperativo, il prosieguo di X. I dati sul programma non bastano a prevalere su questi fattori, e quindi dalla valutazione del programma non si deduce che non si deve rifinanziare il programma. Simili fattori comprendono i casi nei quali le alternative a X richiederebbero installazioni costose, e un certo tempo da usare per la formazione, per i quali non ci sono fondi; i casi nei quali c'è un forte supporto politico esterno per X, in modo che l'agenzia debba scendere a un compromesso e rifinanziare X per fare il miglior lavoro possibile con il proprio portafoglio di programmi, o forse anche per sopravvivere o