

ALBERTO MARRADI

Misurazione e scale: qualche riflessione e una proposta *

- Quanto ci vuole per arrivare a Messene? - chiese il viandante fermandosi presso un vecchio che sonnechiava al ciglio della strada.
- Cammina, cammina, - rispose il vecchio.
- Paziente, il viandante ripeté la domanda.
- E cammina! - si spazientí il vecchio.
- Sconcertato, il viandante riprese la sua strada.
- Ti ci vorranno due ore, - lo raggiunse la voce del vecchio dopo una ventina di passi.
- E perché non me lo hai detto subito?
- Come potevo dirtelo prima di vedere a che velocità cammini?

Da Esopo.

1. Il presente stato di confusione terminologica e concettuale.

1.1. Le riflessioni cui accenna il titolo di questo saggio prendono le mosse da una constatazione: da qualche decennio nelle scienze sociali il termine 'misurazione' viene usato per denotare processi estremamente diversi fra loro: questo progressivo stiramento del termine per coprire nuovi (o vecchi) concetti è ricostruito nelle sezioni 2, 3 e 4. Il fenomeno riguarda le sole scienze sociali, perché il termine non ha subito un analogo processo nel linguaggio delle scienze fisiche (dove non si parla di 'misurazione ordinale' e meno che mai di 'misurazione nominale'), e neppure nel linguaggio comune: le due espressioni appena citate suscitano infatti la meraviglia dei non-specialisti, e disorientano i discenti. Al punto che il caso del termine 'misurazione' sembra costituire - sia detto per inciso - una vistosa smentita alle tesi dei neopositivisti sulla precisione e univocità

* Per i preziosi consigli avuti in varie fasi della sua redazione, l'autore ringrazia i professori Alberto Baldissera (Università di Torino) e Nico Stehr (Università di Alberta) e i dottori Roberto Petrognani (Università di Firenze) e Maria Concetta Pitone (Università di Catania).

del linguaggio scientifico contrapposte alla vaghezza ed equivocità del linguaggio comune; tesi correntemente ripetute nelle scienze sociali malgrado le proteste degli studiosi di orientamento fenomenologico o affine.

Nella sezione 4.2 ci si chiede perché significati del termine 'misurazione' sempre più ampi e lontani da quelli originari sono stati proposti e concordemente adottati da psicologi, sociologi e politologi. E si trova una possibile risposta nell'aspirazione allo status «scientifico», che - per l'acritica identificazione della misurazione con la scienza - ha indotto a celare con un artificio terminologico alcuni fatti sgradevoli (per chi appunto aspirava a conseguire status «scientifico» attraverso la mimesi con le scienze già affermate):

- a) moltissime proprietà importanti per le scienze sociali presentano stati discreti anziché variare in modo continuo;
- b) per molte di queste proprietà, gli stati non sono neppure ordinabili secondo un criterio ragionevole;
- c) anche quando è ipotizzabile che una proprietà vari in maniera continua, le scienze sociali dispongono di vere e proprie unità di misura solo quando la misurazione non richiede alcuna cooperazione attiva da parte dei soggetti il cui stato è misurato; in tal caso, le unità di misura sono derivate dalle scienze fisiche (vedi più diffusamente le sezioni 7.2 e 7.4);
- d) se invece è necessaria la cooperazione attiva dei soggetti, allora anche le tecniche più sofisticate (discusse nella sezione 8) non pervengono a stabilire delle vere unità di misura, e quindi non possono essere definite tecniche di misurazione nel senso in cui quel termine è inteso nelle scienze fisiche (e nel linguaggio comune).

Un altro possibile motivo dell'uso del termine 'misurazione' per denotare ogni procedimento di assegnazione di valori numerici a stati sulle proprietà è - come si accenna nella sezione 9.1 - il mancato accordo su un termine generale che denoti l'insieme di tali procedimenti. Il vuoto terminologico viene riempito promuovendo a etichetta dell'intero *genus* quella che era originariamente l'etichetta di una sola *species* (la misurazione). Poiché la stessa etichetta è ormai usata per tutte le altre *species*

(la classificazione, l'assegnazione a categorie ordinate, il conteggio, ecc.) la confusione terminologica è giunta al colmo.

1.2. Ad essa ha dato e sta dando un notevole contributo la classificazione dei tipi di *scale* (cioè dei prodotti finali dei procedimenti di assegnazione dei valori) proposta da Stevens negli anni '40 e prontamente adottata dai metodologi e dai ricercatori empirici nelle scienze sociali. Stevens ha operato la distinzione fra le categorie basandosi esclusivamente sulle proprietà logico-formali dei vari tipi di scala; ma la scelta di questo criterio (anche se fosse stato applicato in modo ineccepibile, il che non è affatto: vedi sezioni 6.1 e 8.1) è in sé inopportuna e fuorviante, in quanto induce a non tener conto delle concrete operazioni di assegnazione dei valori: non si può dare per scontato che differenze significative in queste operazioni coincidano con differenze significative nelle proprietà logico-formali della scala.

Infatti, come si mostra nella sezione 3, l'etichetta 'ordinali' viene attribuita a scale prodotte con procedimenti diversissimi (e che presentano, fra l'altro, anche proprietà logico-formali del tutto differenti). Inoltre, alcune di queste scale 'ordinali' vengono promosse a 'scale a intervalli' da molti studiosi, che vogliono fruire della legittimazione, conferita dall'etichetta 'a intervalli', a usare tecniche metriche sulle corrispondenti variabili (vedi sezione 8.2). La confusione si manifesta quindi attraverso entrambe le sue possibili incarnazioni: cose profondamente diverse chiamate allo stesso modo, e la stessa cosa chiamata in modi significativamente diversi.

Naturalmente, una confusione di tale portata non può restare confinata al piano terminologico, ma è irrimediabilmente destinata a investire il piano concettuale. In altre parole, il problema non è soltanto una certa dose di incommunicabilità fra ricercatori che compiono operazioni corrette chiamandole in modi diversi (e magari logicamente scorretti); il problema è anche, ancor più scariamente, di ricercatori che – ingannati da etichette falsamente legittimanti – compiono operazioni concettualmente assurde come l'impiego di tecniche metriche su variabili classificatorie. Un esempio (tratto da una diffusa opera di un noto «metodologo», e tutt'altro che isolato nella letteratura) è riportato nella sezione 4.2. E non si può assolvere Stevens – e tutti i metodologi che hanno ripreso e diffuso la sua

terminologia – dalla responsabilità per simili interpretazioni aberranti: l'espressione 'scala nominale' contiene infatti (sia per la mera veste verbale, sia per il conato di mimesi con le scienze fisiche che rozzamente esprime) la legittimazione di tali aberrazioni, e ad esse inesorabilmente doveva condurre – come infatti ha condotto.

1.3. La proposta, che il titolo del saggio annuncia, emerge con naturalezza dalla critica della situazione esistente, ed è una proposta di risistemazione concettuale e terminologica dell'intero arco dei procedimenti con cui vengono assegnati valori simbolici a stati sulle proprietà; essa ha come conseguenza una risistemazione dei tipi di 'scale', cioè dei risultati finali di quei procedimenti.

Volutamente ho parlato di ri-sistemazione, anziché di nuova sistemazione, perché il campo era concettualmente e terminologicamente abbastanza in ordine prima che si cominciasse a stirare il termine 'misurazione', e in tale stato è rimasto nelle scienze fisiche e nel linguaggio comune, in cui si parla di classificare, contare, ordinare, e non solo, ossessivamente, di misurare.

Proprio dal linguaggio comune ho attinto, fin dove era possibile, i termini delle etichette che propongo: conteggio, assegnazione, categorie, ordinate. Piuttosto che coniare neologismi, propongo di rimettere in onore termini classici, come classificazione, di adottare espressioni proposte da altri metodologi, come 'scala assoluta', di definire con più rigore termini già in uso, come 'metrica', e infine di riutilizzare una delle etichette proposte da Stevens (scala a intervalli), ma ridefinendola integralmente al fine di renderla adeguata a coprire una famiglia di tecniche, elaborate in seno alle scienze sociali a partire da Thurstone, che assegnano le fattispecie a intervalli la cui ampiezza si suppone – con ragionevoli fondamenti – nota. Tecnica che invece – a parte le possibili intenzioni di Stevens, e il successivo uso generalizzato – restavano a rigore del tutto fuori dall'etichetta 'a intervalli' come l'aveva definita quell'autore, dato che nessuna di loro soddisfaceva il requisito della presenza di un'unità di misura.

Dopo questa esposizione sinottica degli intenti del saggio, passo ora alla riflessione sui singoli temi, iniziando – come mi

sembra necessario – da un'analisi storica del concetto di misurazione.

2. Due concetti di misurazione: originario e derivato.

2.1. Nelle principali lingue europee, i termini che designano il concetto di misura, e derivati, risalgono tutti al termine latino *mensura*, che a sua volta deriva da una radice indoeuropea M-vocale-N, che designa la luna (*mēnē* in greco, *mond* in tedesco, *maan* in olandese, *miesiasc* in russo, e così via). Cosa c'è in comune fra la luna e l'atto di misurare, o il suo risultato? Il fatto che la luna piena, tornando regolarmente a mostrarsi ogni 28 notti, offrì ai nostri lontani antenati un meraviglioso ancoraggio per sondare col pensiero l'abisso angoscioso del tempo. Attorno al sicuro e immutabile ripetersi delle fasi lunari l'uomo imparò ad organizzare tutta la sua vita sociale ed economica, la semina il raccolto l'accoppiamento il rito la guerra. La stessa radice fu pertanto usata dai popoli indoeuropei per designare la luna e il segmento di tempo intermedio fra due lune piene: il mese (*monadh* in inglese antico, *mānōd* in tedesco antico, *mēn* in greco, *miesiasc* in russo, *mensis* in latino, ecc.); solo in tempi assai tardi, e non in tutte le culture, il mese fu svincolato dal ciclo lunare e adattato al ciclo solare. La stessa radice fu anche usata per designare il processo di suddivisione di un continuum mediante un'appropriata unità convenzionale (*metrēin* in greco, *metiri* in latino, *messen* in tedesco, ecc.). Cosìché, gli scienziati che si riunirono verso la fine del XVIII secolo per stabilire una unità appropriata per misurare un altro fondamentale continuum – lo spazio – fecero ancora ricorso alla stessa radice, e l'unità fu detta metro.

Certamente l'argomento etimologico non è conclusivo, e inoltre esso andrebbe esteso a ceppi linguistici non-indoeuropei. Tuttavia la comune radice dei termini fornisce un consistente indizio del fatto che il concetto originario di misurazione si riferiva a un processo con il quale alcuni segmenti di lunghezza sono nota e costante (la lunghezza dell'unità di misura) venivano individuati ed enucleati da un continuum indefinito, come il tempo.

Non ha particolare rilevanza stabilire se tale continuum fosse

effettivamente percepito come indefinito, o se questo o quel polo ne individuasse il principio e/o la fine in qualche evento mitologico. Chiaramente la distanza da quel mitico inizio non era oggetto di misurazione effettiva, e il mese lunare, come unità e strumento di misura, non veniva proiettato nel passato e nel futuro al di là di quello che i bisogni pratici – non quelli simbolici – richiedevano. Pertanto, l'idea di un inizio assoluto, di un 'punto zero', era estranea al concetto originario di misurazione – anzi, addirittura incompatibile con esso.

2.2. Non mi sento di avanzare congetture sul quando, il dove e il come l'uomo si rese conto che lo stesso strumento concettuale usato per ritagliare sezioni di lunghezza definita ed uguale nel flusso indefinito del tempo poteva anche essere usato con un altro scopo. La vita di un uomo durava un tempo definito; aveva un inizio e una fine. Se l'età era un equivalente finito del tempo indefinito, l'unità di misura del tempo poteva essere usata per misurare anche l'età di una persona: quante volte la luna piena aveva fatto la sua comparsa da quando quel bambino era nato¹?

Per questa via il concetto originario di misurazione ne generò un altro, sensibilmente differente, in quanto si riferiva al processo con cui si confrontava l'ammontare di una proprietà X posseduto da un oggetto A con l'ammontare della stessa proprietà posseduto dallo strumento-unità scelto convenzionalmente per misurare appunto la proprietà X. Se l'oggetto A risultava avere 20 volte di più, di quella proprietà, di quanto ne possedeva lo strumento-unità, era più semplice dire che aveva «20 unità». Quel bambino aveva 7 lune piene (mesi lunari), quel campo era (lungo) 100 piedi. Lo strumento era l'unità di una scala numerica, e la misurazione convertiva una certa proprietà in simboli numerici (multipli di quella unità).

¹ La mia distinzione fra tempo indefinito ed età è analoga a quella di Durkheim fra tempo totale e durata concreta (*Le forme elementari della vita religiosa*, Comunità, Milano 1963, pp. 481-84). Durkheim afferma che l'individuo non potrebbe concepire il tempo totale al di fuori della società, il che potrebbe implicare una priorità temporale del concetto di durata concreta. D'altra parte, io non intendo affermare che l'idea del tempo indefinito è anteriore; soltanto che l'idea di *misurare* il tempo indefinito, o meglio scandirlo, sembra anteriore, almeno per i popoli indoeuropei, dati i termini che essi scelgono per designare il concetto di misura e derivati.

2.3. Verosimilmente questo concetto grezzo si dovette rapidamente raffinare in presenza di due fatti empirici ricorrenti: quantità della proprietà X inferiori all'unità di misura, oppure non divisibili per essa senza un resto tangibile. I nostri lontani antenati dovettero pertanto ammettere dei numeri razionali, o frazioni, come possibile esito di un processo di misurazione. Più arduo dev'essere stato ammettere come esito dei numeri irrazionali, cioè aventi infinite cifre diverse da zero, e non esprimibili come frazioni di numeri interi; il concetto di numero irrazionale turbò profondamente anche matematici provetti come i pitagorici, quando furono costretti a formularlo per esprimere il rapporto fra il lato di un quadrato e la sua diagonale. D'altra parte, gli strumenti di misurazione non erano così sofisticati da porre praticamente il problema dei numeri irrazionali – e questo naturalmente è vero anche per gli attuali strumenti di misurazione: non è possibile spingere oltre un certo livello di precisione il confronto fra l'ammontare della proprietà X posseduto dall'oggetto A e l'ammontare posseduto dallo strumento-unità. Ma se per ipotesi tale confronto potesse essere perfettamente esatto, un numero irrazionale ne sarebbe l'esito normale, mentre un numero razionale o addirittura intero sarebbe un esito del tutto eccezionale.

L'ipotetico esito esatto del confronto, cioè della misurazione, subisce un processo di arrotondamento, nel senso che la serie, quasi sempre infinita, di cifre del numero reale viene ridotta a una serie finita la cui lunghezza dipende dal livello di precisione degli strumenti. Questo processo di arrotondamento è inevitabile, e in effetti indistinguibile dalla misurazione stessa. Inoltre, poiché il livello di precisione degli strumenti è spesso superiore a quello richiesto dai fini conoscitivi di chi effettua o dirige quella particolare misurazione, l'esito del confronto può subire un secondo processo di arrotondamento, questo evitabile e condotto secondo regole esplicite o implicite, che ridurrà ulteriormente la serie delle cifre rilevate alla lunghezza prevista per quella data variabile dalla matrice dei dati.

Naturalmente tutti questi processi (arrotondamento connesso alla misurazione, arrotondamento secondo regole, effettiva registrazione delle cifre) vanno soggetti ad errori, nel senso che un ipotetico esito 2.14150265 può venir letto come

3.1415, deliberatamente arrotondato a 3.141, e poi magari registrato come 3.147. Nell'epistemologia classica, tutti questi errori di varia natura vengono accomunati sotto l'etichetta 'errori di osservazione'.

2.4. L'idea di un punto zero non è estranea al concetto derivato di misurazione come lo era al concetto originario. Tuttavia, l'unità di misura viene *definita* prescindendo da un punto zero: il metro è la quarantamillesima parte del meridiano terrestre, partendo da un punto qualunque sul meridiano; il mese lunare è l'intervallo fra due consecutive comparse della luna piena, o della luna nuova, o comunque della luna nella stessa esatta fase. Un punto zero è necessario solo per *applicare* lo strumento-unità di misura a un oggetto. Il fatto che tale punto zero sia convenzionale o assoluto non ha alcuna influenza sul processo di misurazione, e neppure – come vedremo nella sezione 8.1 – sulle operazioni che le tecniche statistiche correnti compiono sui risultati delle misurazioni, cioè sui dati.

3. Un caso di *stiramento terminologico*: la «misurazione» ordinale.

3.1. I due concetti di misurazione esaminati fin qui differiscono quanto ad alcuni attributi della proprietà che viene misurata, ma fanno entrambi riferimento in modo essenziale a uno strumento-unità di misura. Se tale strumento manca, siamo in presenza di un concetto significativamente diverso, che è stato ricompreso sotto il termine 'misurazione' solo a pena di un serio stiramento di quel termine². Questa estensione non è un fatto recente, perché già nel suo classico trattato, il fisico Norman Campbell denominava 'misurazione ordinale' il processo con

² La locuzione «stiramento dei concetti» è stata introdotta da G. Sartori (*La politica comparata: premesse e problemi*, in «Rivista Italiana di Scienza Politica», vol. I, 1971, n. 1, pp. 7-66 a p. 20). Peraltro, dalla stessa impostazione di quell'autore deriva che i concetti non possono essere stirati orizzontalmente, ma solo aumentati o diminuiti di estensione facendoli salire o scendere lungo una scala di astrazione. Sono i *termini* ad essere stirati allorché sono usati per designare concetti troppo diversi fra loro, anche se collocati allo stesso livello di astrazione.

cui l'ammontare della proprietà X posseduto da un oggetto A è confrontato con l'ammontare della stessa proprietà posseduto da un oggetto B, e il risultato di tale confronto è un giudizio di maggiore/uguale/minore³.

'Misurazione ordinale' è in seguito divenuta un'etichetta che designa vari processi che hanno ben poco in comune fra loro, e altrettanto poco con il processo che più comunemente è inteso come misurazione nel linguaggio corrente, e che corrisponde al concetto qui definito 'derivato'.

Fra questi processi, i due più correntemente impiegati, e più correntemente denominati 'misurazione ordinale', possono essere distinti con chiarezza chiedendosi *perché* manca lo strumento-unità che renderebbe possibile una misurazione nel senso classico del termine. Le risposte – mi sembra – possono essere due, e nettamente diverse, a seconda della natura della proprietà che si considera.

3.2. Supponiamo che la proprietà sia il 'titolo di studio'. Ha senso dire che compariamo i titoli di studio del signor Rossi e del signor Bianchi, e troviamo che il primo ha un titolo più alto, uguale, o più basso del secondo. Ma non ha senso dire che ci manca un'unità di misura. Se infatti, per assurdo, ne avessimo una, potremmo dividere un titolo di studio per quell'unità, trovando un numero reale come quoziente. Tutto questo suona appunto assurdo; e dà questa impressione, a riflettere, perché la proprietà 'titolo di studio' ha solo pochi stati discreti, o categorie: possiamo concepirli come una serie ordinata, non come un continuum. E come Thurstone affermava nel 1931: «un con-

³ N. R. Campbell, *An Account of the Principles of Measurement and Calculation*, Longmans & Green, London 1928. La confusione fra i vari tipi di misurazione ordinale, criticata nel testo, è presente anche in un'opera classica e meditata come *Theory and Methods of Scaling* di Warren S. Torgerson (Wiley, New York 1958). A p. 16 egli afferma che «nella scala ordinale i numeri sono assegnati ai vari stati della proprietà, in modo che l'ordine dei numeri corrisponda all'ordine di grandezza degli stati», cioè si riferisce al processo descritto nella sezione 3.2 del testo. Ma a p. 29 egli definisce «il caso più semplice» di scala ordinale una situazione in cui «la proprietà e le operazioni definitive sono tali che ogni oggetto può essere comparato direttamente con ogni altro oggetto»: in questo caso Torgerson si riferisce al confronto a coppie, descritto nella sezione 3.3 del testo. Ma il confronto a coppie fra oggetti non è «il caso più semplice» di assegnazione di numeri a categorie ordinate, bensì qualcosa di radicalmente diverso.

tinuum lineare... è implicito in ogni misurazione»⁴. Possiamo pensare a un'unità di misura solo se la proprietà che consideriamo può assumere infiniti stati. Un numero finito di stati è incompatibile con un'unità di misura, e quindi con i due concetti di misurazione illustrati nella sezione 2.

Spingendo oltre l'analisi, si può aggiungere che, nel caso del signor Rossi e del signor Bianchi, non stiamo effettivamente comparando *i loro* titoli di studio, perché abbiamo già confrontato tali titoli come categorie astratte, senza far riferimento a individui concreti. Questa comparazione fra i titoli ci può portare a ordinarli tutti secondo un dato criterio (qualora ne troviamo uno che ci soddisfi), e a quel punto non si tratterà più di comparare direttamente i titoli di Rossi e di Bianchi, ma di assegnare ciascun titolo concreto alla corrispondente categoria astratta. Se le categorie sono state ordinate come tali, la comparazione fra i titoli di Rossi e Bianchi sarà già fatta automaticamente. Ma il fatto che le categorie sono ordinate è irrilevante al mero processo di assegnazione dei casi concreti, che sarebbe lo stesso anche se tale ordine non avesse senso, come nel caso di proprietà come 'nazionalità' o 'confessione religiosa'.

Concludendo: se la proprietà consiste in una serie di categorie ordinate, non solo non c'è misurazione, ma non c'è nemmeno una comparazione fra oggetti, come invece voleva la definizione di Campbell. Pertanto, l'etichetta 'misurazione ordinale' sembra del tutto inappropriata, e dovrebbe esser sostituita dall'etichetta 'assegnazione a categorie ordinate'.

3.3. Supponiamo ora che la proprietà sia la 'religiosità'. Ha senso dire che compariamo il grado di religiosità dei signori Bianchi e Rossi, e troviamo che il primo è più, meno, o altrettanto religioso del secondo. Ha senso anche dire che, se troviamo un'unità di misura adeguata, potremmo confrontarla con la religiosità del signor Bianchi, assegnando a quest'ultima come misura il quoziente di tale confronto. Tutto questo non suona assurdo, perché possiamo immaginare la religiosità come un continuum. Appare tuttavia utopico, perché le prospettive di trovare un'adeguata unità di misura della religiosità sono al mo-

⁴ L. L. Thurstone, *The Measurement of Social Attitudes*, in «Journal of Abnormal and Social Psychology», vol. XXVI, 1931, n. 2, pp. 249-69 a p. 259.

mento assai scarse. D'altra parte, la scienza nel suo sviluppo ha trovato unità di misura per proprietà, come il colore, che in precedenza non erano neppure percepite come continue; uno può quindi legittimamente pensare che la mancanza di una unità di misura adeguata è solo uno stato temporaneo.

Se questo ci autorizzi a parlare di 'misurazione ordinale' o solo, più cautamente, di 'confronto quantitativo' è un giudizio che dipende dalla nostra maggiore o minor propensione a stizzare il termine 'misurazione'. I positivisti logici, ad esempio, hanno sempre tenuto fermo il fatto che non si può avere misurazione senza unità di misura⁵. Un'ovvia conseguenza della mancanza di tale unità è che il risultato del confronto non può essere un numero: si possono usare, come infatti si usano, cifre come etichette per le categorie che costituiscono l'esito del confronto; ma al più con tali cifre si potrà costruire una rozza scala ordinale ($x = \text{minore}$, $z = \text{uguale}$, $3 = \text{maggiore}$).

Una seconda conseguenza della mancanza di un'unità di misura è che, in tutte le situazioni dubbie (cioè quando l'ammon-tare della proprietà X è approssimativamente lo stesso nei due oggetti A e B), il giudizio di maggiore/uguale/minore diventa arduo, incerto, e instabile, nel senso che può essere smentito da un altro giudice, o dallo stesso giudice in un altro momento.

Ma l'inconveniente più grave del confronto diretto fra coppie di oggetti si manifesta allorché si considera il fatto che nelle scienze sociali non possiamo certo limitarci a misurare le proprietà di due soli oggetti. Se i nostri oggetti sono 100, misurarne una proprietà mediante il confronto di ciascuno con un'unità di misura richiede 100 atti; ma se dobbiamo ricorrere al confronto fra coppie di oggetti, il totale dei confronti è $(100 \times 99)/2 = 4950$. In una normale ricerca con 2000 casi e 200 variabili, gli atti di misurazione sono 400 000, ma i confronti a coppie sarebbero 200 $(2000 \times 1999)/2 = 399\,800\,000$.

Un altro aspetto importante, che sembra trascurato dalla letteratura metodologica, è che la misurazione mediante confronto con un'unità e la «misurazione» mediante confronto a coppie producono dati organizzati in maniera diversa e del tutto in-

⁵ Vedi ad esempio R. Carnap, *Logical Foundations of Probability*, Chicago University Press, Chicago 1950, p. 9; C. G. Hempel, *Fundamentals of Concept Formation in Empirical Science*, Chicago University Press, Chicago 1952, p. 58.

compatibile. Se la proprietà da misurare è una sola, il confronto con un'unità produce un vettore di valori numerici, mentre il confronto a coppie produce una matrice quadrata, in cui gli oggetti sono teste di riga e teste di colonna, e le celle contengono simboli di maggiore/uguale/minore. Se le proprietà sono più di una, il primo metodo produce una matrice rettangolare in cui gli oggetti sono teste di riga e le proprietà sono teste di colonna (o viceversa); il secondo produce una serie di matrici quadrate – una per ogni proprietà – del tipo sopra descritto. Le tecniche statistiche adatte a trattare matrici rettangolari (analisi tabulari, correlazioni-regressioni, e così via) sono ovviamente inapplicabili a matrici quadrate prodotte attraverso confronti a coppie.

Tali matrici, infine, possono contenere dati incongruenti fra loro. Supponiamo che il signor Bianchi sia stato giudicato più religioso del signor Rossi, e quest'ultimo più religioso del signor Verdi. In teoria dovrebbe conseguire che il signor Bianchi è stato giudicato più religioso del signor Verdi. Ma in pratica può accadere di trovare che è stato dato un giudizio opposto (Verdi più religioso di Bianchi), e quindi incongruente con i primi due. La probabilità di tali giudizi intransitivi aumenta esponenzialmente all'aumentare degli oggetti da comparare, come ben sa chiunque ha lavorato con la tecnica dei confronti a coppie⁶. È ovvio che la presenza di dati incongruenti pone gravi problemi logici e tecnici in sede di analisi.

Concludendo: se la proprietà può essere concepita come un continuum, ma non disponiamo di un'adeguata unità di misura, il ricorso alla «misurazione ordinale», nel senso di confronto

⁶ Vedi ad esempio K. O. May, *Intransitivity, Utility, and the Aggregation of Preference Patterns*, in «Econometrica», vol. XXII, 1954, pp. 1-13; C. H. Coombs, *Inconsistency of Preference as a Measure of Psychological Distance*, in C. W. Churchman e P. Ratoosh (a cura di), *Measurement: Definitions and Theory*, Wiley, New York 1959; V. Capecchi, *L'analisi delle preferenze politiche*, in «Rassegna Italiana di Sociologia», vol. VI, 1965, pp. 199-264; A. A. Weinstein, *Individual Preference Intransitivity*, in «Southern Economic Journal», vol. XXXIV, 1968, pp. 335-43; A. Tversky, *Intransitivity of Preferences*, in «Psychological Review», vol. LXXVI, 1969, pp. 31-48. Per la frequenza di giudizi intransitivi, oltre che per altri fra i motivi menzionati nel testo, la tecnica del confronto a coppie è stata giudicata «del tutto inadatta alla teorizzazione scientifica quantificata»: vedi T. P. Wilson, *Critique of Ordinal Variables*, in «Social Forces», vol. XLIX, 1971, pp. 432-44. Vedi anche le osservazioni, in parte analoghe, di L. Perrone, *Metodi quantitativi della ricerca sociale*, Feltrinelli, Milano 1977, p. 142.

fra coppie di oggetti comporta tali e tanti inconvenienti da renderlo del tutto sconsigliabile nelle scienze sociali, a meno che il numero degli oggetti e delle proprietà da confrontare sia estremamente ridotto.

4. Un caso di contraddizione in termini: la «misurazione» nominale.

4.1. A mia conoscenza, il primo studioso che abbia affiancato il concetto di misurazione al concetto di nome è stato lo psicometrico Stanley S. Stevens, che in un breve articolo del 1946 propose la «scala nominale» come il livello più basso di misurazione. Pochi anni dopo, Stevens coerentemente definì la misurazione come «l'assegnazione di etichette numeriche (*numerals*) a oggetti o eventi secondo una regola – qualsiasi regola». L'idea di una misurazione nominale attraverso scale nominali fu prontamente accettata nelle scienze sociali, e portata a tutte le sue logiche conseguenze. Nel 1953 Coombs scriveva: «La misurazione nella sua forma più semplice consiste nel sostituire oggetti reali con nomi o simboli... questo livello di misurazione è così primitivo che non è sempre riconosciuto come tale». Ellis parlò di «misurazione di identità e differenze». Galtung presentò addirittura il requisito classico della classificazione, la mutua esclusività delle categorie, come la sola caratteristica della misurazione: «Data una variabile, la misurazione è stabilire una corrispondenza fra i valori della variabile e un insieme di numeri. Non si richiede null'altro: solo che ad ogni valore corrisponda uno e un solo numero»⁹.

La magica etichetta è stata estesa anche all'attività pratica di

⁹ S. S. Stevens, *Mathematics, Measurement, and Psychophysics*, in Stevens (a cura di), *Handbook of Experimental Psychology*, Wiley, New York 1951, pp. 1-49 a p. 22. L'articolo del 1946 è *On the Theory of Scales of Measurement*, in «Science», vol. CIII, 1946, pp. 677-80.

¹⁰ C. H. Coombs, *Theory and Methods of Social Measurement*, in L. Festinger e D. Katz (a cura di), *Research Methods in the Behavioral Sciences*, Dryden, New York 1953, pp. 471-535 a p. 473.

¹¹ B. Ellis, *Basic Concepts of Measurement*, Cambridge University Press, Cambridge 1968, p. 42.

¹² J. Galtung, *Theory and Methods of Social Research*, Universitetsforlaget, Oslo 1967, p. 73.

assegnare oggetti a delle classi: «Un linguaggio di misurazione definisce classi di fenomeni fornendo criteri specifici per decidere se un'osservazione può essere assegnata a una particolare classe». Abell, sviluppando quello che era implicito nelle posizioni di Coombs sopra citate, estese l'etichetta oltre la classificazione fino alla denominazione: «Tradizionalmente il livello nominale è stato preso come il livello più basso di misurazione. Ma se vogliamo essere perfettamente rigorosi dobbiamo definire un livello nominale semplice... che è il livello di misurazione meno strutturato: si limita semplicemente a denominare gli oggetti del mondo reale senza alcun tentativo di classificarli»¹¹.

È stupefacente constatare quanto scarse siano state le critiche rivolte a un così artificioso stiramento del termine «misurazione». Le sole due reazioni di cui sia a conoscenza vengono da Torgerson e Sartori. Il primo osserva che «Etichette numeriche possono essere usate per denominare le classi... e ciò accade comunemente. Tuttavia, il fatto che in una biblioteca si assegna il numero 8105 a un libro non vuol dire che il bibliotecario ha *misurato* il libro... Altrimenti, la classificazione, e persino la denominazione dei casi individuali, diventano una forma di misurazione»¹². E Sartori: «Una scala nominale è soltanto una classificazione – una classificazione qualitativa – e dunque non è per niente una scala che misuri qualcosa. S'intende che anche le voci di una classificazione possono essere numerate: ma questo è solo un espediente di codifica che non ha nulla a che fare con una quantificazione»¹³.

¹¹ A. Przeworski e H. Teune, *The Logic of Comparative Social Inquiry*, Wiley, New York 1970, p. 93.

¹² P. Abell, *Measurement in Sociology. I: Measurement Systems*, in «Sociology», vol. II, 1968, pp. 1-20 a p. 3.

¹³ Torgerson, *Theory and Methods* cit., pp. 9 e 14. Corsivo dell'autore.

¹⁴ Sartori, *La politica comparata* cit., p. 23. Durante il dibattito su questa relazione, il professor Ammassari ha osservato che un'estensione alla classificazione del termine 'scala' suona meno assurda ad orecchie anglosassoni che ad orecchie latine, dato che per gli anglosassoni il termine 'scale' vuol dire 'bilancia' (originariamente, 'piatto della bilancia'). C'è probabilmente qualcosa di vero in questa tesi, anche se tutti i dizionari inglesi riportano come correnti entrambi i significati (bilancia e scala nel nostro senso), insieme ad altri ancora. L'osservazione di Ammassari ha inoltrare il merito di richiamare l'attenzione su due fenomeni: a) il fatto che le scelte terminologiche non sono affatto irrilevanti agli sviluppi concettuali, ma contribuiscono spesso ad offuscare in modo assai grave i corrispondenti concetti; b) il fatto che la comunità dei ricercatori

A queste osservazioni si potrebbe aggiungere che l'espressione 'scala nominale' è una contraddizione in termini (tecnicamente, una *contradictio in adjectu*, in quanto l'aggettivo contraddice il nome). Se le differenze fra le voci di una lista sono nominali (cioè qualitative) ma non quantitative o di grado, quella lista non può essere chiamata scala, perché si ha una scala solo quando c'è una differenza fra i gradini. Sarebbe una ben buffa scala quella composta di gradini qualitativamente differenti, ma tutti collocati allo stesso livello! Sotto questo aspetto, una classificazione è ovviamente l'opposto di una scala.

4.2. Quali ragioni possono aver spinto gli scienziati sociali a tendere e stritare un termine fino al punto di usarlo sia per il suo referente originario sia per il suo opposto? Una risposta plausibile sembra debba esser cercata in quella forte, quasi morbosa attrazione verso un certo modello delle scienze fisiche, e conseguente repulsione per la propria eredità filosofica, che ha anche condotto la maggioranza degli scienziati sociali a dare status privilegiato ai compiti nomotetici rispetto a quelli idiografici, e alla spiegazione rispetto alla comprensione; a metter su penose imitazioni di esperimenti; ad analizzare i dati mediante semplici ipotesi di relazioni bivariate controllate con criteri dicotomici (vero/falso) anziché mediante la giustapposizione di più ricchi, e complementari, modelli multivariati di spiegazione, nella consapevolezza che qualsiasi modello è comunque mal specificato, cioè non può che costituire una lontanissima approssimazione alla complessa rete di influenze e interazioni che agiscono su ogni fenomeno studiato dalle scienze sociali.

Dato che la misurazione era identificata con Galilei e Newton, mentre la classificazione era identificata con Aristotele, lo stesso termine 'classificazione' è stato bandito, e i suoi significati sono stati distribuiti fra 'misurazione nominale' e 'scala empirici dell'Europa continentale importa i termini americani (e le connesse confusioni concettuali) senza operare alcun controllo filologico-semantic. Di entrambi i fenomeni si potrebbero naturalmente addurre moltissimi altri esempi, oltre a quello di 'scala nominale'.

¹⁹ Il termine 'classificazione' ha tre principali significati: a) L'operazione intellettuale con cui l'estensione di un concetto è divisa secondo un qualche criterio, e le divisioni (classi o categorie) ricevono etichette che possono essere concetti a più basso livello di astrazione di quello classificato. Questo significato viene attribuito al termine 'misurazione' da definizioni come quelle di Ellis e di Galtung. b) L'insieme delle classi create dall'operazione sub A. Questo è il

nominale'. Tre peccati intellettuali possono essere rilevati in questo atteggiamento e nei comportamenti conseguenti:

- a) dando alla misurazione e alla quantificazione il merito per il progresso delle scienze fisiche, l'analista non si è spinto oltre gli aspetti più superficiali e appariscenti, trascurando il fatto che «la formazione concettuale è preliminare rispetto alla quantificazione... Non possiamo misurare se non sappiamo prima cosa è che stiamo misurando... per questa buona ragione il progresso della quantificazione deve arrivare – in ogni disciplina – dopo il progresso concettuale»¹⁶;
- b) è stato dato per scontato che i problemi delle scienze sociali possono essere risolti con gli stessi strumenti concettuali e operativi dei problemi delle scienze fisiche, e ogni argomentazione circa la possibile specificità di tali problemi è stata scartata come non-scientifica;
- c) strاندolo al di là di ogni limite, il termine 'misurazione' è stato praticamente privato di significato. Se tutto, compreso la classificazione e la denominazione, è misurazione, ognuno si è naturalmente sentito autorizzato a usare il linguaggio quantitativo «non soltanto senza disporre di alcuna misurazione effettiva, ma senza averne alcuna in progetto e, peggio ancora, senza nessuna apparente cognizione di quel che occorre fare prima che una misurazione diventi possibile»¹⁷.

I risultati sono sotto gli occhi di tutti. A furia di parlare di scale nominali e di misurazione nominale, la gente è stata indotta a prendere sul serio, cioè come numeri veri e propri, quelle etichette numeriche che erano state assegnate come contrassegni alle categorie di una classificazione, e lì ha coscientemente sottoposti alle più potenti e «scientifiche» tecniche statistiche. Senza rendersi conto – l'indottrinamento all'ideologia numerologica della scienza può rapidamente tacitare la voce non-

significato corrente di 'scala nominale'. c) L'operazione con cui singoli oggetti vengono assegnati a una delle classi di una classificazione nel senso b. Questo significato viene attribuito a 'misurazione' da definizioni come quelle di Stevens, Coombs, e Przeworski e Teune.

¹⁶ G. Sartori, *Concept Misformation in Comparative Politics*, in «American Political Science Review», vol. LXIV, 1970, pp. 1033-53 a p. 1038.

¹⁷ A. Kaplan, *The Conduct of Inquiry*, Chandler, San Francisco 1964, p. 123.

scientifico del buon senso – del fatto che, visto che le categorie non erano ordinate, le etichette numeriche non potevano che essere assegnate in modo totalmente arbitrario; qualsiasi altra assegnazione di etichette sarebbe stata altrettanto lecita, ma avrebbe cambiato magari radicalmente i risultati della procedura statistica, la cui applicazione era pertanto priva di senso.

Al fine di mostrare come la produzione di *nonsense* numerologico non è affatto limitata ai novizi entusiasti, mi si permetta di tornare a citare un'equazione di regressione presentata in un notissimo libro del «metodologo» americano Rudolph J. Rummel:

Stile sistemico = $2,15 + .36$ (status costituzionale) – $.18$ (carattere rappresentativo) – $.13$ (sistema elettorale) – $.12$ (libertà per l'opposizione organizzata) – $1,83$ (regime non-comunista) + $.31$ (leadership politica) – $.22$ (distribuzione orizzontale del potere) – $.31$ (tipo monarchico) + $.02$ (partecipazione militare) + X (errore statistico)¹⁸.

Come osservavo altrove, «Rummel non si preoccupa di spiegare come ha 'misurato' queste variabili; per la maggior parte esse sembrano assolutamente non-misurabili»¹⁹. Rummel si limita ad annunciare orgogliosamente che R^2 è $.83$, cioè che la sua equazione «spiega» l' 83% della varianza della dipendente. Di fronte a questo abissale nonsenso, non vale neppure la pena di rilevare le altre grossolane sciocchezze manifestate in questa equazione, che sembra davvero un compendio delle assurdità cui conduce l'ideologia della quantificazione.

5. Accento sulla misurazione o accento sulle scale?

5.1. Nella sezione 4.1, le locuzioni 'scala nominale' e 'misurazione nominale' sono state discusse parallelamente, in quanto entrambe sono affette dalla stessa contraddizione in termini. Peraltro, i due termini, con i relativi concetti, dovrebbero essere tenuti accuratamente distinti, e non solo per ragioni logico-formali. Un discorso centrato sulla misurazione, infatti, dirige

¹⁸ R. J. Rummel, *Applied Factor Analysis*, Northwestern University Press, Evanston 1967, p. 203.

¹⁹ A. Marradi, *Concetti e metodi in scienza politica*, Giuntina, Firenze 1980, p. 44.

l'attenzione su una sequenza di operazioni intellettuali e pratiche, con le relative caratteristiche, condizioni e limiti. Un discorso centrato sulle scale, invece, dirige l'attenzione sulle proprietà logico-matematiche del prodotto finale di tali operazioni. Negli anni '20, due trattati di Norman Campbell²⁰ avevano richiamato l'attenzione sulle procedure, influenzando il lavoro di Thurstone e della sua scuola; ma alla fine degli anni '40, i già citati lavori di Stevens invertirono la tendenza, portando alla ribalta le scale e le loro proprietà matematiche.

Inversione gradita ai cultori delle esercitazioni formali: negli anni successivi proliferarono le proposte di nuove tipologie di scale, sempre più sofisticate e avulse dalla pratica della ricerca²¹. Il commento di Torgerson alla proposta di Coombs può essere esteso all'intera famiglia: «Coombs sembra occuparsi soprattutto di quelli che potremmo chiamare metodi per la classificazione sistematica di insiemi *limitati* di oggetti, piuttosto che di metodi per la misurazione di una proprietà»²².

5.2. Mentre le tipologie di scale proposte da Coombs e altri come elaborazioni della tipologia di Stevens non ebbero alcun seguito pratico, quest'ultima, proprio per la sua semplicità (4 tipi: nominali, ordinali, a intervalli, cardinali), si è diffusa fulmineamente dai testi di metodologia al metalinguaggio dei ricercatori nelle scienze sociali, condizionando totalmente le immagini che essi si facevano delle tecniche di misurazione che andavano imparando e applicando, e in una certa misura anche le tecniche stesse. Visto che l'unica cosa su cui si richiamava la sua attenzione erano le proprietà matematiche dei vari tipi di scale, il ricercatore si sentiva autorizzato a trascurare ogni riflessione critica sui procedimenti con cui venivano attribuiti i punteggi individuali su tali scale.

Tutto questo ha avuto conseguenze a mio avviso gravissime

²⁰ N. R. Campbell, *Physics: the Elements*, Cambridge University Press, Cambridge 1920; *An Account* cit.

²¹ C. H. Coombs, *A Theory of Psychological Scaling*, University of Michigan Press, Ann Arbor 1952; Coombs, *Theory and Methods* cit.; C. H. Coombs, H. Raiffa e R. M. Thrall, *Some Views on Mathematical Models and Measurement Theory*, in «Psychological Review», vol. LXI, 1954, pp. 132-44; Abell, *Measurement in Sociology* cit.

²² Torgerson, *Theory and Methods* cit., p. 18. Corsivo dell'autore.

sulla qualità della ricerca sociale, proprio nel trentennio della sua massima diffusione. Un brillante saggio di Deutscher porta prove analitiche di questo scadimento qualitativo²³, che può essere valutato anche confrontando i due giudizi che seguono. Scrivendo nel 1935, lo psicologo Gordon W. Allport affermava che «i successi raggiunti negli ultimi dieci anni in fatto di misurazione degli atteggiamenti possono essere considerati uno dei risultati più lusinghieri della psicologia sociale»²⁴. Una quarantina di anni dopo, lo statistico Hubert M. Blalock era costretto a osservare che «mentre l'importanza di misurare bene è riconosciuta largamente dagli scienziati sociali, scarssissima attenzione è stata dedicata a questo argomento nella letteratura del settore... pochi si sono sforzati di estrarre principi generali da casi specifici; non si sono sviluppate specializzazioni particolarmente orientate a problemi di misurazione». A suo avviso, era possibile individuare «un periodo di ristagno nell'interesse per la misurazione degli atteggiamenti... fra gli anni '50 e gli anni '60... Scale elaborate con grande cura venivano sostituite da rozzi strumenti a due o tre categorie, nella convinzione che essi funzionassero quasi altrettanto bene»²⁵.

Penso che i rilievi di Blalock potrebbero agevolmente essere estesi in varie direzioni: dagli atteggiamenti a gran parte delle variabili delle scienze sociali; dal numero di categorie a ogni altro aspetto del processo di raccolta e controllo di qualità dei dati; dal periodo a cavallo fra gli anni '50 e '60 a tutto il trentennio successivo al 1950. Una quota notevole della responsabilità di questa situazione è da ascrivere allo spostamento di attenzione dai concreti procedimenti alle proprietà formali, determinato da Stevens, e all'adozione della sua classificazione dei tipi di scala come la sistemazione ottimale e definitiva del problema.

Invece, se la categoria 'scala nominale' presenta addirittura

²³ I. Deutscher, *Looking Backward: Case Studies on the Progress of Methodology in Sociological Research*, in «American Sociologist», vol. IV, 1969, pp. 34-42.

²⁴ G. W. Allport, *Attitudes*, in C. Murchison (a cura di), *Handbook of Social Psychology*, Clark University Press, Worcester 1935, cap. 17.

²⁵ H. M. Blalock jr., *Introduction*, in Blalock (a cura di), *Measurement in the Social Sciences*, Aldine, Chicago 1974, p. 1.

una contraddizione in termini, anche le altre tre categorie proposte da Stevens non vanno affatto esenti da critiche sul piano logico, e la classificazione nel suo insieme ha un effetto fuorviante, in quanto i punti di divisione fra le categorie non sono quelli in cui le concrete operazioni di misura – e persino le proprietà formali, in due casi – cambiano in modo significativo. Procedure differenti che sono state inglobate nella stessa categoria dovrebbero essere separate; procedure simili che sono state attribuite a categorie diverse dovrebbero essere assegnate alla stessa: è quanto si cercherà di fare nelle sezioni che seguono.

6. Conteggio: la scala assoluta.

6.1. Secondo Stevens e l'uso corrente, una scala cardinale deve soddisfare due requisiti: avere un punto zero e un'unità di misura. Il punto zero dev'essere assoluto (cioè deve corrispondere a una situazione di assenza della proprietà che viene misurata), altrimenti il rapporto fra un punto e un altro della scala risulta distorto: il termometro non è una scala cardinale perché 20° non sono caldi il doppio di 10° (questo classico esempio è criticato più avanti, nella sezione 8.1). Di conseguenza, la sola trasformazione possibile dei valori lungo la scala è la loro moltiplicazione per una costante, operazione che lascia inalterati i rapporti.

La scelta di queste caratteristiche come requisiti della categoria 'scala cardinale' riflette le preoccupazioni di Stevens per le proprietà matematiche della scala (trasformazioni possibili, rapporto fra punti della scala) piuttosto che per le concrete procedure di misurazione e di analisi dei dati: nessuna procedura statistica richiede che due qualsiasi punti di una scala siano messi in rapporto, cioè divisi l'uno per l'altro.

Comunque, la categoria in questione è insoddisfacente anche dal punto di vista formale. Se infatti si ammettono moltiplicazioni per una costante, ciò significa che l'unità di misura è considerata convenzionale. Ma cosa accade se l'unità *non* è convenzionale? Come ha osservato Galtung, «se stiamo contando mucche, cosa sia 'una mucca' non è un fatto convenzionale. Questa scala la chiamiamo assoluta, perché non è possibile trasformarla

in alcun modo... se ci sono 32 studenti in una classe, il fatto è adeguatamente descritto dal numero 32»²⁶.

In altre parole, ogniquale volta una proprietà consiste nel possesso di, o nella relazione con, un certo numero di oggetti discreti, non abbiamo bisogno di misurare – né potremmo farlo. Non misuriamo il numero di figli di una coppia, o il numero di letti in un ospedale: li contiamo. Da qualsiasi punto lo si consideri, il conteggio è radicalmente differente dalla misurazione. Non ha luogo alcun confronto con uno strumento-unità, o con altri oggetti; non si deve procedere ad arrotondamenti. Il conteggio ha come risultato dei veri e propri numeri interi, la misurazione dei numeri reali arrotondati²⁷. Da un punto di vista epistemologico, non si devono stabilire unità convenzionali, non si deve alterare deliberatamente la realtà: il risultato del conteggio viene registrato direttamente nella matrice dei dati.

6.2. Se Stevens avesse classificato delle forme di misurazione, avrebbe quindi avuto ragione nel lasciare fuori il conteggio. Ma poiché classifica tipi di scale, il suo schema è chiaramente incompleto, perché il conteggio produce un tipo di scala che non ha le caratteristiche matematiche delle scale cardinali, in quanto nessuna trasformazione dei suoi valori è ammissibile. «Nessun cambiamento è possibile. I numeri sono stati originariamente

²⁶ Galung, *Theory and Methods* cit., p. 75.

²⁷ Arrotondati per i motivi e nei modi visti alla sezione 2.3. Come osserva E. W. Adams, «ciò che è irrealistico nelle teorie sulla misurazione è che implicano la possibilità di misurare esattamente quantità che variano in modo continuo – per esempio che sia possibile determinare mediante l'osservazione che un oggetto pesa esattamente due libbre anziché 2.000...1 libbre». Adams, *A Theory of Inexact Measurement*, in «Philosophy of Science», vol. XXXII, 1965, n. 3, pp. 205-28 a p. 205.

Alberto Baldissera mi ha giustamente fatto osservare che non sempre il conteggio produce solo numeri interi: se è infatti impossibile avere 2 figli e $\frac{1}{2}$, è senz'altro possibile avere 2 mele e $\frac{1}{2}$. Ogniquale volta gli oggetti sono divisibili senza essere snaturati, si possono produrre di questi esiti ibridi. Nel caso delle 2 mele e $\frac{1}{2}$, il 2 è frutto di conteggio, e il $\frac{1}{2}$ è frutto di un processo embrionale di misurazione, in cui la porzione di mela in possesso del soggetto è confrontata con una unità di misura implicita (la ideale mela intera), e il risultato del confronto arrotondato a una frazione molto semplice ($\frac{1}{2}$, e simili). Nella pratica, si hanno altri processi di pseudo-misurazione con unità stimate 'a occhio': si pensi alla frase: «quella macchina sarà lunga 4 metri e 20». Il caso non mi sembra affatto pregiudicare il motore dei due concetti di conteggio e misurazione, anche se mostra che in certe situazioni pratiche si può far ricorso ad entrambi (senza confonderli: prima il conteggio, e poi la misurazione, o viceversa).

concepiti a fini di conteggio, cioè per le scale assolute, e solo in seguito estesi a fini di misurazione»²⁸.

Assimilando le scale assolute prodotte dal conteggio alle scale metriche prodotte dalla misurazione, Stevens ofusca la fondamentale distinzione matematica fra discreto e continuo. È vero che tutti i valori registrati in una matrice sono discreti – anzi, molto spesso sono interi²⁹. Ma solo i valori derivanti da conteggi sono genuinamente interi, e rappresentano correttamente lo stato effettivo degli oggetti sulle relative proprietà; i valori derivanti da misurazione sono numeri reali arrotondati (sezione 2.3). La decisione sul numero di cifre da registrare nella matrice, che è strettamente collegata alla decisione sull'unità di misura da usare, può essere considerata una decisione sul numero di classi nella scala che risulterà: nel misurare il peso, se registriamo gli etti avremo dieci volte più classi che se registriamo i chili. Analizzando le fasi di un processo di misurazione (scegliere un'unità di misura e un punto zero, decidere quante cifre vanno registrate e come i numeri reali devono essere trasformati in numeri interi o razionali aventi il numero di cifre previsto, confrontare l'unità di misura con l'ammontare della proprietà posseduto dai singoli oggetti, arrotondare il risultato del confronto a una delle cifre registrabili), troviamo che la seconda e la quarta fase sono concettualmente analoghe a fasi del processo di classificazione: si decide sul numero di classi da creare, e si assegnano casi particolari a una classe o all'altra.

Al contrario, nessuna fase di un processo di conteggio può essere assimilata a una fase della classificazione: le classi sono né più né meno che la serie dei numeri naturali, e quindi non si richiede alcuna decisione in merito (Dio creò i numeri naturali, diceva il matematico Kronecker); l'assegnazione dei casi all'una o all'altra classe, una volta che si è contato, è automatica.

6.3. Presa coscienza di questo fatto, e alla luce delle considerazioni svolte nella sezione 4.2, si potrebbe arguire che la di-

²⁸ Galung, *Theory and Methods* cit., p. 75.

²⁹ Mentre la distinzione fra continuo e discreto è fondamentale, la distinzione fra intero e razionale non lo è. Ogni cifra con un numero finito di decimali può essere trasformata in un intero da una semplice operazione scalare sulla unità di misura. Ma questo è impossibile se il numero di decimali è infinito, come nei numeri irrazionali (e anche nei razionali periodici).

stinzione fra misurazione e conteggio è stata trascurata per il desiderio, più o meno conscio, di non prendere atto della presenza di fasi di tipo classificatorio (cioè pre-scientifico) nel processo di misurazione, santuario della scienza. Ammettere che la misurazione, a differenza del conteggio, implica una serie di decisioni del ricercatore sarebbe equivoale a riconoscere la presenza di un'altra piccola nube nel cielo radioso della scienza oggettiva (cioè automatica, indipendente da decisioni umane) contemplato dall'epistemologia positivista e neopositivista.

Se si discende da questo empireo alla bassa cucina della ricerca quotidiana, minimizzare la distinzione fra continuo e discreto ha significato autorizzare implicitamente i ricercatori a chiamare 'continue' anche variabili³⁰ con 4 o 5 categorie, come le scale Likert, e quindi ad applicare loro tecniche statistiche metriche senza alcuna consapevolezza del fatto di star introducendo una distorsione di grandezza ignota.

7. Misurazione: la scala metrica.

7.1. Tiriamo ora alcune conseguenze di quanto detto sin qui sulla misurazione. Una è che confrontare l'ammontare della proprietà posseduto dall'oggetto con l'unità di misura, e arrotondare il risultato del confronto a una cifra registrabile nella matrice dei dati, equivale a suddividere il continuum della proprietà in categorie discrete di ampiezza uguale³¹. A ciascuna di queste categorie si assegna un numero tra quelli definiti ammissibili quando si è deciso sul numero di cifre da registrare. Trasversando il confine fra una categoria e la successiva, passiamo anche da un numero al successivo nella serie dei numeri ammissibili. In linea di principio, non ci sono limiti al numero delle

³⁰ Se distinguiamo fra proprietà (attributo reale di oggetti reali) e variabile (registrazione degli stati di tale proprietà in una matrice dei dati attraverso una definizione operativa), allora l'espressione 'variabile continua' è un altro caso di *contradictio in adjecto*. Quasi nessuno stato di una proprietà che varia in modo continuo può essere esattamente registrato nel numero finito di cifre ammesso da una matrice dei dati. Parlare di 'variabili continue' è uno dei tanti modi di celare l'intervento di attività di tipo classificatorio nella misurazione.

³¹ Per la precisione, la prima categoria (etichettata con uno 0) sarà più piccola delle altre, a meno che la regola di arrotondamento sia il troncamento dei decimali.

categorie: una volta stabilita un'unità di misura, qualsiasi ammontare della proprietà può esser misurato. L'importanza di questo rilievo diverrà chiara in seguito.

Una seconda considerazione, già anticipata, è che il punto zero, che serve ad applicare l'unità di misura, non deve necessariamente essere uno zero assoluto (corrispondente a mancanza della proprietà). Anche quando uno zero assoluto esiste, si può concepire un'implicita *decisione* di applicare l'unità di misura a partire da esso, così come si può concepire anche un'esplicita decisione di applicare l'unità a partire da un altro punto: decisione che viene in effetti presa, ad esempio, quando si misura la deviazione di una serie di prodotti da uno standard.

Fare della presenza di uno zero assoluto uno dei due requisiti delle scale cardinali implica – anche se non molti lo hanno rilevato – l'esclusione dei numeri negativi da tali scale. Ciò conduce alla conseguenza paradossale che, misurando la stessa proprietà con lo stesso strumento-unità, produciamo una scala cardinale se partiamo da un punto zero assoluto, e una scala a intervalli se partiamo da un qualsiasi altro punto.

L'unico motivo addotto da Stevens, e dai numerosi metodologi che lo hanno seguito, per giustificare la loro distinzione fra scale cardinali e scale a intervalli è il fatto che dividendo fra loro i punti di una scala a intervalli si ottiene un risultato distorto. Ma – come verrà mostrato nella sezione 8.1 – questa difficoltà dev'essere largamente ridimensionata, ed è comunque del tutto irrilevante in sede di analisi dei dati con una qualsiasi delle tecniche statistiche correnti.

Tutto considerato, quindi, la differenza fra scale cardinali e scale a intervalli è certamente minore della differenza, che Stevens trascura, fra scale assolute ottenute contando e scale cardinali ottenute misurando; ed è assai minore della differenza fra la «misurazione» mediante confronti a coppie e l'assegnazione a categorie ordinate (vedi sezioni 3.2 e 3.3), procedimenti che darebbero entrambi origine a scale ordinali secondo la classificazione di Stevens. Le scale cardinali ottenute tramite misurazione e le scale a intervalli dovrebbero pertanto essere riunite nella categoria 'scale metriche', che mette l'accento su ciò che effettivamente distingue queste scale dalle altre, e cioè la presenza di un'unità di misura.

7.2. In che modo sono state stabilite le unità di misura delle variabili metriche nelle scienze sociali? Nella maggior parte dei casi, prendendole a prestito dalle scienze fisiche: l'età o l'anzianità sul lavoro di un individuo; la distanza di una città dalla capitale, sia in chilometri sia in percorrenza media; il livello medio di anidride carbonica nella sua atmosfera, la superficie totale dei suoi parchi; il territorio di una provincia, il suo consumo di energia in kwh, in litri di carburante; la rete ferroviaria di uno stato, la sua produzione annua di acciaio, la stazza complessiva della sua marina mercantile, il suo arsenale nucleare in megatons; ecc.

Alcune variabili metriche sono ottenute con operazioni matematiche su due o più altre variabili metriche: proporzione di territorio coltivato sul totale, percentuale di ore di lavoro perdute per scioperi, ecc. Altre variabili metriche sono ottenute dividendo (o in qualche caso moltiplicando) proprietà discrete (cioè soggette a conteggio) per variabili metriche: densità della popolazione, automobili prodotte al giorno, autobus circolanti per chilometro di rete tranviaria urbana, ecc. Molte variabili metriche sono ottenute con operazioni su due proprietà discrete: prodotto lordo pro capite, tasso di disoccupazione, percentuale dei voti dati a un partito sui voti validi, numero di abitanti per medico o per letto di ospedale, ecc.

Una caratteristica comune di tutte le variabili finora descritte in questo paragrafo è che la loro unità di misura è la cifra 1, ma non nella sua qualità di sottomultiplo di tutti gli interi positivi, che ne fa l'unità del conteggio, bensì come un numero reale con caratteristiche molto speciali (essere la potenza ennesima e la radice ennesima di se stesso, avere logaritmo zero in qualsiasi base, e così via). Invece, se si ottiene una variabile metrica dividendo un'altra variabile metrica per una proprietà discreta, l'unità di misura della metrica originaria varrà anche per la metrica derivata, come risulta anche verbalmente da un qualsiasi esempio (metri quadrati di spazio abitabile per abitante).

7.3. Esempi di unità di misura non prese a prestito dalle scienze fisiche sono assai più difficili da trovare, tenendo conto del fatto che tutte le proprietà collegate al denaro sono discrete,

e quindi contabili, non misurabili – anche se la loro natura è del tutto speciale, perché l'unità di conto è divisibile; mentre le unità di conto delle altre proprietà discrete non lo sono.

A partire dal 1850, psicologi come Fechner, Delboeuf, Cattell, Titchener lavorarono per stabilire unità di misura di varie proprietà sensoriali (sensibilità alla luce, al peso, all'intensità del suono, ecc.), sollevando accese critiche³². Il più noto di questi tentativi è il lavoro di Fechner sulla sensibilità ai cambiamenti di grandezza degli stimoli fisici, che sviluppava il lavoro di Ernest Weber sul potere di risoluzione (capacità di discriminare stimoli di intensità differente). Fechner pervenne alla conclusione che l'unità di misura adeguata per molte variabili 'psicofisiche' era la 'differenza appena percettibile' (*just noticeable difference*, jnd) fra stimoli di intensità differente. Come zero assoluto fu assunto il più piccolo stimolo che veniva percepito (chiamandolo A); la jnd era allora la differenza fra A e il più piccolo stimolo (chiamandolo B) che veniva percepito come «appena percettibilmente più grande» di A. Questa jnd veniva data per eguale alla jnd fra B e il più piccolo stimolo percepito come appena più grande di B, e ad ogni successiva jnd. Come Torgerson ha rilevato, tale assunto non è empiricamente controllabile, perché Fechner lo riferiva non alla grandezza fisica dello stimolo, ma alla variabile psicofisica 'sensibilità allo stimolo'. Pertanto, l'eguaglianza delle jnd doveva «esser presa... come una definizione di ciò che si intende per eguaglianza degli intervalli sul continuum psicologico»³³.

Per apprezzare pienamente la natura di questa definizione, si deve tener presente che le grandezze fisiche del più piccolo stimolo percepito (detto *limen* assoluto) e delle successive jnd variano da persona a persona, e devono essere stabilite su basi statistiche. Per *limen* assoluto fu presa la grandezza dello stimolo «abbastanza forte da esser percepito la metà delle volte»; la

³² Tra i classici della ricerca psicofisica, G. T. Fechner, *Elemente der Psychophysik*, Breitkopf & Härtel, Leipzig 1860; J. R. L. Delboeuf, *Éléments de psychophysique générale et spéciale*, Baillière, Paris 1883; G. S. Fullerton e J. M. Cattell, *On the Perception of Small Differences*, University of Pennsylvania Publ., Philadelphia 1892; E. B. Titchener, *Experimental Psychology*, Macmillan, New York 1905. Tra i critici, si veda J. Jastrow, *Critique of Psychophysical Methods*, in «American Journal of Psychology», vol. I, 1888, pp. 271-309.

³³ Torgerson, *Theory and Methods* cit., p. 133.

jnd fu definita come «quella differenza che è percepita una determinata proporzione delle volte (di solito la metà, oppure a metà strada fra sempre e quanto ci si attenderebbe sulla base del caso)»³⁴. La proporzione variava a seconda della proprietà investigata e delle procedure usate.

È evidente che il *limen* assoluto e la jnd non potevano esser definiti una volta per tutte, ma erano soggetti a variare a seconda del gruppo di soggetti sottoposti all'esperimento. E anche entro la stessa popolazione, «la forma della scala è indipendente dal tempo? e dalla proporzione scelta per definire la jnd? e da variazioni nelle istruzioni, nelle tecniche di presentazione degli stimoli, nelle tecniche di raccolta delle risposte? E sorprendentemente che tali importanti domande siano rimaste senza risposta»³⁵.

Ignorando tutti i problemi posti dalla jnd come unità di misura, Fechner proclamò che la sensibilità agli stimoli stava in relazione logaritmica con la grandezza fisica degli stimoli stessi (legge di Fechner). Altri psicofisici, tra cui Stevens³⁶, provarono in seguito che l'affermazione era empiricamente falsa: la forma logaritmica era una grossolana semplificazione di una relazione più complessa; la «legge» di Fechner era uno dei tanti frutti della tendenza dei positivisti a estrarre formule matematiche semplici ed eleganti da una realtà assai meno semplice ed elegante, chiudendo gli occhi di fronte ad ogni risultanza contraria alle loro tesi.

Ciononostante, gli argomenti discussi finora potrebbero esser considerati insufficienti a scartare definitivamente jnd e simili costruzioni come unità di misura. Si potrebbe sempre ipotizzare un ampio accordo fra scienziati per adottare una particolare serie di jnd, stabilire analizzando dati provenienti da un campione sufficientemente vasto di esseri umani, come scala di misurazione di una certa proprietà psicofisica.

Ma anche in questo caso, tuttavia, lo status della jnd come unità di misura solleverebbe un problema che – a mia conoscenza

³⁴ *Ibid.*, pp. 134 e 132-33.

³⁵ *Ibid.*, p. 131.

³⁶ S. S. Stevens, *A Scale for the Measurement of a Psychological Magnitude: Loudness*, in «Psychological Review», vol. XLIII, 1936, pp. 405-16; *On the Problem of Scales of Measurement of Psychological Magnitudes*, in «Journal of Unified Science», vol. IX, 1939, pp. 94-99. Vedi anche E. G. Boring, *The Beginning and Growth of Measurement in Psychology*, in H. Woolf (a cura di), *Quantification*, Bobbs-Merrill, Indianapolis 1961, pp. 108-27 alle pp. 112-14.

za – non è stato rilevato, e che mi sembra insolubile. Supponiamo di voler misurare la sensibilità di un individuo all'intensità del rumore mediante la relativa scala di jnd. O diamo per scontato che la sua sensibilità è uguale a quella standard, definita dalla scala (e in tal caso non stiamo misurando un bel niente, e avremmo anche potuto risparmiarci tutta la fatica necessaria a costruire la scala stessa); oppure sottoponiamo il soggetto a rumori di intensità variabile (misurata in decibels, o simile unità di misura elaborata dalle scienze fisiche). Dopodiché abbiamo due alternative: o contiamo quante jnd sono comprese nell'intervallo fra il *limen* di quell'individuo e un rumore di una certa intensità, o misuriamo (in decibels) l'intensità di rumore corrispondente alla *n*-esima differenza appena percettibile da quell'individuo. In entrambi i casi, il concetto di jnd ci è servito, ma la scala «universale» delle jnd non ci è servita a nulla, perché abbiamo dovuto ricostruire una specifica scala di jnd per quell'individuo – e così dovremo tornare a fare ogni volta che misuriamo la sensibilità di un individuo diverso. Al massimo, possiamo servirci della scala «universale» per comparare le intensità (in decibels) del *limen* e di ciascuna successiva jnd con quelle empiricamente riscontrate per i vari individui esaminati.

Una seconda e più importante considerazione è che l'unica unità di misura che interviene nel procedimento è il decibel: la jnd ha la funzione di unità di conto. Tra l'altro, a questa conclusione si poteva arrivare anche osservando che la jnd è – per definizione – indivisibile, mentre un'unità di misura dev'essere divisibile teoricamente all'infinito, altrimenti non potrebbe operare su quantità reali.

Ci si può domandare come questi evidenti difetti della jnd come unità di misura abbiano potuto sfuggire a Fechner e ai suoi colleghi. La spiegazione sta nel fatto che per gli psicologi positivisti la misurazione degli stati individuali non era un fine in sé, ma solo un mezzo per stabilire leggi universali. Grazie allo stesso assunto di uniformità della specie umana che spinse Quetelet a cercare l'*homme moyen* e Durkheim a studiare *le cas pur*, si sentivano autorizzati a stabilire «leggi» sulla base delle risultanze empiriche fornite quelle da poche centinaia di individui che passavano per i loro laboratori. La legge di Fechner era una funzione matematica mirabilmente semplice che connetteva una serie di soglie relative a un fenomeno psicofi-

sico con certi punti sulla scala che misurava un fenomeno fisico. Al di fuori di questa relazione, le soglie non avevano alcuna utilità, e in fondo neppure un'interpretazione sostanziale.

7.4. Lo stesso genere di critica si applica ad altre scale costruite in quel periodo, come per esempio la scala di Thorndike di qualità della grafia, la cui «unità di misura» era basata sull'assunto di eguaglianza di differenze notate con la stessa frequenza³⁷. Ma piuttosto che esaminare in dettaglio questa e altre unità di misura proposte, è il caso di svolgere qualche considerazione generale sulla possibilità di generare unità di misura specifiche delle scienze sociali.

Il processo di misurazione nelle scienze sociali ha in genere due attori: la persona le cui proprietà sono misurate (abituamente chiamata S, soggetto) e la persona che sta misurando (abituamente chiamata O, osservatore). Se non si richiede alcuna cooperazione attiva da S (oltre al permesso di lasciar misurare alcune sue proprietà), e nessun intervento personale da O oltre che far funzionare gli strumenti e registrare i risultati, sembra inevitabile che le unità di misura siano prese a prestito dalle scienze fisiche, perché devono essere 'leggibili' da strumenti inanimati.

Se si chiede a S una qualche cooperazione attiva, essa consisterà nel manifestare il suo stato sulla proprietà in questione mediante uno dei seguenti atti:

- a) descrivere il suo stato, verbalmente o in altro modo;
- b) accettare o respingere uno stimolo (verbale o non verbale);
- c) scegliere uno stimolo da un insieme di due o più;
- d) collocare uno stimolo in una categoria scelta tra un insieme di categorie, identificate da un'etichetta verbale o numerica, o dalla posizione spaziale;
- e) valutare uno stimolo assegnandogli un numero, oppure un segmento di lunghezza voluta;
- f) ordinare un insieme di stimoli su un continuum di sua scelta oppure suggerito da O.

³⁷ E. L. Thorndike, *Handwriting*, in «Teaching College Records», vol. XI, 1910, n. 2.

Nel casi *d*, *e*, *f*, lo stimolo può essere anche la sensazione interiore che S ha del suo stato sulla proprietà in questione.³⁸

Se si chiede a O un qualche intervento, esso consisterà nel manifestare la sua valutazione dello stato di S sulla proprietà attraverso atti che possono essere classificati con le stesse categorie usate sopra per le manifestazioni di S³⁹. Pertanto, il fatto che la partecipazione attiva sia richiesta da S o da O, oppure da entrambi, è irrilevante per il nostro discorso.

Nel caso *a*, siamo fuori dal campo di misurazione; lo stesso quando gli stimoli del caso *c*, o le etichette verbali del caso *d*, non sono nemmeno ordinati. Quando etichette o stimoli sono stati ordinati, ma non si sa nulla delle loro distanze reciproche su un eventuale continuum (vedi sezione 9), siamo ovviamente lontani da una situazione che consenta di misurare lo stato di S.

Nel caso *f*, se si richiede a S di ordinare gli stimoli con un criterio che sia legato alle sue preferenze, e adottando assunti assai poco realistici, fra i quali la transitività delle preferenze di tutti gli S (vedi sezione 3.2), la *unfolding technique* proposta da Coombs consente di delimitare lungo il continuum un certo intervallo che comprenda la posizione di S⁴⁰. Ma anche se tutti gli assunti fossero corretti, non perverremmo mai ad accertare la posizione esatta di S.

Nel casi *b* e *c*, anche se abbiamo determinato in precedenza (con tecniche che saranno illustrate nella sezione 8.3) la posizione dello stimolo su un dato continuum, non possiamo ritenere di aver misurato la posizione esatta di S su quel continuum attraverso la posizione dello stimolo che egli accetta o sceglie. Sarebbe legittimo ritenere solo se S potesse scegliere – senza errori, stanchezza o altri fattori di disturbo – uno stimolo da una serie infinita, che coprisse tutte le posizioni sul continuum. Mentre noi, per ovvie ragioni pratiche, possiamo sottoporre a S solo un insieme limitato di stimoli (caso *c*), quando non ne

³⁸ Con l'avvertenza che la descrizione (caso *a*) si riferisce allo stato di S, che gli stimoli che O valuta (*e*) o colloca in una categoria (*d*) sono le sue percezioni dello stato di S, e che gli stimoli che O accetta o respinge (*b*), sceglie (*c*), oppure ordina (*f*) sono possibili descrizioni dello stato di S.

³⁹ C. H. Coombs, *Psychological Scaling without a Unit of Measurement*, in «Psychological Review», vol. LVII, 1950, pp. 145-58. Vedi L. L. Thurstone, *Rank Order as a Psychophysical Method*, in «Journal of Experimental Psychology», vol. XIV, 1931, pp. 249-69.

sotto poniamo addirittura uno solo (caso *b*). Questa argomentazione vale ancora per il caso *d*, qualora le categorie previste siano identificate da etichette verbali.

Se invece esse sono identificate da una serie di numeri interi (come nelle scale «auto-ancorate», o nei «termometri dei sentimenti»), o dalla loro posizione in un diagramma (come nel «differenziale semantico» o nelle varie scale di «auto-collocazione»), possiamo approssimare almeno una delle caratteristiche della misurazione — la possibilità di infiniti esiti numerici — in quanto il numero delle categorie può essere molto aumentato, a differenza del numero delle categorie identificate verbalmente, che deve restare molto limitato altrimenti *S* non può effettivamente considerarle tutte al momento di scegliere⁴⁰. Ad ogni modo, anche nel caso di identificazione mediante cifre o posizioni grafiche, non sappiamo se *S* (oppure *O*), nel collocare lo stimolo, considera effettivamente l'intero arco delle possibilità, cioè l'intera scala, oppure solo una parte (vedi sul punto la sezione 8.3).

Queste difficoltà sono in parte aggirate nel caso *e*: la possibilità di infiniti esiti numerici sarebbe teoricamente raggiunta se non si ponessero limiti al numero che *S* può assegnare allo stimolo; ma questo renderebbe impossibile comparare gli stati di differenti *S*, dato che ognuno userebbe una sua scala numerica di ampiezza a noi ignota. Se fissiamo un arco di numeri accettabili (es., da 1 a 10 000), salviamo la possibilità teorica di esiti molto numerosi, se non infiniti, perché *S* od *O* potrebbero assegnare un numero con molte cifre diverse da zero (anche se in pratica quasi tutti si varranno di cifre «tonde»).

Gli esiti possibili sono, almeno formalmente, infiniti se richiediamo a *S* o ad *O* di tracciare un segmento, pur fissandone la lunghezza massima per preservare la comparabilità: infatti la lunghezza di un segmento corrisponde sempre ad un numero reale. Questa è pertanto la tecnica che più avvicina le condizioni formali della misurazione quando *S* oppure *O* prendono parte

⁴⁰ Le unità di informazione che la mente umana può elaborare contemporaneamente sono in numero limitato. Vedi G. A. Miller, *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, in «Psychological Review», vol. LXIII, 1956, pp. 81-97; R. D. Luce, *Individual Choice Behavior. A Theoretical Analysis*, Wiley, New York 1959, pp. 5 segg.

attiva al procedimento. Resta una differenza: la necessità di fissare la lunghezza massima del segmento introduce un limite estensionale ai possibili esiti — limite che invece non esiste quando si misura senza richiedere la cooperazione attiva degli attori (vedi sopra, sezione 7.1). Inoltre, naturalmente, resta il problema sostanziale delle differenze inter-individuali nella tendenza a considerare effettivamente tutta la lunghezza permessa, oppure solo una parte; tendenza che è una proprietà individuale che non ha niente a che vedere con la proprietà che stiamo misurando (vedi sul punto la sezione 8.3).

8. Assegnazione a intervalli di ampiezza nota: la scala a intervalli.

8.1. Secondo Stevens e i manuali correnti di metodologia, una scala a intervalli ha un'unità di misura ma non ha un punto zero assoluto⁴¹; quest'ultima caratteristica la distingue da una scala cardinale. Nella sezione 7.1 sono state esposte alcune conseguenze paradossali della scelta di considerare discriminante la presenza di un zero assoluto; il paradosso è rilevato anche da Hamblin: «se le scale cardinali fossero definite... sulla base della presenza di un'origine assoluta, variabili come il tempo e la distanza non avrebbero proprietà cardinali»⁴².

Ma anche dal punto di vista formale gli argomenti a favore della scelta di Stevens non sono particolarmente cogenti. È vero che — per riprendere l'esempio portato in tutti i testi — una temperatura di 20° non è due volte più calda di una temperatura di 10°; ma la differenza di temperatura fra 20° e 0° è doppia della differenza fra 20° e 10°. Come osserva Coleman, «la misurazione a intervalli è una misurazione cardinale degli intervalli»⁴³.

⁴¹ Vedi Stevens, *On the Theory* cit.; Blalock, *Statistica per la ricerca sociale*, Il Mulino, Bologna 1970², pp. 31-32; J. S. Coleman, *Introduction to Mathematical Sociology*, Free Press, New York 1964, p. 65; Galtung, *Theory and Methods* cit., pp. 73-75; Ellis, *Basic Concepts* cit., p. 64.

⁴² R. L. Hamblin, *Social Attitudes: Magnitude Measurement and Theory*, in Blalock (a cura di), *Measurement* cit., pp. 61-120 a p. 78.

⁴³ Coleman, *Introduction* cit., p. 65. Analogamente Torgerson, *Theory and Methods* cit., p. 83, e S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York 1956, pp. 26-29.

Stabilito questo, esaminiamo più da vicino l'espressione «10 metri sono il doppio di 5 metri». Essa può essere considerata una versione abbreviata dell'espressione: «l'intervallo fra il punto misurato da 0 metri e il punto misurato da 10 metri è doppio dell'intervallo fra il punto misurato da 0 metri e il punto misurato da 5 metri». In linguaggio matematico, $10 = 2 \times 5$ può essere scritto $(10 - 0) = 2(5 - 0)$. In altre parole, stiamo sempre confrontando intervalli; se a un'estremità dell'intervallo c'è lo zero assoluto, possiamo usare una formula abbreviata; altrimenti, entrambe le estremità dell'intervallo devono essere menzionate.

Pertanto, la differenza posta da Stevens fra scale cardinali e scale ad intervalli si riduce a una questione di maggiore o minore concisione della formula matematica per eseguire un'operazione (la divisione di un punto della scala per un altro punto della scala) che peraltro non è *mai* richiesta in sede di analisi dei dati. Infatti, tutte le tecniche statistiche metriche ignorano l'origine del continuum, assoluta o convenzionale che sia; esse collocano il punto zero sulla media della distribuzione e trasformano ciascun valore in una deviazione dalla media. Anche in questo caso, applicando l'impostazione di Stevens, si arriva a un paradosso: le tecniche statistiche metriche non lavorerebbero mai su scale cardinali!

8.2. Fin qui la critica logico-formale alla scelta di Stevens. Le gravi conseguenze di quell'erronea impostazione emergono puntualmente nella pratica della ricerca. Attribuendo un'importanza decisiva a una circostanza praticamente irrilevante (natura assoluta o convenzionale del punto zero), Stevens ha contribuito a togliere rilievo all'unico criterio davvero discriminante: la presenza o meno di un'unità di misura. L'etichetta 'scala a intervalli' ha subito incontrato una straordinaria popolarità nelle scienze sociali perché è stata letta come una liberalizzazione di entrambi i requisiti di una scala cardinale: presenza di uno zero assoluto e di un'unità di misura. Essa ha pertanto fornito una patina di rispettabilità scientifica ai ricercatori che volevano usare tecniche statistiche metriche su variabili per le quali non esisteva un'unità di misura, consentendo loro di continuare a pensare di star rispettando la «realità oggettiva» dei dati. Secondo Blalock «la maggior parte delle tecniche sino a oggi ela-

borate per la misura degli atteggiamenti danno origine a scale che solo in via di estrema approssimazione possono essere considerate a intervalli; in molti casi il risultato che si ottiene non può neppure considerarsi corretto se usato come scala ordinale»⁴.

Anche se l'ultima parte del giudizio di Blalock è forse troppo rigida, non si può negare che l'etichetta 'scala a intervalli' sia abitualmente applicata a variabili prive della più pallida imitazione di un'unità di misura, mentre esattamente le stesse variabili sono etichettate 'scale ordinali' dai puristi che impiegano tecniche statistiche non parametriche.

Per uscire da questa situazione di incertezza e mistificazione, è importante tornare a sottolineare che l'unico criterio veramente discriminante è la presenza di un'unità di misura, unificando sotto l'etichetta 'scale metriche' tutte le variabili che hanno tale unità, a prescindere dalla natura del punto zero.

D'altra parte, l'etichetta 'scala a intervalli' potrebbe essere riservata per quella famiglia di procedure, di importanza crescente nelle scienze sociali, che dividono – direttamente o indirettamente – il continuum in intervalli di ampiezza considerata nota per ragioni logiche, o determinata con procedure empiriche. Tra l'altro, si tratterebbe di un recupero del significato originario dato al termine 'intervallo' negli anni '30, quando fu introdotto da Thurstone nel metalinguaggio degli psicologi⁴⁵. In effetti, è probabile che Stevens sia stato indotto a dare la sua definizione di 'scala a intervalli' dal fatto che Thurstone – ancora influenzato dall'impostazione di Fechner e degli psicofisici – considerava il suo Method of Equally Appearing Intervals come un modo per *misturare* gli atteggiamenti determinando empiricamente intervalli uguali lungo un continuum. Peraltro, Thurstone era perfettamente consapevole del fatto che «l'unità di misura di questa scala... è più arbitraria... cioè, un decimo dell'estensione del continuum psicologico che si stende fra le

⁴⁵ Blalock, *Statistica* cit., p. 34.

⁴⁶ Il termine fa la sua comparsa quando Thurstone lancia il Method of Equally Appearing Intervals (vedi sezione 8.3). Vedi L. L. Thurstone, *Attitudes Can Be Measured*, in «American Journal of Sociology», vol. XXXIII, 1928, n. 4, pp. 529-54; L. L. Thurstone e E. J. Chave, *The Measurement of Attitude*, University of Chicago Press, Chicago 1929.

due affermazioni che il soggetto considera estreme fra tutte quelle della lista che gli viene presentata»⁴⁶.

Nella successiva sezione si argomenterà la tesi che, nella situazione descritta da Thurstone, è improprio parlare di un'unità di misura, per quanto grande sia stata la cura con cui l'effettiva eguaglianza degli intervalli è stata controllata empiricamente.

8.3. Nella nuova classificazione che sto proponendo, l'etichetta 'a intervalli' è conferita a tre tipi principali di scale. Il tipo A ha le seguenti caratteristiche:

- a) si definisce un arco di valori possibili stabilendo un minimo e un massimo. Entro questi due limiti, si può chiedere a S o ad O di valutare uno stimolo collocandolo in una categoria identificata da un'etichetta numerica o semplicemente dalla sua posizione su un diagramma;
- b) la sola informazione semantica che si fornisce a S (o ad O) concerne i due valori estremi. Per esempio, nelle scale «auto-ancorate», 0 = massima insoddisfazione, 10 = massima soddisfazione; nel «termometro dei sentimenti», 0 = massimo sfavore, antipatia, 10 = massimo favore, simpatia. Nella scala di auto-collocazione politica, la casella ad un'estremità è etichettata 'Destra' o 'Estrema Destra', e la casella opposta 'Sinistra' o 'Estrema Sinistra'. Nel «differenziale semantico», le due caselle estreme sono etichettate con coppie di opposti semantici come buono/cattivo, attivo/passivo, caldo/freddo⁴⁷. Talvolta viene etichettato anche il valore centrale (con 'indifferenza', 'centro', e simili). L'interpretazione semantica di tutte le altre categorie numeriche, o caselle, è lasciata a S, o ad O.
- c) il numero delle categorie è determinato da segni su un diagramma, oppure equivale al numero degli interi inclusi fra il minimo e il massimo (compresi). Per 'unità di mi-

⁴⁶ Thurstone, *Attitudes* cit., p. 533.

⁴⁷ Vedi H. Cantrel e L. A. Free, *Hopes and Fears for Self and Country: the Self-Anchoring Scale in Cross-Cultural Research*, in «American Behavioral Scientist», vol. VI, 1962; C. E. Osgood, *The Nature and Measurement of Meaning*, in «Psychological Bulletin», vol. XLIV, 1952, n. 3, pp. 197-237; D. R. Heise, *Some Methodological Issues in Semantic Differential Research*, in «Psychological Bulletin», vol. LXXII, 1969, n. 6, pp. 406-22.

sura' si intende evidentemente la differenza fra la categoria più bassa e quella immediatamente adiacente, che è supposta uguale alla distanza fra due qualsiasi categorie adiacenti (vedi il punto d). Mancano tuttavia almeno tre caratteristiche della misurazione: la «unità di misura» non viene confrontata con lo stato di S sulla proprietà; il procedimento può avere come esiti solo dei numeri interi (non dei numeri razionali o irrazionali: limitazione intenzionale) compresi fra i due interi scelti come estremi (anche i due estremi sono esiti possibili, ma tutti gli altri numeri interi non lo sono: limitazione estensionale).

Anziché lasciare che un'unità di misura sia confrontata con il suo stato sulla proprietà, S adatta la sua visione della proprietà, e del suo stato su di essa, all'intera struttura (minimo, massimo, e numero di categorie intermedie) che gli viene proposta. Enucleare una «unità di misura» in questa struttura rigida è come affermare che in un pettine la distanza tra due punti adiacenti è l'unità di misura del pettine stesso.

d) si suppone che gli intervalli fra le due categorie siano percepiti come uguali da ogni S, dato che sono rappresentati graficamente come uguali, o che consistono negli intervalli fra due interi adiacenti. Ma questo assunto può non corrispondere alla realtà, come è suggerito dalla letteratura sugli *end effects*, *zero effects*, *median effects*⁴⁸.

Ancora più importante è il fatto che — proprio per quella stessa scarsità di ancoraggio semantico che ci permette di considerare uguali gli intervalli — non sappiamo come S adatti la sua visione della proprietà alla scala numerica o grafica che gli viene proposta. Per esempio, comparando batterie di «termometri» compilate da S differenti, si trovano vistosi indizi di una radicale diversità nei modi di intendere lo strumento. Alcuni S sembrano considerare il massimo e il minimo come riferiti ad ogni possibile S, e quindi evitano di usare tali categorie estreme; altri li considerano come riferiti a loro stessi, e quindi usa-

⁴⁸ Vedi ad esempio K. Hevner, *An Empirical Study of Three Psychophysical Methods*, in «Journal of Genetic Psychology», vol. IV, 1930, pp. 191-212; Galting, *Theory and Methods* cit., p. 98; N. Jordan, *The Asymmetry of 'Liking' and 'Disliking': a Phenomenon Meriting Further Reflection and Research*, in «Public Opinion Quarterly», vol. XXIX, 1965, n. 2, pp. 315-22.

no anche le categorie estreme; per ragioni psicologiche, culturali, o sociologiche, alcuni sono propensi a usare prevalentemente i punteggi alti, altri i punteggi bassi; e così via.

Gli effetti di questa variabilità nei modi di concepire lo strumento possono essere in parte neutralizzati se, e solo se, le categorie sono numerose (come nei «termometri») e se S ha compilato almeno una dozzina di scale relative a proprietà sostanzialmente contrastanti tra loro (per esempio, simpatia per i sindacati, per gli industriali, per il movimento studentesco, per la polizia, per il femminismo, ecc.). Se S assegna punteggi tutti alti o tutti bassi a questi stimoli così diversi, possiamo supporre che ciò non dipenda dallo stato di S sulle relative proprietà, ma dalla sua particolare maniera di vedere lo strumento; di conseguenza, possiamo sentirci legittimati a «normalizzare» la sua visione dello strumento con una procedura analoga alla standardizzazione, che ho chiamato «deflazione»⁴.

Il punteggio medio assegnato da S agli stimoli della batteria – considerato un indicatore della sua tendenza ad assegnare punteggi alti o bassi indipendentemente dalla natura degli stimoli – viene sottratto da ciascuno dei singoli punteggi che, trasformato in una deviazione dal punteggio medio, viene così depurato di quella componente estranea alla proprietà che si intende misurare. Si può anche tener conto della tendenza di S a disperdere o concentrare i punteggi intorno alla loro media, calcolando la deviazione standard dei punteggi assegnati da S e usando per normalizzare i punteggi stessi, depurati dalla media.

Le scale ad intervalli del tipo B hanno le seguenti caratteristiche:

- a) si chiede a S di accettare o respingere ciascuna affermazione compresa in un elenco, oppure di scegliere una o più affermazioni dall'elenco stesso.

⁴ Mentre nella standardizzazione usiamo le medie e deviazioni standard dei punteggi di tutti i S su una sola variabile, nella deflazione usiamo le medie e deviazioni standard dei punteggi di ciascun S separatamente su tutte le variabili della batteria. Il nome «deflazione» è stato scelto per l'analogia con il procedimento con cui gli economisti depurano i singoli prezzi mediante il livello generale dei prezzi o simili indici. Per maggiori particolari, vedi A. Marradi, *Dimensioni dello spazio politico in Italia*, in «Rivista Italiana di Scienza Politica», vol. IX, 1979, n. 2, pp. 263-96 alle pp. 294-96, e *Concetti cit.*, pp. 57-58.

- b) A ciascuna affermazione è stato assegnato un numero intero approssimativamente corrispondente alla sua supposta posizione sulla scala.
- c) A ciascun S si attribuirà un punteggio sulla proprietà che corrisponde al numero assegnato all'affermazione (se ne ha scelto o accettato solo una), o alla media o mediana dei numeri assegnati a tutte le affermazioni che ha scelto o accettato.

- d) Per assegnare a ciascuna affermazione un numero che corrisponda alla sua posizione sul continuum che rappresenta la proprietà si ricorre al giudizio di un largo numero di «esperti». Per esempio, nel Method of Equally Appearing Intervals, che è il prototipo di questa famiglia di scale, ogni esperto assegna a ciascuno stimolo un numero compreso tra 0 (ammontare minimo della proprietà) e 10 (ammontare massimo della proprietà). Le affermazioni valutate in modo troppo differente dai vari esperti sono escluse dalla scala; le altre ricevono un numero che è la mediana dei numeri assegnati dagli esperti. Si usa la mediana anziché la media per evitare esiti decimali, e anche per ridurre l'influenza di assegnazioni troppo devianti dal giudizio prevalente.

- e) Ne risulta una scala formata da un certo numero (di solito limitato) di posizioni, ciascuna rappresentata da una o più affermazioni che sono state collocate in quella posizione dal giudizio «mediano» di un gruppo di esperti.

L'eguaglianza degli intervalli fra due posizioni adiacenti non è stata stabilita per fiat; è stata in un certo senso generata insieme con la scala mediante un sofisticato processo di costruzione. Tuttavia, nessuno potrebbe sostenere che tale eguaglianza è stata dimostrata vera in rapporto a un qualsiasi gruppo di S, e persino al gruppo degli esperti. Non si può nemmeno affermare che un esperto considera un'affermazione, cui abbia assegnato il numero 2, come equidistante da due affermazioni cui ha assegnato i numeri 1 e 3 rispettivamente. Lo si potrebbe dire se un esperto avesse a disposizione infiniti, o almeno moltissimi, numeri corrispondenti ad altrettante posizioni sul continuum; ma poiché di solito deve scegliere fra un insieme limitato di numeri, egli è obbligato ad «arrotondare» la posizione che assegnerebbe

a ciascuna affermazione. Collocherà quindi nella stessa posizione affermazioni che avrebbe differenziato se avesse potuto valersi di uno strumento più preciso.

Dato che le affermazioni scelte per la scala sono poi sottoposte ai S senza informarli delle posizioni loro assegnate, si dovrebbe accertare se i S tendono a scegliere o approvare solo affermazioni collocate dagli esperti in posizioni tra loro contigue, o anche in posizioni distanti, al fine di controllare il grado di accordo fra la visione di S e quella degli esperti. Ma non sono a conoscenza di alcuna pubblicazione sull'argomento. Si è molto dibattuto invece, anche sulla base di ricerche empiriche, sul problema se le posizioni assegnate alle affermazioni facendo la mediana dei giudizi degli esperti risentano o meno delle opinioni personali dei singoli esperti. Sembra che il Method of Equally Appearing Intervals riesca a neutralizzare tale influenza, a meno che gli esperti con opinioni estreme siano troppo numerosi⁵⁰.

Le scale ad intervalli del tipo C differiscono dalle altre in quanto evitano assunti circa l'eguaglianza degli intervalli fra le categorie. Informazioni sull'ampiezza di tali intervalli sono estratte elaborando le informazioni ordinali fornite dagli 'esperti'. Thurstone e i suoi collaboratori proposero tre diverse tecniche per raccogliere l'informazione ordinale e trasformarla in metrica.

Il Method of Paired Comparison chiede a ogni esperto di fare tutti i possibili confronti a coppie fra le affermazioni di un elenco, scegliendo da ogni coppia l'affermazione che corrispon-

⁵⁰ Un seguace di Thurstone, E. D. Hinckley, sostiene che S sudisti bianchi, nordisti e negri avevano costruito, con il Method of Equally Appearing Intervals, scale di 'razzismo' perfettamente equivalenti: Hinckley, *The Influence of Individual Opinion on Construction of an Attitude Scale*, in «Journal of Social Psychology», vol. III, 1932, n. 3, pp. 283-95. Ma questo risultato era stato ottenuto eliminando tutti gli esperti che avevano accumulato troppe affermazioni nelle posizioni estreme della scala; visto che questo comportamento è tipico di persone che si collocano personalmente su posizioni estreme, le conclusioni di Hinckley non possono essere accettate, come hanno sottolineato C. I. Hovland e M. Sharif, *Judgmental Phenomena and Scales of Attitude Measurement: Item Displacement in Thurstone Scales*, in «Journal of Abnormal and Social Psychology», vol. XLVII, 1952, pp. 822-32. È in corso in Italia una replica della ricerca di Hinckley condotta in modo più rigoroso, al fine di dirimere la questione.

de a un maggiore ammontare della proprietà che si sta misurando. Per ogni coppia, la differenza percentuale fra gli esperti che scelgono l'affermazione A e quelli che scelgono la B è presa come misura della distanza fra A e B sul continuum⁵¹. Naturalmente, aumentando il numero delle affermazioni aumentano esponenzialmente il lavoro richiesto agli esperti, il rischio di scelte intransitive (vedi sezione 3.3), e la conseguente difficoltà di rimediare alle incongruenze nei dati.

Il Rank Order Method chiede ad ogni esperto di ordinare tutte le affermazioni di un elenco a seconda della loro supposta posizione sul continuum che rappresenta la proprietà che si sta misurando. Come nel caso precedente, la differenza percentuale fra gli esperti che collocano l'affermazione A prima della B e quelli che fanno l'inverso è presa come misura della distanza fra A e B⁵². Questa tecnica costituisce un notevole miglioramento rispetto alla precedente, in quanto riduce il lavoro degli esperti ed esclude la possibilità di scelte intransitive.

Il Method of Successive Intervals raccoglie l'informazione in modo analogo al Rank Order, dando però agli esperti la possibilità di assegnare la stessa posizione a due o più affermazioni. L'informazione metrica, però, è estratta attraverso una procedura molto complicata, che implica un assunto di distribuzione normale delle posizioni effettive occupate da ciascuna affermazione agli occhi dei vari esperti⁵³.

Le scale costruite con le due ultime tecniche presentate sono chiaramente preferibili, da un punto di vista epistemologico, alle scale di tipo B, in quanto pongono restrizioni assai minori alle valutazioni degli esperti, e da esse estraggono informazione metrica nel modo migliore nella situazione data.

⁵¹ Vedi L. L. Thurstone, *A Law of Comparative Judgment*, in «Psychological Review», vol. XXXIV, 1927, pp. 273-86; *The Method of Paired Comparison for Social Values*, in «Journal of Abnormal and Social Psychology», vol. XXI, 1927, pp. 384-97.

⁵² Vedi Hevner, *An Empirical Study of Scales Constructed by Three Psycho-*

physical Methods, in «Psychometrika», vol. II, 1937, n. 3, pp. 179-98; C. I. Mosier, *A Modification of the Method of Successive Intervals*, in «Psychometrika», vol. V, 1940, pp. 101-7; A. L. Edwards, *The Scaling of Stimuli by the Method of Successive Intervals*, in «Journal of Applied Psychology», vol. XXXVI, 1952, pp. 118-22.

Tuttavia, anche le scale di tipo C non possono esser considerate scale metriche, in quanto solo una serie limitata di affermazioni, la cui posizione è stata predeterminata, viene sottoposta ad A. Non si può dare per scontato che S condividerebbe, se le conoscesse, le posizioni sul continuum assegnate dagli esperti alle varie affermazioni, né si può supporre che il suo stato sulla proprietà sia esattamente rappresentato dalla posizione dell'affermazione che sceglie o accetta – se una sola – o dalla media di tali posizioni, se più di una.

In effetti, S sceglie o approva la o le affermazioni che giudica più vicine al suo stato effettivo sulla proprietà, ma è spesso consapevole di darne una rappresentazione distorta, non di rado se ne lamenta, e talvolta si rifiuta addirittura di scegliere o approvare alcuna delle affermazioni sottopostegli. In ogni caso, ben di rado lo stesso S è in grado di valutare l'entità della distorsione che la sua posizione ha subito; tanto meno potrebbe farlo O, o qualsiasi altro estraneo.

9. Assegnazione a categorie ordinate: la scala (*simplex*).

9.1. Nella sezione 3.2 si è sostenuto che quando la proprietà consiste in una serie di categorie ordinate non si dovrebbe parlare di 'misurazione ordinale', ma di 'assegnazione a categorie ordinate'.

Le categorie possono essere ordinate «naturalmente» (come i livelli gerarchici in un'organizzazione, o i titoli di studio), o possono venire ordinate dal ricercatore in relazione a qualche criterio (ad es., le differenti occupazioni potrebbero venire ordinate in termini di prestigio sociale).

Si assegna abitualmente l'etichetta 1 alla categoria più bassa, la 2 alla successiva, e così via. Tuttavia, come ha osservato Tufte, «l'assegnazione della serie dei numeri naturali (1, 2, 3, 4, ...) alle categorie è di solito inappropriata. Essa non è, comunque, una scelta più saggia o più prudente dell'assegnazione di una qualsiasi altra serie di numeri; anzi, è una scelta chiaramente peggiore se urta contro le nostre conoscenze intorno all'oggetto... [Infatti] una buona ragione per assegnare numeri alle categorie ordinate è che il ricercatore spesso sa attorno al fenomeno qualcosa di più di quanto è implicato dal mero ordine

delle categorie. Assegnare numeri aiuta a sfruttare questa ulteriore informazione nella misurazione»⁸⁴.

Sottoscrivo le tesi di Tufte, con un'estensione e una riserva. Se è saggio il ricercatore che fa uso della «conoscenza tacita» che ha di un fenomeno, anziché esorcizzarla come «soggettiva», è certo ancora più saggio il ricercatore che, anziché fidarsi interamente della propria conoscenza tacita, sfrutta anche quella altrui. In altre parole, l'assegnazione delle etichette numeriche alle categorie ordinate dovrebbe essere fatta nel modo più intersoggettivo possibile, dato il patrimonio di conoscenze ambientali cui il ricercatore può attingere.

Fin qui l'estensione; vediamo ora la riserva, che è terminologica. Pur essendo consapevole quanto nessun altro della differenza fra misurazione e assegnazione di etichette numeriche, nel brano citato Tufte usa il termine 'misurazione' nel solito modo estensivo. Questa cattiva abitudine è forse anche dovuta alla scarsa diffusione di un termine generale che designi il processo di assegnazione di etichette numeriche ai casi concreti al fine di registrarli in una matrice dei dati – processo che è un *genus* rispetto al quale la misurazione è solo una delle *species*. Un termine generale esiste, ed è 'operativizzazione', purtroppo assai goffo, anche se meno del suo gemello 'operazionalizzazione', che è un calco esatto dell'originale inglese.

Probabilmente siamo in presenza di un clamoroso esempio di come una carenza del patrimonio terminologico si ripercuota sul patrimonio concettuale (contribuendo a provocare non solo un indebito stiramento del termine 'misurazione', ma anche un grave offuscamento del relativo concetto) e di lì sulle concrete procedure di ricerca di una generazione di studiosi.

9.2. L'uso delle categorie ordinate non è limitato al caso in cui gli stati degli oggetti sulla proprietà sono discreti: in molte situazioni la proprietà può essere concepita come un continuum, ma non possiamo misurarla per mancanza di un'unità; dobbiamo pertanto dividere il continuum in segmenti corrispondenti a categorie. Nella sezione 8.3 sono state presentate varie tecni-

⁸⁴ E. R. Tufte, *Improving Data Analysis in Political Science*, in Tufte (a cura di), *The Quantitative Analysis of Social Problems*, Addison-Wesley, Reading 1970, pp. 437-49 alle pp. 441 e 440.

che per stimare l'ampiezza dei segmenti su basi logiche (scale di tipo A) o empiriche (scale di tipo B e di tipo C). Ma se non possiamo o vogliamo – per un qualsiasi motivo – ricorrere a queste tecniche, la nostra stima dell'ampiezza dei segmenti, e pertanto la nostra assegnazione di etichette numeriche alle categorie ordinate, dev'esser basata sulla conoscenza tacita del fenomeno in questione.

Invece, nella maggior parte dei casi, non si fa alcun tentativo di stimare l'ampiezza dei segmenti sul continuum, ricorrendo come si è visto alla serie dei numeri naturali; l'eguaglianza degli intervalli fra le categorie è solo una conseguenza meccanica, non proclamata né cercata. Peraltro, la tecnica attualmente più diffusa per costruire scale – quella proposta da Likert – assegna agli interi contigui come etichette alle categorie in base a un esplicito assunto di eguaglianza degli intervalli. L'assunto è basato sulla stessa strategia seguita dagli inventori delle scale di tipo A (sezione 8.3), cioè ridurre al minimo la caratterizzazione semantica delle categorie al fine di poter sostenere che solo la posizione spaziale, o l'etichetta numerica, distingue una categoria dall'altra.

Tuttavia, mentre nelle scale a intervalli del tipo A il contenuto semantico di tutte le categorie non estreme è effettivamente ridotto a un numero o a una posizione, le categorie delle scale Likert hanno un contenuto assai maggiore (approvo pienamente, sono parzialmente in disaccordo, ecc.), il che a mio avviso è sufficiente per rovesciare l'onere della prova. In altre parole, mentre è ragionevole attendersi che la categoria '50' sia vista come equidistante da '45' e da '55', e chi sostiene il contrario deve provarlo, non c'è ragione per cui si debba credere che agli occhi della gente 'approvo parzialmente' appare equidistante da 'approvo pienamente' e da 'incerto', o che 'mi piace' e 'non mi piace' sono percepiti alla stessa distanza da 'mi è indifferente'. Qui è l'equidistanza a dover essere provata; nelle ricerche empiriche svolte ad hoc, è stato se mai riscontrato il contrario³⁵.

Ancor meno giustificato sarebbe un assunto di equidistanza delle categorie nelle 'scale di frequenza': infatti, tale assunto

³⁵ Vedi Jordan, *The Asymmetry* cit.; Galtung, *Theory and Methods* cit., p. 98.

non è mai esplicitato, e l'equidistanza deriva meccanicamente, come si diceva, dall'uso della serie dei numeri interi. Tuttavia, si potrebbe provare a seguire il consiglio di Tufte, usando la propria conoscenza tacita al fine di essere «saggiamente arbitrari». Prendiamo ad esempio la proprietà 'frequenza in chiesa'. Operativizzata attraverso la domanda: «Quanto spesso va in chiesa?», e confrontiamo tre risposte: «mai», «a Natale e Pasqua», «una volta al mese circa». Assegnando etichette 1, 2, 3 ignoreremo il messaggio affidato a quel rigido «mai», che dovrebbe essere chiaramente differenziato dall'atteggiamento più possibilista manifestato dalle altre due risposte.

Categorie costituite da frasi compiute anziché da espressioni brevi ed ellittiche dovrebbero ovviamente offrire maggiore spazio per un'analisi semantica, al fine di assegnare etichette numeriche che rispecchino le distanze fra le categorie lungo il continuum. Tuttavia, non si dovrebbe mai dimenticare che le percezioni degli scienziati sociali, singoli o in gruppo, non sono affatto necessariamente le stesse delle percezioni della popolazione su cui si svolge la ricerca. Il disaccordo può benissimo investire non solo l'ampiezza degli intervalli fra le categorie, ma lo stesso ordine delle categorie sul continuum.

Per aggirare questa difficoltà, Guttman ricorse a brevi battorie di affermazioni, ciascuna delle quali S doveva accettare o respingere. L'ordine delle affermazioni sul continuum non era pre-determinato, ma veniva stabilito a posteriori sulla base della percentuale di accettazioni raccolta da ciascuna affermazione. Quanto all'ampiezza degli intervalli fra le affermazioni sul continuum, Guttman rinunciava a qualsiasi tentativo di stimarla. Tuttavia, questo criterio apparentemente elementare implicava un assunto di cumulatività delle risposte di S: se S accettava un'affermazione «difficile», doveva accettare anche tutte quelle più «facili»; altrimenti, la percentuale delle accettazioni non poteva esser messa in rapporto con il loro ordine lungo il continuum. Questa condizione rende la tecnica inadatta per tutte quelle proprietà che non possono esser concepite in termini di maggiore o minore *difficoltà* per S. Il campo di applicazione ideale sono quindi le capacità fisiche o psichiche.

Ma la tecnica di Guttman è stata largamente applicata anche in ricerche sugli atteggiamenti e i valori. Naturalmente si è riscontrato che in tali settori l'ordine delle affermazioni non è lo

stesso per tutti i S: molti respingono frasi che, in base alle altre risposte date, dovrebbero accettare, e viceversa. Guttman ha etichettato come 'errori' queste differenze inter-individuali, e ha mostrato come si può minimizzare l'incidenza di tali errori sperimentando ordini alternativi delle frasi ed eliminando frasi e soggetti dalla matrice⁵⁶. Ma in tal modo si tende a privilegiare le affermazioni estreme, che sono troppo «facili» per essere respinte da qualcuno, o troppo «difficili» per essere accettate: una scala composta esclusivamente da affermazioni del genere produrrebbe ovviamente un'incidenza minima di errori⁵⁷. Inoltre, aumentare il numero di affermazioni nella scala significa, a parità di altre condizioni, aumentare esponenzialmente la probabilità di errori: questo tende a eliminare le scale più ricche e informative.

Ma il maggiore difetto della tecnica di Guttman emerge quando, dopo aver ordinato le affermazioni in modo da minimizzare gli errori, si tratta di assegnare i punteggi individuali. Se S non ha commesso «errori», il suo punteggio coinciderà con la posizione della più «difficile» delle frasi che ha accettato; ma se ha commesso «errori», in molti casi il suo punteggio non sarà determinabile⁵⁸.

Il sostanziale fallimento del tentativo di Guttman fa pensare che la sua strategia di indebolire e de-strutturare lo strumento (rinunciando a ogni valutazione di tipo metrico, o anche a un ordine a priori delle affermazioni) non sia il modo giusto di affrontare il problema della mancanza di unità di misura specifiche delle scienze sociali.

Facoltà di Scienze Politiche, Università di Bologna.

⁵⁶ L. A. Guttman, *A Basis for Scaling Qualitative Data*, in «American Sociological Review», vol. IX, 1944, pp. 139-50.

⁵⁷ B. W. White e E. Saltz, *Measurement of Reproducibility*, in «Psychological Bulletin», vol. LIV, 1957, n. 2, pp. 81-99.

⁵⁸ A. F. Henry, *A Method for Classifying Non-Scale Response Patterns in a Guttman Scale*, in «Public Opinion Quarterly», vol. XVI, 1952, n. 1, pp. 94-106.