

Claudio La Rocca

Presupposti epistemologici della valutazione/misurazione

Il titolo del mio contributo, che suona forse un po' enfatico, vuole proporre una breve riflessione sul rapporto tra le procedure istituzionalizzate di valutazione della ricerca e la dimensione più generale della valutazione in quanto tale. Vorrei chiedermi cioè in che misura quella che per chiarezza andrebbe chiamata neo-valutazione¹, e che in questa sede ho proposto di indicare con "valutazione/misurazione", si ponga in continuità (o addirittura esprima) le pratiche consuete di valutazione esercitate nelle comunità scientifiche e, più in generale, in molti aspetti della nostra vita. Anticipo da subito che la tesi sarà negativa, in favore cioè di una chiara distinzione e netta discontinuità, che credo meriterebbe di essere oggetto di riflessione più di quanto solitamente avvenga.

La questione del rapporto tra neovalutazione e la tradizionale dimensione del valutare riguarda più da vicino la valutazione in ambito umanistico, perché l'uso di tecniche non-bibliometriche sembra conservare o voler conservare, rispetto alla bibliometria, una maggior continuità, appunto, con quella che è una consueta e inevitabile pratica nelle comunità di ricerca. Seppure ottenuti con metodi diversi dalla bibliometria, anche in ambito umanistico-sociale si mira, tuttavia, alla produzione di indicatori quantitativi. Se si guardano i criteri indicati nell'ultimo bando VQR, si vede come questi siano riferiti a fattori quali originalità, rigore metodologico, impatto, intesi non come "proprietà" del prodotto di ricerca, ma come "il livello al quale" (è questa la non elegantissima espressione ivi ricorrente) il prodotto realizza determinate prestazioni.

Prima di entrare in questo ordine di problematiche, consentitemi di ricapitolare delle semplici ovvietà, che possono però aiutarci a entrare nel nostro problema "dal basso", senza schemi prefissati. E forse a percepire il carattere problematico di ovvietà solo apparenti.

¹ Ho usato questo termine in LA ROCCA 2013. Al momento della stesura di quell'articolo non avevo letto BAEHLHAUSER – GORI – SAURET 2011, nel quale Roland Gori propone il termine *néo-évaluation* (già usato in un suo libro del 2010). Mi dispiace di non avervi fatto riferimento in quella occasione, ma sono contento della convergenza: come scriveva Kant, «se si vuole essere inventori, si pretende di essere i primi; se si vuole solo la verità, si richiedono precursori» (KANT 1924, 255).

1. Ricerca, verità, criteri

La prima considerazione ovvia è che l'obiettivo di ogni ricerca – nell'ambito delle scienze naturali come delle discipline “sociali” – è il raggiungimento della verità. La “ricerca” si chiama così perché cerca la conoscenza, e di questa è componente imprescindibile, secondo il modello classico della conoscenza come “credenza vera giustificata”, la verità. La verità – sembra – non ha gradazioni. Una tesi, un teoria è vera o falsa, *tertium non datur*. O meglio, il terzo c'è, ma è un terzo epistemico, non logico: la probabilità. La probabilità non indica quanto siano vere le nostre conoscenze, ma quale sia il grado del nostro possibile *accesso* alla verità. Riguarda il terzo elemento della triade ricordata, la giustificazione. Di probabilità si parla nel caso che questo grado sia misurabile, altrimenti si possono usare altre espressioni, si può parlare, ad esempio, di plausibilità, verisimiglianza.

Nel caso immaginario – è quello che i filosofi chiamano un “esperimento mentale” – in cui avessimo strumenti che garantissero la verità delle nostre teorie (e dunque che la ricerca abbia “trovato” con certezza), meriterebbe il titolo di ricerca solo quella vera. E in questo caso – sembra, ed è, paradossale – la ricerca non sarebbe valutabile. O sarebbe ricerca o non lo sarebbe. Dunque – il paradosso lo si può formulare anche così – una valutazione obiettiva della ricerca ha come sua condizione che non vi sia ricerca obiettivamente valutabile; almeno se come ideale valutazione obiettiva della ricerca (o, come si dice oggi, del suo “prodotto”)² si intenda quella che la valuta rispetto al suo obiettivo, riguardo la sua verità o falsità.

La valutazione della ricerca (la visione condivisa su di essa) si muove tuttavia sul ragionevole presupposto secondo il quale la ricerca resta in linea di principio sempre tale, *ricerca*, ossia una attività finalizzata alla conoscenza ma priva di un accesso assoluto alla verità. Presupporre il contrario vorrebbe dire ammettere dei dogmi, riferirsi non alla ricerca ma ad una *dottrina*, e la valutazione della ricerca non si distinguerebbe da un tribunale di difesa della verità – insomma, da una Santa Inquisizione³. Ciò presupposto, dunque, il focus si sposta – appunto, del tutto ragionevolmente – dalla

² Qui ci sarebbe un'altra ovvietà da ricordare: essendo la ricerca un processo, gli attuali sistemi di “valutazione della ricerca”, che sono incentrati sui “prodotti” della stessa, non valutano la ricerca stessa in quanto tale. La valutano piuttosto in modo indiretto: dalla qualità dei prodotti si inferisce la qualità del processo, una inferenza anche questa niente affatto priva di problemi. Si pensi, per fare un solo e il più semplice esempio, al caso di complesse procedure sperimentali che devono servire a verificare un'ipotesi e che danno – può avvenire per anni – risultati negativi: si tratta di un processo di ricerca assolutamente rispettabile, utile, per molti aspetti “produttivo”, ma che solo in alcuni casi si traduce in “prodotti”. Sulla non coincidenza tra attività di ricerca e prodotti ha insistito PINTO (2012, in particolare pp. 486 ss).

³ Non si tratta solo di un ragionevole presupposto, ma del modello costitutivo della Università humboldtiana. Come ricorda PINTO 2018, Humboldt fonda l'università sul «principio» «di considerare la scienza come qualcosa di non ancora del tutto trovato e che mai possa essere interamente scoperto, e di ricercarla incessantemente come tale».

qualità della ricerca nel senso stretto ed estremo (ideale) che ho ricordato ad una valutazione delle *procedure* di ricerca. Qui bisogna aprire però una distinzione, che ci porta più vicino al nostro tema specifico.

L'ambito dove il consenso circa le procedure è maggiormente generalizzato, è quello delle scienze cosiddette "dure". Qui il metodo diffuso di (neo)valutazione è – a torto o a ragione – quello "bibliometrico". La bibliometria nella sostanza *non* è un metodo di valutazione della qualità della ricerca, ovvero lo è in modo marcatamente indiretto e inferenziale. In modo diretto, la bibliometria misura semplicemente la reputazione (in modo solo relativamente diretto, o solo più diretto, perché la reputazione è a sua volta inferita dal cosiddetto "impatto", che di nuovo è inferito dal numero di citazioni)⁴. La reputazione viene considerata – nell'orizzonte di una ampia condivisione di metodologie e procedure – un indice di *rilevanza* della ricerca, che a sua volta, se deve contribuire a valutare la ricerca rispetto alla sua finalità più propria, conoscitiva, dovrebbe segnalare *anzitutto* qual è l'ampiezza di ulteriori conoscenze (o ricerche?) che viene resa possibile. Si badi bene che anche in questo caso la verità della conoscenza perseguita e acquisita è solo sullo sfondo, non costituisce ciò che fa la differenza. In altri termini, non è il fatto che una ricerca abbia raggiunto o meno la verità – abbia prodotto conoscenza – ad essere in primo piano, ma la sua fruttuosità (uso il termine per non usare quello di "produttività") conoscitiva⁵. La differenza tra un articolo poco citato e uno molto citato non sta in quanta conoscenza rispettivamente *contengano*, ma in quanta conoscenza *producano*. Per fare due esempi su poli opposti: la formulazione della teoria della relatività o una ricerca sulla composizione della lava del Vesuvio hanno fruttuosità scientifica molto diversa, pur potendo avere lo stesso grado di accesso alla verità, di "conferma" (la seconda presumibilmente, in una prima fase almeno, anche decisamente maggiore). Questo ci ricorda che la valutazione bibliometrica premia un tipo di ricerca a svantaggio di altra, del tutto indipendentemente dalla sua qualità, quanto meno dalla sua sola *qualità conoscitiva*.

Se veniamo alla valutazione non-bibliometrica, che qui più direttamente ci interessa, possiamo ripartire dai criteri dell'ultima VQR per quest'ambito, cui prima

⁴ «L'origine epistemica della citazione» sta nel suo essere «condizione di accesso alla comunicazione scientifica», in quanto serve a render conto dello stato dell'arte e con esso della originalità della ricerca (cf. BONACCORSI 2015, 20). Altra cosa è la funzione della citazione in quanto indicatore di impatto e quindi indirettamente di qualità. In questo caso non è la circostanza del citare, ma quella dell'*essere citato* a essere in gioco. Sulle complesse problematiche riguardo le citazioni cf. *ibid.*, 31 ss.

⁵ Prescindo qui, per non complicare il quadro rispetto ai punti che interessa sottolineare, da fattori ulteriori che possono condizionare e condizionano l'impatto di una ricerca, dalla sua rilevanza economica al suo essere all'interno di un progetto di ricerca più ampio o di un corrente di ricerca *mainstream*, o altro. Tutti fattori assolutamente rilevanti per una discussione sull'esito dei meccanismi di valutazione e di quelli bibliometrici in particolare, ma qui mi interessa rivolgere l'attenzione sulla misura nella quale tali meccanismi già solo in linea di principio mantengono il nesso con l'obiettivo conoscitivo della ricerca.

facevamo cenno. La questione se un prodotto di ricerca raggiunga il suo obiettivo cognitivo è approssiata anche qui, come dicevamo, con le buone ragioni ricordate, in modo indiretto. La *koiné* congetturale e fallibile che vale nel campo della scienze naturali vale con ragioni più marcate nelle scienze “umane” o dello spirito, in cui le procedure argomentative e di prova producono un grado minore di affidabilità delle nostre credenze. Dunque i criteri di valutazione sono fin dal principio anche qui indiretti, a dispetto dell’approccio meno inferenziale dato dalla tecnica della *peer review*. Leggiamo i criteri della ultima VQR nelle aree non bibliometriche:

Il giudizio di qualità si riferisce ai seguenti criteri:

- a) originalità, da intendersi come il livello al quale il prodotto introduce un nuovo modo di pensare in relazione all’oggetto scientifico della ricerca, e si distingue così dagli approcci precedenti allo stesso oggetto;
- b) rigore metodologico, da intendersi come il livello al quale il prodotto presenta in modo chiaro gli obiettivi della ricerca e lo stato dell’arte nella letteratura, adotta una metodologia appropriata all’oggetto della ricerca e dimostra che gli obiettivi sono stati raggiunti;
- c) impatto attestato o potenziale nella comunità scientifica internazionale di riferimento, da intendersi come il livello al quale il prodotto ha esercitato, o è suscettibile di esercitare in futuro, un’influenza teorica e/o applicativa su tale comunità anche in base alla sua capacità di rispettare standard internazionali di qualità della ricerca⁶.

Pochi, credo, metterebbero in dubbio l’importanza dell’originalità e del rigore metodologico nella ricerca. Sull’impatto il discorso sarebbe più problematico e più

⁶ I criteri del GEV area 10 sono ripresi da quelli del bando:

«7.1. Ai GEV è affidata la responsabilità di valutare la qualità di ciascuno dei prodotti di ricerca selezionati dalle Istituzioni. Il giudizio di qualità si riferisce ai seguenti criteri:

- a. originalità, da intendersi come il livello al quale il prodotto introduce nuove prospettive critiche o nuove acquisizioni in relazione all’oggetto scientifico della ricerca, e si distingue così in modo significativo da precedenti lavori sullo stesso argomento;
- b. rigore metodologico, da intendersi come il livello al quale il prodotto presenta in modo chiaro gli obiettivi della ricerca e lo stato dell’arte nella letteratura, adotta una metodologia appropriata all’oggetto della ricerca e dimostra che gli obiettivi sono stati raggiunti;
- c. impatto attestato o potenziale nella comunità scientifica internazionale di riferimento, da intendersi come il livello al quale il prodotto ha esercitato, o è suscettibile di esercitare in futuro, un’influenza teorica o applicativa su tale comunità anche in base alla sua capacità di rispettare standard internazionali di qualità della ricerca.

7.2. A seguito di tale giudizio di qualità ogni pubblicazione sarà attribuita a uno dei seguenti livelli di merito, così come disciplinati dal DM e dal bando VQR: a. Eccellente b. Elevato c. Discreto d. Accettabile e. Limitato f. Non valutabile».

complesso. Tuttavia, già sulla definizione di originalità si possono avere dubbi. Che tipo di originalità è un requisito per una buona ricerca? Originalità nei risultati oppure «nel modo di pensare»? Una buona e importante edizione critica offre senz'altro contributi conoscitivi nuovi, non necessariamente o direi raramente «un nuovo modo di pensare». Nelle scienze “dure”, più legate ad un paradigma, il nuovo modo di pensare è l'assoluta eccezione. Nelle scienze “multiparadigmatiche”, se vogliamo chiamarle così, ossia nell'ambito delle scienze umane, il nuovo modo di pensare è spesso, in un senso, carattere di ogni contributo scientifico, e, in un altro, è un requisito di pochissimi lavori, tali da introdurre davvero un nuovo approccio. L'interpretazione di un evento storico è nuova per definizione, nel senso più minimale (altrimenti è un plagio); se invece ci si riferisce alla proposta di una nuova metodologia, questo è un caso che riguarda una porzione molto limitata e per così dire emergente del lavoro di ricerca.

Naturalmente la sola originalità, non accompagnata da plausibilità minima di quanto viene proposto, non è un valore. Proprio tenendo conto di ciò, è previsto, sembra, il secondo criterio: il rigore metodologico. Anche qui la definizione data si presta a perplessità: parlare di un livello che misuri se e quanto sono presentati «in modo chiaro gli obiettivi della ricerca e lo stato dell'arte nella letteratura» sembra corrispondere ad una mera prescrizione espositiva; il riferimento alla «metodologia appropriata» sembra una tautologia, si presta poco a fungere da criterio (qual è il criterio per una metodologia appropriata? che sia appropriata). Più solido sembra essere il riferimento alla dimostrazione del raggiungimento degli obiettivi della ricerca, che, trattandosi di obiettivi conoscitivi, mette in gioco il riferimento alla verità e ai modi in cui la verità di quanto si espone cerca di essere fondata o giustificata. L'evocare la “dimostrazione” porta maggiormente in primo piano la valutazione *del contributo cognitivo*; anche qui la forza valutativa del criterio è tuttavia debole, perché il fatto che il risultato cognitivo debba essere corroborato è una condizione minima perché si possa parlare di ricerca, non un criterio di qualità. Anche se è vero che il criterio proposto riguarda, così come è formulato, «il livello al quale» questa condizione è soddisfatta.

Non approfondisco il terzo criterio, che mi sembra, come accennavo, il più discutibile. Segnalo solo che il riferimento all'impatto potenziale o influenza futura richiede discutibili doti divinatorie, e il riferimento agli standard internazionali è tanto indeterminato quanto fuori luogo in un criterio che dovrebbe *definire* standard e non rimandare ad altri (ammesso siano identificabili). In generale, il terzo criterio tende a restaurare, in modo fumoso, il peso dell'elemento reputazionale che è centrale nelle valutazioni bibliometriche, come se questo da indice, strumento, fosse diventato fine in sé. C'è da riflettere quale senso abbia – in una procedura come la revisione tra pari che avrebbe il vantaggio di un approccio diretto e intrinseco ai testi ed alla loro qualità – il riferimento ad una qualità putativa, indiretta ed estrinseca. Lascio qui da parte anche la

questione dell'interazione tra loro dei tre criteri, che pure sarebbe interessante discutere⁷.

2. *Valutare non è commisurare*

È possibile che questi cenni diano l'impressione di obiezioni sofistiche, di un approccio pregiudizialmente non simpatetico con il testo del bando VQR⁸. Tutti sappiamo cos'è una ricerca originale, rigorosa e influente, e cosa non lo sia. E tuttavia è proprio il passaggio da questo sapere condiviso alla sua applicazione e declinazione nella valutazione/misurazione ciò che crea precomprensioni in parte illusorie, su cui vorrei richiamare se possibile l'attenzione.

Il punto decisivo – al di là o prima di altre obiezioni – è il fatto che non sono in gioco solo originalità, rigore, influenza, ma le loro gradazioni o “livelli”. Questo davvero non è né un dettaglio né un cavillo. È invece uno slittamento semantico decisivo che non dobbiamo farci sfuggire. Quando valutiamo, attribuiamo una qualità positiva o negativa. Un quadro è bello, un formaggio è buono, un coltello taglia bene. Non attribuiamo livelli. Quando questo avviene, è una operazione differente. Questo vuol dire che segue una logica differente, finalità differenti, ha condizioni differenti, conseguenze differenti. Sottolinearlo può avere qualche senso.

Non è questo il luogo, né avrei il tempo di sviluppare una analisi adeguata del rapporto tra valutazione e quella che viene chiamata “commisurazione”. Vorrei proporre solo alcune considerazioni semplici, ma forse utili.

⁷ C'è da notare che il bando ultimo definisce il livelli di “merito” facendo convergere i tre criteri. Ad esempio: «Elevato (peso 0,7): la pubblicazione raggiunge buoni livelli in termini di originalità e rigore metodologico, e ha conseguito o è presumibile che consegua un impatto significativo nella comunità scientifica di riferimento a livello internazionale e/o nazionale. Idealmente, essa si colloca nel segmento 10-30% della distribuzione della produzione scientifica internazionale dell'area cui appartiene». Il bando precisa però in seguito che: «Le definizioni dei livelli di qualità hanno carattere esemplificativo in quanto fanno riferimento esclusivamente ai casi in cui le valutazioni attribuite ai tre criteri risultino concordanti. Per i casi in cui i criteri siano invece discordanti (ad esempio per prodotti che raggiungano i massimi livelli sul piano dell'originalità, pur avendo un livello di rigore metodologico non altrettanto elevato, e ottenendo magari un impatto ridotto) le regole di classificazione saranno definite dai GEV». I GEV di area 10 e 11 *non hanno definito affatto* queste regole di classificazione nei criteri pubblicati (cf. <http://www.anvur.it/wp-content/uploads/2016/02/DefCriteri%20GEV%2010.pdf>; <http://www.anvur.it/wp-content/uploads/2015/11/Criteri%20GEV%2011a.pdf>); nel documento di area 10 sono ripresi semplicemente i criteri generali del bando; in quello di area 11 non vi è alcuna indicazione di criteri. Questa circostanza, più che segnalare una omissione, segnala una obiettiva difficoltà di definire sub-criteri di classificazione, in presenza di criteri di classificazione già generici o discutibili.

⁸ Un'analisi un po' più ravvicinata dei criteri della precedente VQR (analoghi a quelli della successiva) è in LA ROCCA (2013, 78 ss.).

In un importante saggio di Espeland e Stevens la *commensuration* è definita come «l'espressione o la misurazione delle caratteristiche normalmente rappresentate da diverse unità in base a una metrica comune»⁹. È un processo che realizziamo in una molteplicità di casi, al punto da costituire – dicono stessi autori – «un tratto fondamentale della vita sociale», che «cambia i termini di ciò di cui si può parlare, come valutiamo e come trattiamo ciò che noi valutiamo. È simbolica, intrinsecamente interpretativa, profondamente politica e troppo importante per essere lasciata implicita».¹⁰ È altrettanto evidente come un tratto ancora più fondamentale della vita sociale, più diffuso e quotidiano della stessa commisurazione, sia la valutazione, intesa nel senso più basilare, distinta dalla neovalutazione (che è un tipo di commisurazione). È solo sullo sfondo della valutazione che la commisurazione può emergere.

Se i criteri VQR parlano di livelli, è perché il processo di *peer review* prescritto dal bando deve condurre all'attribuzione di “livelli di merito” che si traducono a loro volta in voti espressi numericamente¹¹. Siamo così abituati prima, da giovani, a “subire” voti e poi, per chi lavora nell'Università, a darli, che questo non ci salta all'occhio. Eppure *non* è quello che facciamo nella maggior parte dei casi quando valutiamo. Quando ci piace un quadro, apprezziamo un film, ci piace un piatto, e – anche – leggiamo con ammirazione l'articolo di uno studioso, non attribuiamo voti. La nebbia, il pulviscolo di numeri e loro simulacri che troviamo sempre più spesso (le stellette e le statistiche di gradimento che usiamo per il film, il ristorante, persino per il museo o il monumento, ormai; i “like” per le opinioni) ci fanno dimenticare questa banalità. E distolgono la nostra attenzione dalle condizioni che devono darsi, e si danno implicitamente, quando dall'apprezzamento o dal giudizio di valore (la valutazione vera e propria) si passa alla commisurazione.

Sottolineo che il primo prodursi della commisurazione non ha bisogno necessariamente di complicati apparati di quantificazione, o anche solo di una “metrica comune”, come abbiamo letto nella definizione di Espeland e Stevens. C'è un piano che definirei di “proto-commisurazione” che non necessita di una metrica così intesa, e consiste nell'istituzione di una classificazione ordinale minima, costituita da due classi. Possiamo chiamarlo il *rating*, però non ancora nel significato tecnico-economico moderno, ma facendo riferimento invece al significato del verbo inglese *to rate*, che è insieme *valutare e classificare*¹².

⁹ ESPELAND – STEVENS (1998, 315).

¹⁰ *Ibid.*

¹¹ Si può osservare incidentalmente che il bando parla di un “giudizio di qualità”.

¹² Il sostantivo *rate* significa “quota”, “tasso”, “tariffa”, ha dunque un significato immediatamente quantitativo. *To rate* può voler dire però, in forma intransitiva, anche “valere”, “contare”, “essere importante”. Il che rimanda ad una dimensione ancora precedente alla valutazione-classificazione commisurativa. Cf. *infra*, la n. 22.

Come accennavo, una parte preponderante delle nostre valutazioni, quotidiane e anche professionali, non si traduce in “livelli”, non mette capo a commisurazioni. Mi piacciono Mozart e Stravinsky, Picasso e Schiele, amo Leopardi e Dickinson, la lasagna e la pastiera, trovo affascinanti Heidegger e Dennett, senza istituire per questo commisurazioni tra di loro. La nostra vita valutativa non ne ha *in ogni caso* bisogno. Quando scelgo un ristorante non sto attribuendo stelle e neanche creando un *ranking* mentale, ma sto facendo convergere una serie di obiettivi, gusti, considerazioni, bisogni, orientamenti, e lo sto facendo in un contesto definito, in un certo spazio e in un certo tempo. Lo stesso vale quando scelgo un progetto di ricerca, o decido della pubblicazione di un saggio in una rivista. O quando scelgo il partito per il quale votare. Insomma, in un ampio e relevantissimo ambito delle nostre valutazioni *non* istituamo graduatorie, ma ci affidiamo ad una logica valutativa più complessa, sensibile al contesto, “plurifinale”, se posso dire così, ossia che tiene conto di più finalità allo stesso tempo. Anche quando vi sono più alternative, quando la valutazione è preconditione per una scelta (non avviene sempre)¹³, la logica della scelta non è quella di una classificazione ordinale, di un *rating*.

L’obiettivo di queste semplici considerazioni non è quello di ritornare ad un tema molto discusso e rilevante, quello della (possibile, presunta) incommensurabilità della qualità. Mi interessa piuttosto portare l’attenzione sui presupposti e le precondizioni della commensurabilità o della commisurazione. Data la appena e rozzamente accennata proprietà “diffusa”, non ordinale, delle nostre valutazioni¹⁴,

¹³ L’apprezzamento per un quadro, una canzone, un amico non è collegata comunque ad una scelta. Ne può essere la condizione. In questo senso la questione delle attribuzioni di valore è ovviamente più ampia di quella che si pone in ambito etico, e anche la questione delle scelte non si esaurisce ovviamente nel contesto etico. Estenderei per questo la distinzione tra due «atteggiamenti valutativi costitutivi» – quello che assume qualcosa come commisurabile e quello che la considera fuori di questo rapporto – anche al di là della dicotomia persone/cose, prezzo/dignità, misurazione/rispetto, di origine kantiana, che Bagnoli fa valere in ambito pratico (BAGNOLI 2006, 65-79).

¹⁴ Questa visione naturalmente non è indiscussa. Fa parte anzi di un filone importante della cultura occidentale l’idea di una commensurabilità universale dei valori, che si può far risalire a Platone, il cui presupposto è la possibilità di assumere una *omogeneità qualitativa* che consenta poi di “soppesare” i beni come “minori” o “maggiori”. Questa assunzione di fondo presente in Platone è ricostruita da NUSSBAUM 1984, che mostra come il progetto platonico sia quello di un intervento radicale di *trasformazione* delle nostre credenze ordinarie. Nussbaum non manca di riferirsi a dinamiche e teorie attuali ed esprime la estrema preoccupazione secondo cui «certain proposals in ethics and social choice theory that present themselves as innocuous extensions of ordinary belief and practice could actually lead, followed and lived with severity and rigor, to the end of human life as we currently know it» (p. 56). Sul lato opposto starebbe un modello aristotelico, per il quale la fragilità del bene sarebbe minacciata da una visione che vede appunto il valore come qualcosa di generale ed omogeneo, contro il suo carattere plurale e incommensurabile. Dei limiti alla comparabilità dei valori sono posti anche da Husserl, che prevede solo una sorta di “comparabilità regionale”: «Fin da principio è opportuno dire che non possiamo pretendere che i valori, presi nel senso più ampio del termine e sotto ogni riguardo, possano sempre essere confrontati secondo relazioni di comparazione e che dobbiamo piuttosto limitarci ai valori di una

naturalmente si può sostenere – ha difeso molto bene questa posizione Bonaccorsi nel suo libro *La valutazione possibile* – che in ogni caso è possibile ricondurre giudizi di valore “idiosincratici e taciti” in giudizi discorsivi, intersoggettivamente approcciabili (se non validabili) e infine a scale numeriche. Quali sono le condizioni per farlo?

Anche qui, vorrei prescindere per un attimo dalle complesse procedure istituzionalizzate che conosciamo (in particolare nell’ambito specifico della valutazione della ricerca) per tornare per così dire alle origini epistemiche. In quali condizioni si produce una “proto-commisurazione”? quando emerge un *rating* (o forse possiamo dire un *proto-rating*) nel senso non tecnico prima detto?

Credo si possa dire che l’esigenza e la produzione di classificazioni ordinali – la cui unità minima è di due elementi – emerge quando la nostra normale vita valutativa (quella della *Lebenswelt*, se vogliamo usare un riferimento filosofico autorevole), che non è esaurita nelle attribuzioni di valori relative a scelte, e a scelte “duali”, si trova di fronte ad *un set di scelte* in relazione a *finalità definite*. Se devo scegliere stasera spaghetti o linguine, carne o pesce, non mi muovo istituendo una comparazione di valori con una scala ordinata. Qualcosa di simile ad un *rating*, una classificazione in categorie cui attribuisco un valore relativo ordinato, diventa invece utile e si produce di fatto quando devo orientare una serie di scelte rispetto ad *un fine prevalente e costante*. Se il fine è la salute, può crearsi una scala ordinata tra pesce e carne, o tra carni rosse, bianche e pesce, ecc. Questo non vuol dire che nella valutazione quotidiana non vi siano in gioco più valori e un peso maggiore di alcuni su altri, ma il ragionamento che qui ha luogo – non necessariamente implicito, anche molto esplicito (anche se la mera ragione valutante, nel suo strato originario, come scriveva Husserl, «non vede, non comprende, non esplicita, non predica») ¹⁵ – non consiste nello stabilire delle scale, perché di esso fa parte anche il variare e il confrontarsi a loro volta di punti di vista e di finalità diverse e compresenti. Nella scelta se lasciare o meno un fidanzato per uno nuovo è solo una caricatura l’uso di una lista comparativa di qualità su un foglietto, e non solo perché vi è l’elemento sentimentale imponderabile, le ragioni del cuore. Se, per prendere un esempio più vicino alla nostra tematica, chi dirige una rivista compara tra loro due articoli è perché ha, ad esempio, l’alternativa circa quale pubblicare prima o in quale

regione del valore (o se preferite di una categoria di valore)» (HUSSERL 2002, 106). Una posizione differente è quella di Max Scheler, che in generale ritiene che i valori siano «in quanto pure qualità incommensurabili» e una «quantificazione dei valori» sia sempre «artificiale»; è possibile tuttavia per Scheler una «indiretta misurazione del valore», ma solo (analogamente a Husserl) all’interno della stessa *Wertmodalität* («sistema di qualità dei valori materiali»). Tra queste «modalità di valore» sussiste in ogni caso un «ordinamento di rango» che è un aspetto essenziale dei valori in quanto tali e che Scheler considera invariabile, mentre al suo interno sono storicamente variabili le «regole di preferenza» (*Vorzugregeln*). Cf. SCHELER (1916, 84 ss., in particolare pp. 86, 103). Per una posizione contemporanea che considera il problema della comparabilità tra *tipi* diversi di valore cf. NAGEL 1986.

¹⁵ HUSSERL (2002, 86).

posizione, ma normalmente – «innanzitutto e per lo più» – non fa considerazioni comparative.

3. Costi epistemologici della commisurazione

Se la proto-commisurazione è un processo che si svolge su poche cose e anche – ma non prevalentemente – implicito, la commisurazione prevede invece, oltre alla «metrica comune», procedure complesse, scale ordinali articolate, metodologie di formazione di consenso e di controllo. Ma in ogni caso la commisurazione acquista un senso e soprattutto seleziona i suoi strumenti in base alle finalità che persegue. È questo il punto che mi interessa mettere particolarmente in rilievo. Più che il tema – rilevante e serio – del fatto che alla misurazione sfugge qualcosa di incommensurabile (è quasi tautologico, ma il punto naturalmente è *cosa e quanto* sfugga), credo sia importante rilevare che una commisurazione ha la sua ragione non nella conoscenza del suo oggetto (anche se produce una qualche forma di conoscenza), ma nella tipologia di scelte che deve aiutare a produrre. Dunque trattarla come se fosse in gioco solo la sua adeguatezza rispetto all’oggetto – la capacità di istituire classificazioni “oggettive” – può essere profondamente fuorviante. Una commisurazione comporta una procedura che è sempre anche una scelta di *fini*, e non solo di “oggetti”¹⁶. Una visione “tecnica” della valutazione/commisurazione concentrata sugli strumenti manca fin dall’inizio di rendere visibile ciò che conta. I costi epistemici – non quelli finanziari – di una valutazione così intesa rischiano di sfuggire¹⁷.

Si possono riassumere alcuni di questi costi nel modo seguente.

1) Il primo è *l’introduzione in molti casi di quelli che possiamo chiamare “operatori di decisione”*. Vi sono fenomeni che si prestano di più e altri di meno ad una commisurazione. Per stare all’esempio di un economista, Mathias Binswanger, che ha parlato di *Messbarkeitsillusion*¹⁸, i cento metri di corsa si prestano più facilmente del pattinaggio su ghiaccio o dei tuffi ad una commisurazione. Ma credo si possa

¹⁶ La commisurabilità, dove è possibile, non è un dato o una pre-condizione, ma il risultato di un processo (cf. MILLGRAM 1997, 151 ss.), su cui si innesta poi il processo di commisurazione. Per arrivare a commisurare è necessaria una indagine preliminare rispetto a ciò in base a cui intendiamo o vogliamo o dobbiamo commisurare, e dunque una considerazione e, se possibile, una scelta di fini.

¹⁷ Questo non significa riferirsi ad un «pre-given order of things» che costituirebbe una vita “autentica”: le modalità “diffuse” di valutazione sono senz’altro culturalmente condizionate e trasformabili. Questa obiezione all’argomento della valutazione come antitetica da una “vita autentica” è considerato da DAHLER-LARSEN 2015. È piuttosto la “cultura della valutazione” ad appellarsi spesso ad una naturalità ed inevitabilità della (neo)valutazione, identificandola con la valutazione *tout court*.

¹⁸ BINSWANGER 2010. Di particolare interesse è l’analisi degli effetti controfinali del tentativo di misurare prestazioni individuali in sport cooperativi come il calcio.

aggiungere che vi sono anche fenomeni per i quali la commisurabilità offre margini di possibilità piuttosto stretti e deve essere costruita – se è indispensabile – in modo particolarmente artificiale. Dove vi è necessità di istituire agonismo, la possibilità di decidere chi è il migliore è frutto di una costruzione, non necessariamente arbitraria, ma che deve inserire artificialmente dei fattori di decisione netti. È il caso di dire che questi fanno parte del *gioco*, e non pretendono di offrire un reale contributo “conoscitivo”. Anche se mantengono un nesso con qualche qualità. Si pensi al caso di una partita di calcio, e al fatto che capita non di rado che una squadra abbia giocato, a giudizio di tutti i commentatori, meglio di un'altra, e tuttavia abbia perso. Naturalmente anche nello sport, in alcuni tipi di competizione (è il caso dei “campionati”), l'aggregazione di numerosi risultati può ridurre la discrepanza tra la qualità e la commisurazione. Ma, ripeto, il punto che mi interessa in prima istanza non è il possibile *deficit* di oggettività e la sua rilevanza, ma il legame tra la commisurazione e la sua finalità. La commisurazione per sua natura – proprio perché legata a finalità non puramente conoscitive, ma sempre e originariamente a *decisioni*, deve introdurre limiti netti anche o soprattutto dove la natura della cosa non ne presenta, e presenta piuttosto contorni sfumati e sfocati. Questo distingue valutazione e commisurazione in modo radicale, e non solo per gradi o per tecniche applicate¹⁹.

2) Il secondo punto è quella che chiamerei una *dialettica della commisurazione*, nel senso kantiano del termine, per cui dialettica è quella logica che produce una illusione inevitabile (nel nostro caso direi solo parzialmente inevitabile). La commisurazione, che può essere pienamente razionale in relazione ad una finalità definita, produce dei risultati che per così dire si autonomizzano, vengono assunti come dotati di senso anche di fuori del contesto in cui ne avevano, e producono ulteriori commisurazioni, che costituiscono una sorta di escrescenze valutative. Ha luogo una sorta di produzione di commisurazioni per mezzo di commisurazioni, esplicita e consapevole in alcuni casi, implicita in altri, che, anche in questo caso, non è la semplice estensione di una procedura, ma una sua radicale trasformazione.

¹⁹ Va osservato incidentalmente che proprio negli ambiti in cui la complessità di ciò che è da valutare e le qualità delle prestazioni salgono, le possibilità di commisurazione si attenuano. Chi segue i *talent show* culinari sa che è facile distinguere tra un cuoco che brucia un filetto e uno che non lo fa, ma diventa problematico giudicare comparativamente interi menù di più portate presentati da due o più cuochi molto capaci. Quando per qualche ragione si deve/vuole farlo, si introducono fattori arbitrari o che misurano abilità effettive (ad esempio: la velocità di sfilettatura di una cernia), ma che finiscono per avere un legame fortemente allentato – o nessun legame – con la qualità che in primo luogo interessa. Il cuoco più veloce non è necessariamente il cuoco migliore, se interessa la bontà del piatto. Se interessa la complessiva “abilità” del cuoco, allora evidentemente il maggior peso di una abilità o un'altra è stabilito arbitrariamente, o quanto meno in modo non univoco e intrinseco. Una commisurazione che persegue grane sempre più fini (e dunque meno percepibili) di elementi da considerare paradossalmente spesso si allontana – invece di avvicinarvisi – da ciò che è il fine originario della valutazione, il “valore” che è originariamente in questione.

3) Terzo, ma forse primo in ordine di importanza, è il fatto che l'istituzione di un *rating*, di una valutazione/commisurazione, se è funzionale e indispensabile nel caso di *scelte* e in relazione ad una singola scelta (se ho due dentisti disponibili di sabato scelgo il migliore dei due), è discutibile che sia altrettanto indispensabile e soprattutto funzionale *in relazione alla conoscenza di realtà complesse*. Se ho un *rating* o un *ranking* dei dentisti della città, è discutibile che la mia conoscenza della realtà della medicina odontoiatrica di Pisa possa essere ritenuta più approfondita rispetto a quella di chi ne ha fatto uno studio non commisurativo²⁰. La valutazione/commisurazione è, insomma, la costruzione di un complesso gioco sociale, dove il fine non è però quello di divertirsi. Non è la prosecuzione con altri mezzi e su grande scala della stessa cosa che facciamo valutando nella nostra vita sociale, e in particolare in quella scientifica: è *un altro gioco*, ha – direbbe Wittgenstein – una “grammatica” del tutto differente. La retorica della “cultura della valutazione” tende sistematicamente ad occultare questa differenza.

Sulla base di queste osservazioni, si possono fare alcuni rilievi sulla difesa della commensurabilità che Bonaccorsi – appoggiandosi alle tesi di Ruth Chang – propone, in vista di una difesa della valutazione commisurativa. Non mi sembra essere decisivo l'argomento (che forse è affrettato definire come un «risultato») per cui sarebbe falso che (cito Bonaccorsi che ricostruisce Chang) «date due alternative A e B, non esiste nessuna relazione di valore in base alla quale si potrebbero comparare»²¹. È probabilmente sempre possibile istituire un gioco valutativo che renda possibile una comparazione; ma questo dipende costitutivamente dalla fissazione di una finalità cui fa riferimento una scelta (azioni che non perseguono un fine sono arbitrarie e in quanto tali sfuggono ad una logica cognitiva): non corrisponde però alla logica prevalente e diffusa delle nostre valutazioni e non dà necessariamente un contributo di conoscenza apprezzabile, o comune maggiore rispetto alle nostre valutazioni plurifinali e plurivaloriali. Il problema non è la comparabilità, sono i costi cognitivi della commisurazione²².

²⁰ Anche se questo *ranking* è il risultato di una aggregazione di un ampio numero di giudizi individuali. L'aggregazione di un ampio numero di giudizi non è un processo cognitivo superiore a quello svolto da un singolo individuo – è semplicemente una rilevazione di opinioni, che naturalmente può avere un significato, in relazione anzitutto ai processi di formazione di *opinioni*, ma che non ha alcun titolo per pretendere una superiorità conoscitiva rispetto all'*oggetto* delle opinioni in gioco.

²¹ BONACCORSI (2015, 80). Il saggio di Chang cui Bonaccorsi si riferisce è CHANG 1997a.

²² Non credo che sia decisiva una distinzione tra commisurazione (come comparazione in relazione ad una scala di misura) e comparazione. Come accennavo in precedenza, il “rating” minimo che istituisce appunto una priorità tra A e B istituisce una differenza di valori in base ad un criterio (se non arbitraria), e questa è già una scala ordinale. Se «la commensurabilità non è una condizione necessaria per la creazione di ordinamenti» e «ciò che è invece assolutamente necessario è preservare la comparabilità»

Le procedure poi di aggregazione di giudizi – tali da trasformare non tanto «giudizi qualitativi in ordinamenti»²³, ma giudizi individuali già *commisurativi* in ordinamenti²⁴ – possono avere maggiore o minore razionalità in rapporto al modo in cui sono in grado di render conto dei giudizi individuali che ne costituiscono la base, ma questa razionalità non è riferita ad un più adeguato approccio cognitivo-valutativo all’oggetto. La questione, per dirlo in modo chiaro, *della verità* della valutazione resta del tutto sullo sfondo: in questione è la affidabilità delle procedure rispetto ad una *rilevazione di opinioni*, importante per indagini di mercato e per processi di decisione collettiva, ma tutta interna alla *doxa* e non all’*episteme*. La tensione ad un giudizio valutativo il più possibile adeguato alla “cosa stessa” è del tutto assente. La polarità tra *affidabilità* e *verità* (che si pone per classificazioni scientifiche, come quelle dei disturbi psichiatrici) è qui del tutto spostata verso la sola affidabilità²⁵. Come sottolineano Espeland e Stevens, «la commisurazione può essere intesa come un sistema per scartare informazione e organizzare quanto rimane in nuove forme. Astraendo e riducendo informazione, il legame tra ciò che è rappresentato e il mondo empirico è oscurato e l’incertezza assorbita»²⁶. Nel quantificare gli esiti di giudizi per poterli aggregare ci si allontana da elementi essenziali dell’atteggiamento valutativo: «L’esperienza quotidiana, il ragionamento pratico e l’identificazione empatica divengono basi per il giudizio sempre più irrilevanti in quanto il contesto è strappato via e le relazioni vengono rappresentate in modo più astratto da numeri»²⁷. La distruzione del *contesto* è il primo passo per la neutralizzazione della *comprensione*.

(BONACCORSI 2015, 79), dalla creazione di ordinamenti scaturisce immediatamente una commisurazione: anche se privo di punteggi, un ordinamento è subito traducibile (seppure con minore precisione di una classifica con punteggi basati su una scala di misura) in misure di valore. È curioso che Bonaccorsi si appoggi su Bagnoli per sostenere che non sarebbe «necessario richiedere la commensurabilità come condizione per generare degli ordinamenti», in quanto avrebbe mostrato che «non solo la commensurabilità non è necessaria, ma neppure sufficiente a dare determinazione alla scelta» (BONACCORSI 2015, 80, n. 15). Bagnoli, introducendo il concetto di “importanza” come alternativo alla commensurabilità sottolinea, mi sembra, che essa *non conduce* ad un ordinamento dei nostri beni e fini, insistendo anche sulla necessità di «mantenere aperte le relazioni di importanza», che possono generare una predominanza – il prevalere di un valore – solo «contestuale» (cf. BAGNOLI 2006, 75).

²³ BONACCORSI (2015, 78).

²⁴ La procedura di Baliski e Laraki che Bonaccorsi riprende (BONACCORSI 2015, 81 ss.) prevede giudizi espressi in un ordine di valore (“eccellente”, “molto buono”, etc.) che prescindono dall’istituzione preliminare di una scala di valori *tra gli oggetti comparati*, il che consentirebbe di evitare i noti paradossi (Condorcet, Arrow) di aggregazione di scelte di preferenza su più elementi. Ma è evidente che non si tratti per questo né di giudizi “qualitativi” né di un giudizio «assoluto, non relativo» (p. 83).

²⁵ Allen Frances, curatore del DSM-IV, sottolinea una tale polarità tra “affidabilità” e “validità” (dove la prima è relativa al grado di consenso conseguibile in una comunità scientifica) in relazione alle classificazioni dei disturbi in psichiatria, che da «disciplina delle sfumature» si sarebbe trasformata in pratica «standardizzata e semplificata» (FRANCES 2013).

²⁶ ESPELAND – STEVENS (1998, 317).

²⁷ *Ibid.*

Anche se si assume – dimenticando questa distanza tra *doxa* ed *episteme* – che il termine di riferimento debba essere il consenso e non la verità, permane il problema che la valutazione/commisurazione non esprime le forme di apprezzamento e di consenso reali di una comunità scientifica. L'idea di fondo di Bonaccorsi è che «la valutazione è un esercizio di esplicitazione, formalizzazione e aggregazione di giudizi già presenti nelle comunità competenti»²⁸. La valutazione/commisurazione non è a mio parere affatto una “esplicitazione” di graduatorie «latenti nella coscienza»²⁹, ma, come dicevo, un gioco con una grammatica radicalmente diversa, che deve giustificarsi di volta in volta in relazione alle sue finalità. La reale pratica di valutazione presente nelle comunità scientifiche è più simile a quella che si svolge nelle redazioni delle riviste o nelle case editrici: quella finalizzata a giudicare (nell'ambito e nelle forme delle diverse discipline e delle loro diverse metodologie e orientamenti) se un contributo di ricerca è effettivamente tale, se contribuisce all'accrescimento del sapere nelle più diverse forme (e dunque risulta pubblicabile). Questa valutazione comporta valutazione in relazione agli obiettivi conoscitivi, argomentazione nel merito, modifica migliorativa, interazione “scientifica” autentica. Non sfocia in commisurazione né *in ranking*.

4. Per esempio

Perché il discorso non resti solo sul piano teorico, provo a proporre infine qualche esemplificazione riguardo a ciò che è concretamente in gioco in particolare nell'ambito della ricerca “umanistica”.

In generale, prima di entrare nello specifico, va ricordato che l'impianto della valutazione della ricerca in Italia è strettamente e, direi, originariamente finalizzato ad un meccanismo automatico di assegnazione di risorse. In questione è dunque il senso che assume una commisurazione come quella delle VQR. La produzione di indici numerici è finalizzata ad una quantificazione di fondi attribuiti a strutture che prescindono da considerazioni di politica della ricerca di cui un decisore si assuma di volta in volta la responsabilità. Non discuto di questa connessione in questa sede, che pure non è affatto necessaria, né indiscutibile, né indiscussa, come ha ricordato di recente Viesti³⁰. Questa finalità si estende ad altre forme di decisione meccanizzata, se così vogliamo chiamarla, come gli indicatori di soglia per la Abilitazione Scientifica Nazionale. Questo fine a sua volta rende necessari dei mezzi. In questo tipo di gioco valutativo, i

²⁸ BONACCORSI (2015, 89).

²⁹ Usa molto appropriatamente questa espressione BANFI (2011, 12).

³⁰ «In Italia i processi di valutazione non sono un esercizio i cui esiti vengono sottoposti a governo e Parlamento per le scelte più opportune: essi producono effetti diretti, automatici, sul sistema» (VIESTI 2018, 93).

mezzi non sono neutrali, perché dipendono dal fine reale o prevalente da raggiungere, che non è la valutazione della qualità, ma la decisione meccanizzata.

In Italia abbiamo due esempi clamorosi di prevalenza dei mezzi sul fine apparente (valutazione della qualità), in favore del fine reale o prevalente (decisione meccanizzata). Uno è la decisione, presa in occasione della protesta realizzata da una parte dei docenti italiani tramite astensione dalla VQR, di considerare la mancata presentazione di prodotti, dovuta a ragioni appunto “sindacali” di protesta, come equivalente ad una valutazione negativa. Credo che sia una macchia sulla valutazione della ricerca italiana e sulla politica universitaria italiana il cui significato negativo è difficilmente sottovalutabile e che resterà nel suo libro nero. Ha avuto naturalmente ragioni anche politiche, e purtroppo la connivenza di molti rettori, ma mostra nel modo più evidente come la finalità di consentire decisioni meccaniche prevalga su ogni considerazione di obiettività cognitiva. Dati *consapevolmente falsi* rispetto alle loro finalità valutative vengono utilizzati per valutare, perché il fine di *questo* gioco valutativo prevale. Un esempio analogo è la circostanza, per certi versi anche più discutibile, perché priva di ragioni “politiche”, della esclusione delle monografie dai dati per la valutazione dei componenti dei collegi dei docenti dei dottorati. L’utilizzo del riferimento alle sole pubblicazioni in riviste di fascia A con esclusione delle monografie – quanto meno per un indice molto rilevante³¹ – non ha alcuna immaginabile e presentabile razionalità rispetto alla valutazione della qualità di un lavoro di ricerca, mentre è finalizzata alla sola meccanizzazione, consentita solo

³¹ Durante il convegno in cui è stato presentato questo contributo Maria Luisa Meneghetti ha contestato questa affermazione, nella forma generale in cui l’avevo presentata («esclusione delle monografie dai dati per la valutazione» dei collegi dei dottorati), sottolineando come la valutazione dei collegi di dottorato preveda anche un riferimento a indicatori della VQR, che comprendono anche la valutazione delle monografie. A questo proposito si può osservare quanto segue. 1) Gli indicatori della VQR (di cui alla complessa condizione n. 1 prevista dalla Linee guida del MIUR del 14/4/2017) sono basati sui “prodotti” presentati, i quali costituiscono per lo più solo una minima parte della complessiva produzione scientifica di un ricercatore; questi, anche se ha pubblicato monografie, può scegliere di presentare articoli – in base a considerazioni relative alle finalità della VQR, ad esempio in omaggio alla diffusa tendenza a prediligere per pubblicazioni in lingua straniera o articoli su riviste di fascia A. Per l’accreditamento dei dottorati i componenti del collegio non hanno la possibilità di indicare pubblicazioni diverse, ma vengono appunto considerati i punteggi dei prodotti presentati per la VQR. Le monografie possono dunque restare del tutto fuori da tali indicatori. 2) La condizione n. 3 che le *Linee guida* prevedono («Indicatore quantitativo di attività scientifica») prescrive che «Tutti i componenti [*sic*] del collegio devono aver pubblicato [...] nei settori non bibliometrici, negli ultimi dieci anni un numero di articoli in riviste di classe A almeno pari alla soglia prefissata per i professori associati nel proprio settore concorsuale». Questo significa che per l’«indicatore quantitativo di attività scientifica» le monografie non contano affatto (come del resto, curiosamente per un indicatore “quantitativo”, saggi pubblicati in sedi diverse da riviste di fascia A); chi avesse prodotto una o più monografie, ma non articoli in riviste di classe A, comprometterebbe dunque solo per questo una delle quattro condizioni previste per la «qualificazione del collegio dei docenti» (di cui almeno tre devono verificarsi). Le monografie restano clamorosamente sottovalutate.

dall'indicatore relativo ai lavori su rivista di fascia "A", in quanto non esistono – per fortuna – "classi" di monografie.

Questo tipo di osservazioni sono legate al fattore che ho indicato come la creazione di "operatori di decisione". Circa il secondo aspetto, che ho chiamato "dialettica della commisurazione", ossia il crearsi di apparenti "dati" valutativi utilizzabili come fatti, slegati dal loro contesto di origine e dalla loro logica, gli esempi potrebbero essere molti. Uno potrebbe essere l'uso di valutazioni che hanno la finalità di valutare strutture (possono aver senso in *quel* gioco valutativo) per la valutazione di individui. Si può sostenere che si tratti di semplici storture, anche perché l'ANVUR stessa ha messo in guardia da questo uso. Questo non ha impedito a molti solerti rettori e amministratori di far riferimento in più casi ai dati VQR fuori dal loro ambito, di richiederli esplicitamente a singoli ricercatori, e non ha impedito neanche all'ANVUR l'applicazione di dati quantitativi nell'accREDITAMENTO dei dottorati, che di fatto diventa una valutazione di individui, perché ha conseguenze sulla scelta dei componenti del collegio dei docenti. L'illusione di trovarsi di fronte a dati svincolabili dal loro contesto di produzione, "oggettivi", è stata irresistibile in moltissimi campi. Un diverso esempio può essere quello della classificazione delle riviste ERIH³². Nata in modo ambiguo, con una classificazione che si presentava come inevitabilmente anche valutativa, si è sviluppata – in seguito a diffuse critiche – come classificazione che esplicitamente dichiarava di non voler costituire un *rating*, ma solo uno strumento di informazione. Ciononostante è stato spesso presa, a dispetto di tutto, come punto di riferimento per questioni di classificazione in classi di merito delle riviste (anche da parte di GEV dell'ANVUR), fino alla sua ultima e attuale trasformazione in ERIH PLUS, nella quale si è infine consapevolmente rinunciato ad *ogni classificazione*, limitandosi ad un *bibliographic approach* con finalità informative.

Un intreccio tra i tre "costi" cognitivi che abbiamo ricordato – l'inadeguatezza di un *rating* per realtà complesse, l'intervento di operatori di decisione artificiali (e potenzialmente dannosi) e la creazione di apparenze illusorie di oggettività – è ben esemplificato dalla classificazione italiana delle riviste in classi³³. È probabilmente un meccanismo sfuggito di mano anche a chi si è trovato ad idearlo, se, in relazione al suo uso per la VQR, si sosteneva: «va ribadito ancora una volta che non si crea nessun automatismo tra qualità del singolo articolo e qualità della rivista, o come si dice, tra contenuto e contenitore. Gli articoli su riviste italiane vanno in *peer review*, e vengono quindi valutati per il loro valore intrinseco, indipendentemente dal rating della rivista su

³² *European Reference Index for the Humanities*, prima curato dall'*European Science Foundation* e poi dal 2014 dai *Norwegian Social Science Data Services*.

³³ Su questa e sul suo carattere «non solo inutile, [...] ma anche pernicioso», ha detto parole molto chiare BANFI 2011.

cui sono contenuti»³⁴. Se questo era parzialmente vero per la VQR³⁵, l'uso delle classi di riviste per le "soglie" (prima per le "mediane") dell'Abilitazione Scientifica Nazionale, così come per l'accreditamento dei dottorati, è evidentemente un trasferimento automatico dalla presunta qualità del contenitore a quella del contenuto. L'idea che l'introduzione di un rating con pesanti conseguenze su gruppi, individui, dinamiche culturali intere, processi editoriali, reclutamento accademico, possa essere in qualche modo l'espressione – l'"esplicitazione" – di modi di attribuzione di valore (e direi di senso) di comunità scientifiche di ricerca, una «formalizzazione e aggregazione di giudizi già presenti nelle comunità competenti»³⁶ (non so se Bonaccorsi applicherebbe oggi questa definizione anche a *questa* pratica valutativa) offusca completamente differenze abissali tra le valutazioni presenti nelle comunità, in cui si intrecciano una pluralità di prospettive, punti di vista, finalità, orientamenti culturali, e un gioco commisurativo artificiale che ha l'aggravante di essere gestito da un organismo collegato al potere politico. È un gioco che, con quanti sforzi si faccia per trovare criteri, finisce per costituire una istanza valutativa dall'alto, che non può avere nessun rapporto con le procedure e lo spirito "mertoniano" della ricerca, in cui conta la libera interazione tra gli attori di un gioco ben diverso, con un solo fine chiaro, quello dell'acquisizione di conoscenza. Se c'è un modello che può servire a indicare la differenza tra i giochi commisurativi e la dinamica delle valutazioni che si svolgono nelle comunità scientifiche è forse quello illustre, ma ancora attuale, offerto dalla distinzione che Immanuel Kant faceva tra uso pubblico e uso privato della ragione. Mentre l'uso privato era quello di funzionari, al servizio di finalità predefinite – anche di uno Stato, a dispetto del termine "privato" – quello *pubblico* era per Kant l'uso della ragione rivolto al pubblico dei lettori (oggi diremmo alla comunità scientifica) e sottoposto al solo vincolo della ricerca della verità. A questo gioco più grande – che un ministro britannico dell'istruzione, del Labour Party, Charles Clarke, chiamò sprezzantemente «la medievale ricerca della verità»³⁷ – le tecniche di commisurazione dovrebbero cercare di non porre ostacoli.

³⁴ BONACCORSI (2012, 95).

³⁵ Che si presentava nelle aree non bibliometriche come *informed peer review*, dove l'elemento di informazione è dato dalla classificazione delle riviste.

³⁶ BONACCORSI (2015, 89).

³⁷ «I argue that what I described as the medieval concept of a community of scholars seeking truth is not in itself a justification for the state to put money into that. We might do it at, say, a level of a hundredth of what we do now and have one university of medieval seekers after truth [...] as an adornment to our society» (<https://www.theguardian.com/uk/2003/may/10/highereducation.politics>).

Cf. anche "Independent", May 17, 2003: «The "medieval concept" of the university as a community of scholars is only a very limited justification for the state to fund the apparatus of universities». Alla distinzione kantiana ha fatto riferimento Slavoj Žižek in riferimento alla «riduzione del compito dell'istruzione universitaria alla produzione di un sapere competente e utile», come «forma paradigmatica dell'"uso privato della ragione" nel capitalismo globale contemporaneo» (ŽIŽEK 2012, 23).

È pensabile una valutazione della ricerca che faccia uso di commisurazioni in campi specifici, limitati, dove dati quantitativi aggregati possano avere un senso *in quanto dati*, gli indicatori in quanto indicatori³⁸, senza sostituirsi a decisioni e neanche a valutazioni? Che lasci spazio a pratiche di valutazione e autovalutazione effettivamente funzionali al miglioramento collettivo della ricerca – e dunque pensate e sviluppate per questo, e non per fini diversi? Che escano fuori dal grande gioco della commisurazione universale che le tecniche di valutazione della ricerca contemporanee, delle neo-valutazione cercano di realizzare? Se fosse davvero possibile «un disegno istituzionale della valutazione di tipo fallibilista e aperto, non burocratico, non tecnocratico»³⁹, queste domande dovrebbero interessare anche le istituzioni valutative. In attesa, credo che il compito delle riflessioni sul fenomeno della valutazione sia particolarmente pregnante per chi lavora nelle discipline “umanistiche”, nelle quali la sensibilità per la pluralità e differenza degli apprezzamenti, dei valori, dei fini è in qualche modo costitutiva, e dunque ritengo particolarmente opportuno che siano queste aree di ricerca a farsene carico in prima persona: «the social, political and philosophical story-telling about evaluation should not be left to evaluators»⁴⁰.

³⁸ Per l'essenziale precisazione che gli indicatori di performance *non* sono dati, ma «differiscono dalle consuete informazioni statistiche sullo stato di organizzazioni e servizi innanzitutto per il fatto che implicano il riferimento ad un obiettivo (uno standard, un *benchmark*, una tendenza etc.)», così come per altre importanti precisazioni sul tema, cf. VERGANI 2018.

³⁹ BONACCORSI (2015, 92).

⁴⁰ DAHLER-LARSEN (2015, 21).

Riferimenti bibliografici:

BAELHAUSER – GORI – SAURET 2011

A. Baelhauser, R. Gori, M-J. Sauret, *La folie Évaluation. Les nouvelles fabriques de la servitude*, Paris.

BAGNOLI 2006

C. Bagnoli, *Deliberare, comparare, misurare*, «Ragion pratica» XXVI, 65-80.

BANFI 2011

A. Banfi, *Il resistibile fascino delle classifiche di riviste*, «Rassegna Italiana di Valutazione» XV, 51, 9-21.

BINSWANGER 2010

M. Binswanger, *Sinnlose Wettbewerbe. Warum wir immer mehr Unsinn produzieren*, Freiburg i.B.

BONACCORSI 2012

A. Bonaccorsi, *Ancora sulla valutazione nelle aree umanistiche e sociali*, in P. Miccoli, A. Fabris (a cura di), *Valutare la ricerca? Capire, applicare, difendersi*, Pisa, 85-98.

BONACCORSI 2015

A. Bonaccorsi, *La valutazione possibile*, Bologna.

CHANG 1997

R. Chang (ed.), *Incommensurability, Incomparability and Practical Reasoning*, Cambridge (Mass.)-London.

CHANG 1997a

R. Chang, *Introduction*, in CHANG 1997, 1-34.

DAHLER-LARSEN 2015

P. Dahler-Larsen, *The Evaluation Society: Critique, Contestability and Skepticism*, «SpazioFilosofico» I, 21-36.

ESPELAND – STEVENS 1998

W.N. Espeland, M.L. Stevens, *Commensuration as a Social Process*, «Annual Review of Sociology» XXIV, 313-43.

FRANCES 2013

A. Frances, *Primo, non curare chi è normale*, Torino.

HUSSERL 2002

E. Husserl, *Lineamenti di etica formale*, a cura di P. Basso e P. Spinicci, Firenze.

HUMBOLDT 1809/10

W. v. Humboldt, *Über die innere und äussere Organisation der höheren wissenschaftlichen Anstalten in Berlin*, 1809/10 (<https://edoc.hu-berlin.de/handle/18452/5305>).

KANT 1924

Kant's gesammelte Schriften, hrsg. von der Königlich Preußischen Akademie der Wissenschaften, Bd. XVI, Berlin-Leipzig.

LA ROCCA 2013

C. La Rocca, *Commisurare la ricerca. Piccola teleologia della neo-valutazione*, «aut aut» CCCLX, ottobre-dicembre 2013, 69-108.

MILLGRAM 1997

E. Millgram, *Incommensurability and Practical Reasoning*, in CHANG 1997, 151-69.

NAGEL 1986

T. Nagel, *La frammentazione del valore*, in *Questioni mortali*, Milano, 127-39.

NUSSBAUM 1984

M. Nussbaum, *Plato on Commensurability and Desire*, «Proceedings of the Aristotelian Society» LVIII, 55-79.

NUSSBAUM 2001²

M. Nussbaum, *The Fragility of Goodness*, Cambridge.

PINTO 2012

V. Pinto, *Larvatus prodeo*, «Iride» XXV, 475-91.

PINTO 2018

V. Pinto, *Cultura della valutazione e governamentalizzazione della conoscenza*, in TURCO 2018, II, 91-107.

SCHELER 1916

M. Scheler, *Der Formalismus in der Ethik und die materiale Wertethik*, Halle.

TURCO 2018

A. Turco (a cura di), *Culture della valutazione*, Roma.

VERGANI 2018

G. Vergani, *La valutazione tra ricerca, didattica e indicatori: uno sguardo dall'interno*, in TURCO 2018, I, 69-84.

VIESTI 2018

G. Viesti, *La laurea negata. La politiche contro l'istruzione universitaria*, Roma.

ŽIŽEK 2012

S. Žižek, *Benvenuti in tempi interessanti*, Firenze.