

Digital culture, Turing Test and epistemological criteria of critical thinking

Alfonso Di Prospero
(Università degli Studi di Chieti)
diprospero2017@gmail.com

Articolo sottoposto a *double blind peer review*

Title: Digital culture, Turing Test and epistemological criteria of critical thinking

Abstract: The Turing Test can be interpreted in behaviouristic terms, but the behaviouristic psychology entails assumptions on the epistemological meaning of intersubjectivity that are disputable. My hypothesis is that: 1) the analysis of intersubjectivity should be connected to a non-monotonic logic, 2) Piaget's genetic epistemology could be used with this goal, 3) the troubles that in these years we deal with in relation with the explosion of digital communication should be seen within a theoretical framework where the notion itself of intersubjectivity should be re-defined. Also the notion of intersubjectivity in Nietzsche's philosophy is considered.

Keywords: Intersubjectivity, Turing's test, Non-monotonic logic, Critical thought, Induction.

1. *Modernità, comunicazione e critica*

Per tentare di reperire indizi che servano a capire come si stia configurando la condizione dell'esistenza umana nell'epoca attuale, in cui le tecnologie del digitale impongono in maniera sempre più massiccia la loro influenza, può essere naturale rivolgere l'attenzione ad alcune idee di uno tra gli scienziati che hanno contribuito in modo più decisivo al percorso che ha portato al loro sviluppo: nel suo celebre "test", Alan Turing¹ ha cercato di condensare e venire a capo di una serie di problemi quanto mai impegnativi, che in questo testo sosterrò essere significativi anche per l'analisi dei processi socio-culturali oggi in corso.

Gli obiettivi di fondo di questo articolo – in quanto relativi all'analisi di fenomeni di tipo storico e sociale – rendono opportuno che si esplicitino alcune assunzioni di base. Quelli che qui si indagheranno sono gli effetti che può avere il condizionamento sociale sul pensiero delle persone, anche – più nello specifico – sul piano della codificazione dei procedimenti di indagine intellettuale.

¹ A. Turing, *Computing Machinery and Intelligence*, in: «Mind» 59/236 (1950), pp. 433-460.

A monte, autori di riferimento sono Émile Durkheim², Karl Mannheim³, Peter Berger e Thomas Luckmann⁴. David Bloor e la “scuola di Edimburgo”⁵ hanno provato a radicalizzare le idee di Mannheim, cercando di mostrare per esempio che anche le conoscenze della matematica possono essere studiate da un punto di vista sociologico. Una delle premesse della mia argomentazione è che sia possibile sostenere che anche una caratteristica fondamentale del metodo scientifico, ossia il suo costituirsi come insieme di saperi e di metodi intersoggettivamente condivisibili⁶, può essere analizzata come derivata socialmente. La mia ipotesi fa appello all’epistemologia genetica di Jean Piaget, re-interpretata in termini induttivistici. Il bambino molto piccolo non possiede ancora il senso della reciprocità razionale, che svilupperà solo progressivamente: nella mia ricostruzione, questo avviene appunto grazie a esperienze di rapporto sociale che consolidano nell’individuo l’attesa che un criterio per parlare di vero o falso debba essere quello dell’accordo sociale⁷. All’interno di questo quadro, come caso particolare, si può arrivare a costruire la nozione (molto specifica) di intersoggettività che sottende la pratica della ricerca scientifica, qualora si operi in presenza di bisogni e di esperienze che spingano concretamente in questa direzione. Questo significa che un sapere e delle esperienze che sono individuali “fondano” il sapere “pubblico” della scienza (oltre che in genere quello del “senso comune” condiviso), e non è – viceversa – il sapere scientifico, in quanto pubblico, a detenere il diritto di trattare i contenuti dell’esperienza individuale come un materiale da ridursi ai codici, le procedure e le prassi della ricerca scientifica (si ricordi qui soprattutto la *Crisi delle scienze europee* di Husserl).

Porre in apertura del testo, come premessa di quanto seguirà, questa tesi, può sembrare una grave ammissione di debolezza: si tratta infatti di una concezione molto impegnativa, che non potendo venire qui ulteriormente discussa per ovvie ragioni di spazio⁸, può fornire motivi per pensare che la costruzione argomentativa che segue possa essere sostanzialmente arbitraria. Si può però anche intendere in modo più prudente questa scelta. Da una parte in generale anticipare subito alcuni degli esiti del ragionamento può rendere tutto il discorso più chiaro. Si deve chiedere in sostanza al lettore la pazienza di sospendere il giudizio sulla plausibilità della tesi ora esposta: al termine dell’articolo, riconnettendo con calma i vari passaggi dell’esposizione, sarà più semplice esprimere una valutazione

² *De la division du travail social*, Alcan, Paris 1893; ed. it. *La divisione del lavoro sociale*, Il Saggiatore, Milano 2016.

³ *Ideologie und Utopie*, Cohen, Bonn 1929; ed. it. *Ideologia e utopia*, Mulino, Bologna 1957.

⁴ *The Social Construction of Reality*, Anchor Books, New York 1966; ed. it. *La realtà come costruzione sociale*, Mulino, Bologna 1969.

⁵ D. Bloor, *Knowledge and Social Imagery*, Chicago University Press, Chicago 1976.

⁶ Vedi però P. Rossi, *La nascita della scienza moderna in Europa*, Laterza, Roma-Bari 2000; S. Schaffer, S. Shapin, *Leviathan and the Air-Pump*, Princeton University Press, Princeton 1985.

⁷ J. Gibbs, *Moral Development and Reality: Beyond the Theories of Kohlberg and Hoffman*, Allyn & Bacon 2009.

⁸ Ho però cercato di approfondirla in precedenti pubblicazioni, A. Di Prospero, *Linguaggio e ordine del mondo*, Il Sileno, Cosenza 2020.

che globalmente li tenga ciascuno nel giusto conto. Dall'altra, nel nostro caso, è soprattutto la varietà estrema di manifestazioni con cui si possono dispiegare i rapporti tra individuo e società (sia sul piano descrittivo sia su quello normativo), che ci consiglia di far valere da subito simili indicazioni, che servono indirettamente a dare una definizione di ciò che intenderemo per "intersoggettività", evitando in particolare di appiattire una simile nozione su solo alcune situazioni (comunità scientifica o consenso elettorale o appartenenza a una tradizione culturale comune etc.). Per come è qui intesa, la disposizione dell'individuo verso l'intersoggettività e la condivisione sociale possono rimanere confermate anche nei contesti e nelle situazioni più diverse, comprese anche – per esempio – quelle in cui l'individuo si mostri irriducibile oppositore contro elementi della società in cui vive, ma dimostrando proprio così il proprio attaccamento ad altri aspetti di esse che, *in nuce*, contengono i motivi dell'opposizione. Si pensi a come nelle società occidentali, fortemente pervase dai valori dell'individualismo, possa finire per essere sentito quasi come un obbligo quello di protestare anche contro le scelte pubbliche più ovvie: un tale sentimento di obbligo, anche se non percepito come tale, è comunque espressione e conseguenza di un'appartenenza sociale.

Rispetto ai sociologi sopra ricordati, la posizione che qui si difenderà *non* è né di un tipo "sociocentrico"⁹ né incline alle forme di relativismo che facilmente si associano agli approcci più fortemente sociologisti. Il riferimento che farò all'epistemologia genetica di Piaget – data l'importanza che questa associa all'esistenza di un punto di vista che originariamente è caratterizzato proprio dall'egocentratozza – dovrebbe infatti contenere, *in nuce*, le componenti teoriche che possono servire a evitare questi rischi.

Fatte tutte queste precisazioni, si può forse meglio intendere il significato della critica che muoverò alla concezione di Turing, sostenendo da una parte che essa «offre una definizione operativa dell'intelligenza proprio nello spirito della psicologia comportamentale dell'epoca del dopoguerra»¹⁰, ma che al tempo stesso, dall'altra, è proprio in un tale modello epistemologico che affondano le proprie radici i limiti della sua impostazione. Diane Proudfoot definisce «canonical»¹¹ l'interpretazione comportamentista del test di Turing. Il consenso su questo punto però non è unanime¹². Qui non servirà in ogni caso entrare in questo dibattito: non è in generale mia intenzione concentrarmi su presupposti teorici che sono

⁹ D. Wrong, *The Oversocialized Conception of Man*, in «American Sociological Review», n. 26/2, 1961, pp. 183-193.

¹⁰ D. Bolter, *Turing's Man. Western Culture in the Computer Age*, University of North Carolina Press, Chapel Hill, 1984; ed. it. *L'uomo di Turing. La cultura occidentale nell'età del computer*, SPE Pratiche, Parma, 1985, p. 235; cfr per esempio J. Searle, *The Turing Test: 55 Years Later*, in R. Epstein, G. Roberts, G. Beber (a cura di), *Parsing the Turing Test. Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Springer, Dordrecht, 2008, pp. 139-150; T.E. Wahlen, *A Computational Behaviorist Takes Turing's Test*, in R. Epstein et al., cit., pp. 343-357.

¹¹ D. Proudfoot, *Rethinking Turing's Test*, in: «The Journal of Philosophy», 110/7 (2013), pp. 391-411, p. 391.

¹² D. Proudfoot, *Turing's Concept of Intelligence*, in J. Copeland, J. Bowen, M. Sprevak, et al. (a cura di), *The Turing Guide*, Oxford University Press, Oxford 2017, pp. 301-307.

solo del comportamentismo. Intendo invece sostenere che il comportamentismo porta alle estreme conseguenze un concetto di verità che è stato estremamente influente nella nostra storia culturale, per il quale un attributo di una credenza “vera” dovrebbe essere quello del poter essere accettata e condivisa socialmente, accettando poi eventualmente una specifica perimetrazione del gruppo sociale di riferimento (per esempio la comunità scientifica, il contesto culturale di appartenenza o lo stesso genere umano), o più spesso lasciando in sospeso una tale scelta teorica, che chiaramente è carica di problemi e difficoltà. Più che di una tesi precisa che ricorrerebbe nella storia della filosofia, si può qui parlare più generalmente di una istanza tesa ad affermare la natura sociale della conoscenza, che peraltro ha trovato espressioni significative anche in etologia¹³.

Tali affermazioni andrebbero precisate molto più in dettaglio. Ad esempio, l'appello alla condivisibilità di principio delle proprie credenze, quando siano vere, non implica naturalmente che in una disputa si procederà necessariamente a stabilire quale delle parti riscuota più consenso. Ognuno dei contendenti però cerca di sfruttare gli spazi che ha a disposizione per rivendicare che o anche altri riconoscono la validità delle sue tesi oppure che, se ciò non avviene, è perché vi sono altri fattori che interferiscono (opportunismo, paura etc.). Inoltre è ben chiaro che la ricerca del consenso sociale, per come si manifesta per esempio nella ricerca scientifica, ha caratteristiche molto diverse da altri ambiti in cui pure la dimensione sociale e partecipativa dell'attività che si svolge è preminente. Tutte queste clausole, però, possono facilmente essere messe in conto proprio partendo da una descrizione dei rapporti – sul piano epistemico – tra individuo e società, del tipo di quella già esposta: si sostiene qui infatti che è l'esperienza di vita delle persone a creare il senso che la condivisione delle credenze – a seconda degli ambiti e delle circostanze – dia a esse un sostegno. Ne segue che se in un certo ambito della nostra esistenza la tendenza di fondo è a fare certe esperienze e non altre, anche il concetto di intersoggettività che ne deriva sarà corrispondente (si pensi ad esempio a ciò che avviene nell'ambito dell'arte).

La modernità storicamente ha rappresentato se stessa secondo un criterio che vede nella capacità di avere un pensiero critico un tratto essenziale per la sua identificazione¹⁴. Ma questo vuol dire che una definizione che fosse *puramente* intersoggettiva (o secondo la terminologia più diffusa, di tipo consensualista) del concetto di vero, nel caso in cui si pretendesse di farla accettare su di un piano esplicito e cosciente, e non su quello ben più elusivo e sotterraneo delle dinamiche sociologiche che si sono già richiamate, potrebbe creare perlomeno delle frizioni proprio rispetto all'ammissione del valore del pensiero critico. Non è mia intenzione affermare che la condivisione sociale di una credenza non possa in

¹³ N. Humphrey, *The Social Function of Intellect*, in P.P.G. Bateson, R.A. Hinde (eds), *Growing Points in Ethology*, Cambridge University Press, Cambridge 1976, pp. 303-317

¹⁴ I. Kant, *Was heißt sich im Denken orientieren?*, in *Kants Gesammelte Schriften*, De Gruyter, Berlin 1900 ss., VIII, pp. 129-147; ed. it. *Che cosa significa orientarsi nel pensiero*, Adelphi, Milano, 2006.

genere essere considerata come un criterio importante per stabilire la sua verità, ma vorrei piuttosto suggerire che potrebbe essere interessante approfondire le implicazioni della tesi che la condivisione sociale di una credenza sia essenzialmente un indizio di tipo *empirico* della sua verità, da valutarsi nel suo peso contestualmente e globalmente insieme a tutti gli altri fattori che possono apparire rilevanti: in questo modo anche l'epistemologia della critica alle idee socialmente maggioritarie dovrà essere corrispondentemente diversa. Come esempio del tipo di concezione che qui critico, si può prendere (tra le molte possibili) quella difesa da Jürgen Habermas¹⁵ (senza che però ci si possa qui soffermare su questo autore nello specifico), ma il fatto che Habermas non si lascerebbe mai identificare come un "comportamentista" fa capire che i presupposti epistemologici che intendo tematizzare vanno assai al di là sia dell'opera di Turing sia degli autori rappresentativi del comportamentismo.

L'intensificazione delle comunicazioni resa possibile dalle tecnologie digitali, la possibilità di connettere tra loro parlanti provenienti da ogni parte del pianeta, sono circostanze che rendono intuitivo che una definizione del vero di tipo "intersoggettivista" in senso troppo forte e rigido possa essere oggi problematica: le identità che si confrontano nel teatro degli scambi comunicativi – e che noi comunque *vorremmo* che potessero continuare a confrontarsi in questo agone, evitando ripiegamenti ispirati a un comunitarismo troppo soffocante – sono tra loro troppo diverse e distanti per poter far sentire le persone vincolate a un concetto di vero che corrispondentemente rischia di diventare del tutto astratto, formale e di fatto vuoto (secondo il tipo di obiezione che in modi diversi è stata portata sia contro Kant sia contro Habermas).

Sul piano formale, cercherò nell'opera di un altro grande pioniere dell'Intelligenza Artificiale, John McCarthy¹⁶, gli strumenti concettuali per articolare delle ipotesi che permettano di giustificare le mie obiezioni al test di Turing. Sul piano etico, sociale e culturale, gli strumenti da doversi prendere in considerazione sono di tipo diverso: essendo qui il concetto di critica uno dei perni della questione, è al pensiero di Nietzsche che mi rivolgerò. Cercherò di sostenere che – nel quadro storico e sociale che si va oggi delineando – le forme del "sospetto" e della "demistificazione" che hanno trovato espressione nell'opera di Nietzsche¹⁷ possono essere sottoposte esse stesse a una critica che può farne apparire le valenze in modo molto diverso da come inteso dal filosofo di Röcken.

Prima di Nietzsche, Cartesio aveva rappresentato – nella lettura storiografica più diffusa, risalente a Hegel, che qui si accetta – il momento in cui la soggettività acquista, con la modernità, una posizione e un ruolo del tutto nuovi

¹⁵ J. Habermas, *Theorie des kommunikativen Handelns*, Suhrkamp Verlag, Frankfurt am Main, 1981; ed. it. *Teoria dell'agire comunicativo*, Mulino, Bologna, 1986.

¹⁶ J. McCarthy, *Formalizing Common Sense*, Ablex, Norwood, NJ, 1990.

¹⁷ Mannheim come è noto considera Nietzsche un precursore della sociologia della conoscenza (Mannheim, cit., ed. it. p. 303). Per questa via, le mie considerazioni su Nietzsche possono applicarsi, con i necessari – ampi – aggiustamenti, anche a posizioni come quelle di Mannheim.

nella *Weltanschauung* europea, istituendo al tempo stesso da subito un nesso inscindibile tra esercizio della *individualità* e necessità di reagire con la propria intelligenza critica contro le possibilità di rimanere vittima di una manipolazione. Nietzsche segna il momento in cui più diventa drammatico il rapporto individuo-società che con Cartesio si era inaugurato sotto un segno ancora incautamente ottimistico oltre che in maniera ancora molto indiretta: di questo autore si dovranno ricordare – come uno sfondo implicito che definisce la problematica entro la quale le mie riflessioni si collocano – le posizioni espresse in particolare in un’opera come *Verità e menzogna in senso extra-morale*¹⁸. Anche in questo caso, come per Cartesio, un pensatore come Nietzsche è storicamente troppo imponente per poter essere fatto oggetto di una trattazione riduttiva, ma è bene portare almeno alcuni passi come esempi di come la tematica del rifiuto critico (la “demistificazione”) delle convinzioni più diffuse socialmente figuri nella sua filosofia. Il peso da lui dato a una comprensione di tipo psicologico dei comportamenti umani¹⁹ fa sì che le sue analisi siano dotate fenomenologicamente di una particolare ricchezza, ma questo rende difficile venire a capo di un carattere sfuggente che finiscono per avere se si cercano chiavi interpretative più precise e sufficientemente generali. Più per far capire a quali idee sono qui interessato, senza tentare una lettura complessiva del filosofo (è noto infatti quanto storicamente le interpretazioni della sua opera siano numerose e tra loro divergenti), si possono ricordare passi come il seguente:

da tempo immemorabile si è stati pochissimo “disinteressati”, pochissimo “altruisti” nella lode delle virtù! [...] Se hai una virtù, una virtù reale, tutta d’un pezzo (e non soltanto un esiguo impulso verso una virtù) – allora ne sei la *vittima*! Ma è ben per questo che il vicino loda la tua virtù! [...] La lode delle virtù è la lode di qualcosa che è in privato dannoso – la lode d’impulsi che tolgono all’uomo il suo più nobile egoismo e la forza della suprema cura di se medesimo²⁰.

In questo stesso ordine di idee si può leggere il paragrafo su *Il martire contro volontà*, contenuto in *Umano, troppo umano*:

In un partito c’era un uomo che era troppo timoroso e pavido per contraddire mai i suoi camerati: se ne aveva bisogno per ogni servizio, da lui si otteneva tutto, perché egli temeva la cattiva opinione dei suoi compagni più della morte; era una misera anima debole. Essi si accorsero di ciò, e, sfruttando le suddette qualità, fecero di lui un

¹⁸ F. Nietzsche, *Über Wahrheit und Lüge im außermoralischen Sinne*, in *Werke*, vol. III, tomo 2, De Gruyter, Berlin, 1967-77, pp. 367-384; ed. it. *Su verità e menzogna in senso extramorale*, in *La filosofia nell’epoca tragica dei greci e Scritti 1870-1873*, Adelphi, Milano, 1991, pp. 225-244.

¹⁹ F. Nietzsche, *Menschliches, Allzumenschliches I*, in *Werke*, vol. IV, tomo 2, De Gruyter, Berlin, 1965-1967, pp. 3-375; ed. it. *Umano, troppo umano*, II voll., Adelphi, Milano, 1965, vol. I, pp. 3-305, pp. 43-47.

²⁰ F. Nietzsche, *Die fröhliche Wissenschaft*, in *Werke*, vol. V, tomo 2, de Gruyter, Berlin, 1973, pp. 13-335; ed. it. *La gaia scienza*, in *La Gaia scienza e Idilli di Messina*, Adelphi, Milano, 1977, pp. 19-335, tr. it. 1977, pp. 72-73.

eroe e alla fine addirittura un martire [...] finanche sul patibolo, quando morì per le convinzioni del suo partito: al suo fianco stava cioè uno dei suoi vecchi compagni, che con lo sguardo e la parola lo tiranneggiò talmente, che egli veramente patì la morte nella maniera più compita. Viene da allora celebrato come martire e grande carattere²¹.

Il richiamo – anche se rapido – all’insieme di queste coordinate serve a cercare di ricostruire una “genealogia” di questioni che generalmente sono percepite come molto distanti, ma che nella mia analisi affondano le proprie radici proprio in una impostazione del problema della conoscenza e della sua legittimazione, che storicamente nella filosofia occidentale ha avuto ampia diffusione.

Secondo questa lettura è lo stesso Turing che, con il suo celebre “test”, implicitamente conferma i presupposti di una epistemologia che rende costitutivamente difficile venire a capo dei problemi di ordine culturale che oggi la diffusione del Web impone all’attenzione di tutti. Per Turing si potrebbe giudicare che un computer è in grado di “pensare” nel caso in cui un essere umano possa arrivare a scambiarlo per un altro essere umano, nel corso di un dialogo che avvenga attraverso dispositivi che focalizzino l’attenzione esclusivamente sul contenuto semantico delle frasi, e non sull’aspetto esteriore dei modi usati per produrle²². Ma questo significa che si sceglie di identificare gli *indizi* esterni del pensiero per il pensiero stesso, che è invece in genere – molto di più degli indizi – ciò che in realtà ci interessa nel corso del processo di comunicazione.

Un confronto tra Turing e Nietzsche non può prescindere dal contrasto tra l’atteggiamento dei due autori verso la “verità” in generale e verso la scelta di adottare un metodo scientifico nell’indagine intellettuale: questo punto, anche se sicuramente vero, è abbastanza evidente, e di per se stesso non sembra poter promettere acquisizioni euristiche significative. È invece più interessante notare che l’influenza del comportamentismo su Turing permette di ritrovare, in lui come in Nietzsche, una concezione della “verità” come essenzialmente e irriducibilmente *sociale*, salvo che il filosofo tedesco vede questo come un motivo per dover procedere in un’opera di demistificazione, mentre il matematico inglese considera questa una scelta indispensabile per sfuggire alle assurdità del «solipsism»²³. Il punto più interessante, però, a mio avviso, non sta ancora in questo rilievo, che da solo non permette ancora di orientarsi nel dilemma che si crea, ponendoci di fronte a queste due posizioni così diverse e irriducibili. Piuttosto, può essere trovato in un’implicazione di tipo

²¹ F. Nietzsche, *Menschliches...*, cit., ed. it. pp. 62-63.

²² S. Shieber (a cura di), *The Turing Test. Verbal Behavior as the Hallmark of Intelligence*, MIT Press, Cambridge, Ma, 2004; R. Epstein, G. Roberts, G. Beber (a cura di), *Parsing the Turing Test. Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Springer, Dordrecht, 2008; C.T. Rubin, *Robotic Souls*, in «The New Atlantis», (2019) 57, pp. 75-82; S. Bringsjord, P. Bello, D. Ferrucci, *Creativity, the Turing Test, and the (Better) Lovelace Test*, in: «Mind and Machines», 11 (2001), pp. 3-27; J. Moor, *The Status and Future of the Turing Test*, in: «Mind and Machines», (2001) 11, pp. 77-93.

²³ A. Turing, cit., p. 447.

storico e sociologico che deriva da questa cornice concettuale: rispetto agli anni in cui è vissuto Nietzsche, oggi il dilatarsi ulteriore della rete dei rapporti di comunicazione – il suo estendersi a coprire relazioni tra soggetti sempre più distanti, portatori di identità diverse, in un contesto frammentato e variegato, privo di centri di riferimento comuni – ha *acuito* dinamiche sociali che creano una disposizione al disincanto e allo scetticismo, che sono state già nell'Ottocento il terreno su cui è sorta la filosofia nietzschiana. Vorrei quindi sostenere che proprio il radicalizzarsi dei processi caratteristici della modernità può metterci oggi in grado di sviluppare una immagine dell'umano che vada *oltre* quella di Nietzsche, in effetti confutando premesse che per quella filosofia sono essenziali. Si possono cioè re-interpretare i fenomeni psicologici la cui osservazione è stata il presupposto delle riflessioni di Nietzsche, utilizzando una teoria logica che fondamentalmente, almeno in prospettiva, è di un carattere assai più formale e sistematico, tale cioè da muoversi secondo una strategia essenzialmente diversa da quella del filosofo tedesco, per quanto interessata a trattare problemi simili.

Il compito che mi do è di tipo ricognitivo ed esplorativo: cercherò quindi solo di identificare alcune linee di carattere generale che una successiva ricerca dovrebbe approfondire in modo più rigoroso. L'idea di fondo è che sia necessario soffermarsi sulle caratteristiche del ragionamento non-monotono: John McCarthy, lo scienziato celebre anche a un pubblico più vasto per aver introdotto l'espressione "intelligenza artificiale", ha cercato di codificare le intuizioni soggiacenti al senso comune (umano) attraverso una teoria logica di tipo non-monotono, da lui definita «circumscription logic»²⁴. L'intenzione era quella di poter arrivare a realizzare un programma per l'I.A. che fosse in grado di guidare un computer nel "ragionare" come un essere umano. In questo modo sembrava in effetti possibile affrontare il problema che è stato posto da Hubert Dreyfus²⁵, che ha basato la sua critica del programma dell'A.I. esattamente sull'impossibilità per un computer di compiere ragionamenti che, come quelli umani, tengano conto del numero indefinito di possibili variazioni dei diversi contesti di azione:

Le macchine mancano di intelligenza pratica. Esse sono "esistenzialmente" stupide nel senso che non possono fronteggiare situazioni specifiche. Così esse non possono accettare l'ambiguità e la rottura delle regole, fino a che le regole per trattare le deviazioni siano state così completamente specificate che l'ambiguità sia scomparsa²⁶

La mia prospettiva di ricerca è però differente da quella di McCarthy (è questo che mi obbliga a potermi rifare solo molto parzialmente alla sua opera) o degli altri studiosi, come Raymond Reiter, che hanno proseguito sulla sua stessa strada. McCarthy si poneva obiettivi di tipo ingegneristico: dava per scontato che pos-

²⁴ J. McCarthy, cit. p. 142.

²⁵ H. Dreyfus, *What Computers Can't Do*, Harper & Row, New York, 1972; ed. it. *Che cosa non possono fare i computer*, Armando, Roma, 1988.

²⁶ Ivi, p. 285.

sediamo una discreta comprensione intuitiva di come ragionano gli esseri umani. Non si inoltrava in un'analisi del ragionamento non-monotono (umano) che potesse avere il taglio e i risvolti potenziali di un'antropologia filosofica. Cercava in sostanza solo di fornirne una codifica che potesse poi essere implementata con successo sulle macchine. L'obiettivo era quello indicato da Turing con l'idea del suo celebre test: ottenere che un computer rispondesse a un dialogo in un modo che lo rendesse indistinguibile da un essere umano.

La direzione della mia indagine sarà invece proprio quella che nei testi di McCarthy rimane più in ombra. Non è infatti ovvio che le implicazioni del fatto che la mente umana ragioni in modo non monotono siano concettualmente e fenomenologicamente chiare. Riprendendo – purtroppo sempre in modo solo molto rapido – la posizione di Nietzsche, si può accostare la forma complessiva del ragionamento non-monotono all'idea che nei nostri pensieri siamo guidati da un «mobile esercito di metafore»²⁷? Ma soprattutto – al fine di rendere meno generico il discorso – che implicazioni precise si dovrebbero ricavare da questo accostamento – anche sul piano della storia del pensiero?

Per Dreyfus l'essere umano è sempre radicato in un contesto: forme di sapere impersonale e atemporale, quali quelle previste da Platone con la sua “teoria delle idee”, non sono ammissibili. In altre parole, utilizzando una concettualizzazione diversa, il pensiero umano è temporale e non-monotono – così come Heidegger (alla cui opera Dreyfus si richiama) aveva sostenuto che l'essenza dell'uomo è la temporalità. Secondo Dreyfus si può ritenere che una macchina – un computer – che esegue algoritmi, possa arrivare a pensare come un essere umano soltanto se si è convinti che l'essere umano sia esso stesso un ente in grado di pensare in modo atemporale.

Naturalmente il libro di Dreyfus che qui consideriamo è assai datato e soprattutto è stato superato dai successi effettivi che l'ingegneria nell'I.A. ha ottenuto negli anni successivi alla sua pubblicazione. È da notare che i progressi nelle ricerche sul *machine learning* hanno fatto largo uso dell'induzione (si ricordino *Progol* di Stephen Muggleton e *Aleph* di Ashwin Srinivasan). Ma l'inferenza induttiva è uno strumento logico di tipo monotono o non-monotono? La posizione che qui si assume sull'induzione – che sarà fondamentale per il proseguimento di queste riflessioni – è quella di Isaac Levi, che vede la logica induttiva come non-monotona e che ritiene che molti filosofi della scienza imposterebbero i loro studi sull'induzione cercando di evitare il ricorso teorico a «una forma di soggettività nella ricerca che darebbe spazio agli eccessi del relativismo culturale e dell'irrazionalismo»²⁸, mentre

altri, come Popper, Carnap e i loro seguaci anti-induttivisti, coltivano e proteggono l'obiettività della scienza negando la possibilità dell'espansione induttiva nella

²⁷ F. Nietzsche, *Über Wahrheit...*, cit., ed. it. p. 233.

²⁸ I. Levi, *For The Sake of The Argument*, Cambridge University Press, Cambridge, Ma, 1996, p. 160, trad mia.

scienza. Consigliare di astenersi dall'induzione in questo modo è un'ammissione di fallimento del tentativo di portare la fissazione della credenza sotto un controllo critico. Dato che i ricercatori estrapoleranno dai dati e aggiungeranno congetture all'informazione di sfondo come risorse per le successive indagini senza riguardo per quello che gli anti-induttivisti sostengono, l'ammissione di fallimento concede ai relativisti sociali e agli anti-razionalisti che l'induzione è materia per psicologia, sociologia e storia e che non c'è modo per tentare di identificare standard prescrittivi per asserire la sua legittimità²⁹.

In effetti il riferimento a psicologia e sociologia nella riflessione sull'induzione non deve esser svalutato, ma lo spirito di fondo di Levi, teso a salvaguardare il senso di razionalità che deve riconoscersi ai procedimenti induttivi, viene qui accettato e difeso. In questa direzione si muove anche Carlo Cellucci³⁰, che difende sistematicamente il carattere non-monotono della conoscenza umana (peraltro *anche* nel caso del sapere strettamente deduttivo).

Seguirebbe da queste considerazioni che in altre parole Dreyfus sarebbe nel vero nel ritenere che l'essere umano è radicato sempre in un contesto (dato che l'induzione è una forma di inferenza non-monotona), ma (se si accetta la validità di fondo del ragionamento induttivo) sbaglierebbe nel sostenere che per questo esso debba seguire forme di pensiero irriducibilmente diverse da quelle di un computer. Al tempo stesso, in una direzione di pensiero opposta, si può osservare che lo statuto del pensiero non-monotono non è del tutto chiaro e privo di ambiguità. Nel parlare di "verità" e "conoscenza" non intendiamo di fatto implicitamente che debba esservi qualcosa in una credenza che la renda attendibile in un modo che sia *stabile* e universale? Ma come intendere qui il carattere della stabilità? Si dovrà concludere che l'unica conoscenza reale debba essere di tipo monotono? Un tale esito rischierebbe di essere totalmente disastroso: per stare sempre all'induzione, noi sappiamo bene da un lato che le nostre conoscenze induttive sono logicamente rivedibili (in quanto non-monotone), ma pure non siamo presi da nessuna ansia nell'apprestarci a mangiare il pane che ci ha sempre nutrito o a bere l'acqua che ci ha sempre dissetato. Trattiamo cioè le nostre conoscenze *non-monotone* come vere e proprie conoscenze, per quanto esse siano per definizione *non* stabili. In realtà è assai sintomatico il fatto che in filosofia è ripetutamente comparsa la difficoltà a dar conto della conoscenza dell'*individuale* in quanto tale, vale a dire di ciò che è più instabile e corruttibile per definizione: in Aristotele, per il quale vi era conoscenza scientifica solo dell'universale, come in Hegel o come anche nell'idea che vi sarebbe una unità dell'"intelletto possibile" per tutti gli uomini, come posta nelle dispute medievali sull'averroismo latino. Già l'eterogeneità di questi soli tre riferimenti mostra come il problema tenda a riproporsi all'interno dei modelli di pensiero più diversi.

²⁹ *Ibidem*.

³⁰ C. Cellucci, *Le ragioni della logica*, Laterza, Roma-Bari, 1998.

Il mio proposito è di mostrare che si può forse evitare questa aporia. Si osservi che in riferimento a problemi come questi può tornare in mente nuovamente Nietzsche: se base del sapere è l'esperienza, che è sempre individuale, e se l'esperienza viene generalizzata per induzione, con una forma di inferenza che è non-monotona, la "verità" concepita come universale e monotona non può che essere il frutto di un accordo o di una convenzione, con i limiti e il tipo di status epistemologico che ciò comporta³¹.

L'idea di fondo è che lo schema teorico generale da adottarsi sia derivabile dall'epistemologia genetica sviluppata da Jean Piaget³²: secondo questo grande studioso il neonato non dispone delle nozioni di spazio e di tempo, di oggetto semi-permanente, della distinzione tra soggetto e mondo. La sua attività cognitiva è centrata sul presente e sui dati immediati, che sono costituiti da un insieme di unità che continuamente compaiono e scompaiono, senza un ordine e delle regolarità che rendano possibili le previsioni che fa invece l'adulto. Questo quadro è stato sottoposto a molte critiche (per esempio da parte di Margaret Donaldson, Andrew Meltzoff, Elizabeth Spelke, Renée Baillargeon e molti altri). Le ragioni per cui qui sosterrò che può comunque essere accettato e difeso dipendono in parte dal fatto che se ne può in effetti dare una re-interpretazione³³, meno orientata verso i principi dello strutturalismo fatti propri da Piaget, e più affine invece a un induttivismo empirista. L'idea è appunto che il neonato raccolga momento per momento contenuti di esperienza che – proprio per il fatto di essere in una condizione di radicale, anche se inconsapevole, ego-centratezza – generalizza induttivamente, scambiandoli per una immagine precisa di tutta la realtà. Si potrebbe dire che, in un senso ovvio, è proprio questa – la tendenza a generalizzare per induzione, secondo una logica non-monotona – l'origine della possibilità di sviluppare *bias* che equivalgono alla *mancanza* di senso critico, ma al tempo stesso essa è anche la radice delle generalizzazioni successive che portano alla costruzione progressiva della visione del mondo comunemente accettata dagli adulti, e con ciò alla possibilità in linea di principio della critica di quei pregiudizi che essa stessa in un primo momento aveva portato a legittimare.

Entro questa prospettiva teorica il senso dell'intersoggettività e il senso della reciprocità razionale nascono per effetto di generalizzazioni compiute su esperienze di vita individuali che li rendono possibili e credibili. Ne segue però che non si può assumere l'intersoggettività come un *prius* metodologico, come accade nel comportamentismo: essa è invece piuttosto proprio il risultato di un apprendimento (individuale) dovuto all'esperienza³⁴.

³¹ Potrebbe essere interessante porre in relazione questo punto con le questioni sollevate da D. Xu, Y. Wang, S. Bhattacharya, *Thinking Beyond Means-Ends Analysis: The Role of Impulse-Driven Human Creativity in the Design of Artificially Intelligent Systems*, in D.N. Hart, S.D. Gregor, *Information Systems Foundations: The Role of Design Science*, ANU Press, Canberra, 2010, pp. 213-232.

³² J. Piaget, *La construction du réel chez l'enfant*, Delachaux et Niestlé, Neuchâtel, 1937; ed. it. *La costruzione del reale nel bambino*, La Nuova Italia, Firenze, 1973.

³³ A. Di Prospero, cit.

³⁴ Per una ricostruzione storica molto critica delle ambiguità presenti nei testi degli autori

Parlare della “verità” come di una convenzione sociale – paradossalmente – significa collegare le paure che opprimono Nietzsche ai principi del comportamentismo, che questa convenzione eleggono a principio che deve orientare ogni ricerca razionale e scientifica.

È possibile però disporre all’interno di una concettualizzazione coerente e intuitiva le diverse istanze teoriche che così si sono raccolte. Le esperienze fatte dal bambino nel corso dello sviluppo lo convincono – in forza dell’applicazione di una logica non-monotona – che si può avere fiducia in messaggi e in contenuti di credenza ottenuti applicando forme di pensiero monotono. Nietzsche si chiede come possa nascere «quell’enigmatico impulso alla verità»³⁵ che si trova negli esseri umani, rispondendo:

L’intelletto, come mezzo per conservare l’individuo, spiega le sue forze principali nella finzione. Questa infatti è il mezzo con cui gli individui più deboli e meno robusti si conservano [...] Nell’uomo quest’arte della finzione raggiunge il suo culmine: [...] il far la commedia dinanzi agli altri e a se stessi [...] costituisce a tal punto la regola e la legge³⁶.

Il concetto di verità cui si deve metter capo è inoltre «una designazione delle cose uniformemente valida e vincolante [in cui] la legislazione del linguaggio fornisce altresì le prime leggi della verità»³⁷.

La dialettica che qui ho cercato di porre in evidenza tra pensiero non-monotono e monotono, vorrebbe servire però a dimostrare che l’analisi di Nietzsche è unilaterale: è lui stesso che utilizza un criterio di giudizio implicitamente monotono e universalista, per cui non può che finire per condannare come un «continuo svolazzare intorno alla fiamma della vanità»³⁸ le inclinazioni e i prodotti attribuibili all’intelletto umano. Una epistemologia del pensiero non-monotono sarebbe invece per sua natura meglio in grado di accettare le diversità in un universo contenente una pluralità di poli di riferimento.

2. Conoscenza e standard sociali

Che cosa significhi qui mettere in relazione il test di Turing con l’idea che la verità sia una sorta di artefatto sociale, lo si può scorgere – in una prima approssimazione – considerando le tesi di Brian P. Bloomfield e Theo Vurdubakis, che interpretano le varie competizioni oggi organizzate tra uomo e computer, come «an organized rehearsal and dramatisation of a wide range of moral conflicts

comportamentisti che qui più sono rilevanti, cfr D. Dennett, *Skinner Skinned*, in *Brainstorms. Philosophical Essays on Mind and Psychology*, Bradford Books, Montgomery, VT, 1978, pp. 53-70.

³⁵ F. Nietzsche, *Über Wahrheit...*, cit., ed. it., p. 229.

³⁶ Ivi, p. 228.

³⁷ Ivi, p. 229.

³⁸ Ivi, p. 228.

characteristic of occidental modernity»³⁹, relativi secondo gli autori al rapporto che in generale gli esseri umani hanno con le macchine, o quelle di Mary McGee Wood, che si richiama invece a Roland Barthes. Nei termini della teoria semiotica di quest'ultimo

The very concept of myth, of a second-order sign system, of *signification*, affirms the importance of representation, and the primacy of the level of cultural appearance and suggestion over literal depiction, literal meaning. This gives us a principled basis for an endorsement of the position assumed by Alan Turing. If a machine can appear to be intelligent, Turing is prepared to say that it is [...] The dominance of *signification* confirms the validity of the apparent intelligence of the machine, the independent viability of the simulation, the representation⁴⁰

Secondo l'autrice «the Turing simulation is the extreme logically possible instantiation of Barthes's principles»⁴¹. È interessante che McGee Wood spinga la sua interpretazione fino a vedere nell'intelligenza artificiale una sorta di forza creatrice ed espansiva a oltranza, con una netta divergenza rispetto alle posizioni di Dreyfus e con una scelta teorica difficilmente confrontabile con l'atteggiamento intellettuale e metodologico degli studiosi più impegnati sul fronte ingegneristico. Il modo in cui Barthes descrive il Testo sarebbe

virtually a literal description of the Turing simulation. A network, which expands through systematic combinatorial principles—it would be hard to describe the simulation better [...] The Turing simulation has in perfect measure the flexibility, the “give” or “play” which Barthes is looking for here. It shifts, moves; it will not give us the same answer twice⁴².

C'è un nucleo di verità di estremo interesse che si può scorgere se ci si inserisce in questo ordine di pensieri. Il comportamentismo non vede la “verità” (scientifica) come *convenzione*, ma riducendola alla condivisione sociale delle procedure per la sua definizione, di fatto dimostra di essere talmente *immerso* in quella che – in ogni caso – è una convenzione, o comunque l'effetto di un accordo che ha margini molto ampi di arbitrarietà, da prenderla per un dato di realtà. Con questo non mi sto spingendo verso un convenzionalismo che faccia perdere di vista nichilisticamente il rapporto con le intuizioni del realismo: la realtà sono le esperienze della persona, da cui la persona trae le basi e gli ingredienti per costruire e definire anche una “verità” concepibile come convenzione, ma che comunque si presenta come una “convenzione”

³⁹ B.P. Bloomfield, T. Vurdubakis, *The Revenge of the Object? On Artificial Intelligence as a Cultural Enterprise*, in: «Social Analysis», 41/1 (1997), pp. 29-45, p. 29.

⁴⁰ M. McGee Wood, *Signification and Simulation: Barthes's Response to Turing*, in: «Paragraph», 11/3 (1988), pp. 210-226, p. 216.

⁴¹ Ivi, p. 217.

⁴² Ivi, p. 220.

scelta del tutto ragionevolmente per la sua forza epistemica, cioè per la sua congruenza con le esperienze dell'individuo, che gli insegnano che il sapere socialmente condiviso è un criterio di accettabilità delle credenze che – almeno generalmente – può mettere sulla buona strada nei tentativi di orientarsi nel mondo. Si tratta di una convenzione perché formalmente la si sarebbe potuta istituire secondo un criterio anche completamente diverso, ma è ragionevole e fondata sull'esperienza perché – come in genere avviene – le convenzioni sono scelte per una loro particolare funzionalità sul piano concreto: più esattamente, non può neppure parlarsi di “scelta” in senso proprio, dato che il processo che porta alla sua implementazione è in buona parte implicito e inconsapevole.

Si osservi che Nietzsche 1) parlando della nascita dei “concetti” come dovuta al fatto che una parola «non è destinata a servire eventualmente per ricordare l'esperienza primitiva, non ripetuta e perfettamente individualizzata, ma deve adattarsi al tempo stesso a innumerevoli casi più o meno simili, cioè – a rigore – mai uguali, e quindi a casi semplicemente disuguali»⁴³ sta aderendo di fatto implicitamente a un'epistemologia rigorosamente *empirista* e *nominalista*⁴⁴. «Ogni concetto sorge con l'equiparazione di ciò che non è uguale»⁴⁵: per questo la “verità” non può che essere convenzione.

2) Qualifica l'atteggiamento psichico pre-razionale e pre-analitico in termini che in effetti possono far pensare proprio alla vivezza e alla fondamentale *volatilità* dei contenuti dell'attenzione del neonato secondo Piaget, ad esempio quando scrive che l'essere umano «Come essere razionale [...] non ammette più di essere trascinato dalle impressioni istantanee e dalle intuizioni»⁴⁶:

Solo quando l'uomo dimentica quel primitivo mondo di metafore, solo quando la massa originaria di immagini – che sgorgano con ardente fluidità dalla primordiale facoltà della fantasia umana – si indurisce e si irrigidisce, solo quando si crede, con una fede invincibile, che questo sole, questa finestra, questo tavolo siano veri in sé [con Piaget potremmo dire: quando si costituisce la nozione di oggetto]: in breve, solo quando l'uomo dimentica se stesso in quanto soggetto, e precisamente in quanto soggetto *artisticamente creativo*, solo allora egli può vivere con una certa calma, sicurezza e coerenza⁴⁷.

⁴³ F. Nietzsche, *Über Wahrheit ...*, cit. ed it. p. 232.

⁴⁴ Interessante al riguardo P. Gori, *Il meccanicismo metafisico. Scienza, filosofia e storia in Nietzsche e Mach*, Mulino, Bologna 2009.

⁴⁵ *Ibidem*.

⁴⁶ Ivi, p. 234.

⁴⁷ Ivi, p. 237. Per colmare lo iato teorico che sembra insinuarsi nel passare dall'epistemologia genetica a un utilizzo così sistematico e impegnativo della nozione di metafora, ci si può limitare qui a rimandare a tutta l'opera di George Lakoff, che associa da vicino le sue importanti ricerche sulla metafora con la dimensione evolutiva e genetica e con quella senso-motoria (centrale anche nelle ricerche di Piaget). In questo senso la metafora non è un dispositivo che *indebolisce* retoricamente il rigore di una tassonomia degli enti già preesistente, ma è invece il *primo* modo di riunire tra loro in un'unica categoria entità che sono comunque irriducibilmente separate e diverse nella realtà.

Inoltre, dato che qui interessa anche il rapporto con i problemi posti dall'informatica, è interessante ricordare l'importante contributo di Antonio Imbasciati⁴⁸, che cerca di elaborare una teoria psicogenetica e psicoanalitica in grado di descrivere il funzionamento della mente nel neonato e poi nell'adulto in termini compatibili con quelli appunto dell'informatica. Turing stesso in realtà osservava che un computer dovrebbe essere guidato da un programma assimilabile non tanto a quello che guiderebbe un adulto umano nei suoi pensieri e nei suoi comportamenti, ma semmai a quello che dovrebbe guidare un neonato⁴⁹, in modo che possa poi verificarsi anche per il computer un processo di apprendimento simile a quello che vede come protagonista il bambino.

In realtà Turing, risentendo in modo così evidente dell'influenza del comportamentismo, si mostra come un erede di una tradizione culturale – quella che Dreyfus fa risalire a Platone – che ha avuto un ruolo storico fondamentale nel pensiero occidentale, ma che non costituisce un presupposto indispensabile da assumersi necessariamente in ogni filosofia. La tradizione del pensiero monotono – che raggiunge uno dei suoi vertici negli *Analitici* di Aristotele, con la teoria del sillogismo – si presenta come metodo per attingere un sapere perfetto e non modificabile. Naturalmente nel suo ambito legittimo di applicazione ha effettivamente le proprietà logiche ed epistemologiche che la rendono di importanza così straordinaria, ma non può essere scambiata per un modello di *ogni* sapere. Un sapere cui sistematicamente l'individuo attinge è quello che, a partire dalla sua esperienza, sviluppa per induzione. L'opposizione tra pensiero monotono e non-monotono può essere proiettata su quella tra conoscenze universalizzabili socialmente e conoscenze di carattere individuale. Se si accetta l'interpretazione qui proposta dell'epistemologia genetica, è però il sapere non-monotono a “contenere” e legittimare quello descritto come monotono, e non viceversa.

Gli standard sociali per poter parlare di “verità” senza rischiare di venire corretti dai propri interlocutori, sono variabili di contesto in contesto e di epoca in epoca. In questo senso, come si può evincere partendo dalle riflessioni di McGee Wood e di Bloomfield e Vurdubakis, anche Turing con il suo test non sfugge a questa logica – anche se gli standard socialmente condivisi per poter parlare di “scienza” aspirano a uno status epistemico privilegiato. La mia prospettiva non contesta che il metodo scientifico possa effettivamente legittimare in modo caratteristico le pretese di sapere che avanza, ma si può sostenere che possa riuscire a farlo non *negando* i contenuti del sapere individuale, ma semmai individuando descrizioni dei contenuti di esperienza che *soddisfanno* le aspettative degli individui in quanto individui, nel loro insieme. Se nel loro insieme gli individui che si trovano nelle condizioni di poter giudicare nel merito di una questione sentono di dover dare l'assenso a una teoria perché corrisponde e spiega in modo adeguato i loro contenuti di esperienza, si avrà una spiegazione “intersoggettiva”, cioè condivisa, ma *non* ponendo l'intersoggettività come una dimensione supposta

⁴⁸ A. Imbasciati, *Nascita e costruzione della mente: la teoria del protomentale*, UTET, Torino, 1998.

⁴⁹ A. Turing, cit., p. 456.

superiore a quella delle esperienze individuali. La critica di Nietzsche può servire proprio a cogliere e dare espressione a questo timore: che l'intersoggettività, da convenzione, si trasformi in dogma. Al tempo stesso, l'utilizzo che qui si propone di fare dell'induzione permette di inquadrare in modo assai diverso le questioni da lui sollevate. Dal punto di vista del pensiero monotono, l'individuo sottoposto, per esempio nella sua educazione, a esperienze idiosincratiche (sue proprie o di una minoranza sociale cui esso appartiene) rischia di apparire come qualcuno che è stato soggetto a un inganno o a una mistificazione. Dal punto di vista del pensiero non-monotono invece l'individuo si *identifica* con le sue esperienze e con i giudizi che da esse – induttivamente – promanano. Il campo cognitivo ed esperienziale dell'individuo ne definisce l'identità, né ha senso – per *quell'individuo* – parlare di sé in termini che cerchino di prescindere da ciò che costitutivamente lo realizza nel suo proprio essere. Deve essere posta perciò su basi totalmente diverse la questione della possibilità di una “mistificazione” – e con essa quella dei rapporti da istituirsi tra portatori di identità più o meno profondamente diverse, centrale anche per una adeguata impostazione del dibattito sul rapporto tra tecnologie digitali e forme del dibattito pubblico e in generale della comunicazione in un mondo sempre più globalizzato. È di estremo interesse che Nietzsche, pur muovendo una critica così radicale contro le strutture epistemologiche più fondamentali del sapere tradizionale, si mostri in effetti come vittima lui stesso di un processo di interiorizzazione di un concetto di “verità” di tipo monotono, che lo porta poi inevitabilmente a vedere nelle varie pretese alla “verità” soltanto un inganno: se una verità, per essere tale, deve essere rigorosamente universale, l'individuo guidato dalla propria esperienza non potrà che sentirsi drammaticamente disorientato di fronte a un compito – quello dell'universalizzazione – che è intrinsecamente contraddittorio⁵⁰. L'individuo *sa* di agire sulla base delle proprie esperienze: il compito di una verifica in linea di principio *infinita*, da *tutti* i possibili punti di vista, della verità di questi contenuti (suoi e individuali) di esperienza, è un compito che inevitabilmente percepirà come *inutile e impossibile* a realizzarsi. Psicologicamente nella realtà lo tratterà di fatto come qualcosa di simile a un “esercizio di scuola”, che serve solo a confermare agli occhi dei nostri vicini che pronunciamo i giusti enunciati, conformi alle convenzioni del nostro gruppo sociale. Come dice Isaac Levi, la persona *comunque* procede nelle sue convinzioni per induzione e per esperienza. L'obiezione di Turing – che allora a questo punto non si potrebbe più sfuggire ai paradossi del solipsismo – è naturalmente importante, ma l'opera di Piaget (re-interpretata induttivamente) ci mette in grado di indicare una possibile soluzione del problema: la dimensione dell'intersoggettività è costruita essa stessa grazie alle

⁵⁰ Su di un piano esegetico, per motivi di sintesi si fa forse qui torto a Nietzsche, che per esempio nello *Zarathustra* parla spesso in modo da lasciar intendere che ha presenti i problemi qui discussi. In ogni caso, data la diversità di fondo dell'approccio qui difeso, un confronto testuale puntuale diventerebbe in effetti molto arduo. Si può pensare però che sia l'impostazione generale del suo pensiero che gli rende infine impossibile elaborare una risposta organica alle questioni qui poste.

esperienze individuali e per generalizzazioni induttive compiute dall'individuo. Il sentimento e la conoscenza che abbiamo dell'esistenza e del valore degli altri *si identificano* con un sistema di esperienze, che in senso tautologico sono le *nostre* e sono "individuali".

La riflessione sul ragionamento non-monotono ci permette di guardare con occhio molto diverso al concetto di mistificazione in Nietzsche: ciascun osservatore, procedendo nelle proprie generalizzazioni induttive (portatrici di un sapere soltanto *non* monotono), nega implicitamente le pretese di verità degli *altri* punti di vista (cosa che, a rigore, è ovvia: vedendo la tal cosa come la tal cosa, non posso vederla anche in un *altro* modo). In altri termini, ciò che vedono (in "prima persona") gli altri, inevitabilmente mi sfugge, ma dato che mi posso orientare in modo del tutto efficiente in un mondo intersoggettivamente condiviso servendomi di strumenti diversi, che prescindono da ciò che l'altro – nella sua singolarità e unicità – esattamente vede, questa affermazione non implica quello che Turing vede come una forma di solipsismo.

Sul piano dell'analisi psicologica, *ognuno* – nell'istante in cui pensa qualcosa – deve implicitamente trattare i punti di vista diversi dal proprio come meritevoli di una qualche critica: anche quello che lui stesso pensava pochi istanti prima, se potesse per assurdo essere compreso anche ora con piena e totale esattezza, apparirebbe come una *travisazione* della verità, perché tale da partire da esperienze che – in qualcosa – sono diverse. Utilizzare una logica non-monotona però significa poter *accettare* con estrema serenità quelle che solo in apparenza sono contraddizioni: ragionare con una logica monotona (come fa anche Nietzsche) significa dover pensare che siano in atto inganni e trappole, mentre l'atteggiamento psichico del pensare non-monotono è quello di percepire i comportamenti altrui soltanto come tali da seguire il proprio corso, diverso dal nostro.

In termini intuitivi, la critica che può farsi a Nietzsche è che il suo modo di vedere le possibilità della mistificazione rimanda implicitamente (e in molti passi anche esplicitamente, per quanto in modo solo accennato) a meccanismi di tipo induttivo: sono le abitudini sedimentate, le azioni pedagogiche degli adulti rivolte a plasmare il carattere dei più giovani, i rapporti di potere che stabilmente costringono qualcuno all'obbedienza, a poter produrre effetti eventualmente deformanti sulla volontà. Il problema però è che in *ogni* generalizzazione induttiva – anche quelle che fossero in assoluto le più originarie e le più prive di collegamento con le costellazioni dei vari rapporti interpersonali – è insito questo meccanismo. Nell'induzione accade sempre che ciò che vediamo viene preso per una immagine attendibile anche di cose che non vediamo. Non si può pensare di demistificare un livello delle costruzioni sociali, risalendo a quello precedente o procedendo verso uno successivo, quando anche questi si basano su di una estrapolazione che parte da un insieme limitato di informazioni e quindi è esposta per principio agli stessi rischi di deformazione prospettica dovuta all'induzione.

Il riferimento alla possibilità della mistificazione rimane come *mossa polemica* in concreto legittima all'interno di una pratica di confronto che sia volta a sostenere le diverse ragioni dei differenti *livelli* dell'identità: un individuo può

essere così convinto – ad un livello metalinguistico L – che a un livello-oggetto L', alcune sue precedenti convinzioni erano state il risultato di un ottenebramento indotto da centri di potere che avevano spinto la sua attenzione a essere occupata unilateralmente da un solo tipo di stimoli, e quindi a sviluppare quelle convinzioni corrispondenti che ora riconosce come false. Ma non si può pensare del proprio sé più profondo, delle convinzioni che definiscono la propria identità nel senso più pieno, che si tratti del risultato di una mistificazione – anche se il punto di vista impersonale e universale *from nowhere* del pensiero monotono, se venisse abbastanza interiorizzato, obbligherebbe in pratica sempre a pensare che l'individuo, con i limiti che lo definiscono, sia in definitiva la vittima di un tale inganno, o se si vuole (dato che questo inganno lo costituisce) qualcosa di molto simile a un nulla: questa impossibilità sussiste persino se sul piano causale qualcosa di simile a una mistificazione vi è effettivamente stato, e anzi persino se intellettualmente e in astratto si è portati a pensare di se stessi che sia avvenuto ciò.

3. Intersoggettività, conoscenza e metodo sperimentale

Riflettendo in termini del tutto interni al dibattito sull'I.A., si può osservare che la strategia teorica di Turing è esposta a dei rischi che confermerebbero la mia analisi. Le ricerche in filosofia della scienza e della tecnologia da diversi anni hanno evidenziato le specificità del metodo sperimentale in quanto applicato nel campo dell'I.A.⁵¹. Una macchina che fosse in grado di “pensare” dovrebbe essere in grado di produrre risposte che siano *originali*⁵², ma al tempo stesso, dato che la realizzazione di un programma e della macchina su cui implementarlo devono essere processi scientificamente accettabili, sembra che dovrebbe farsi valere un requisito di *riproducibilità* dei test che servono a dimostrare il successo dell'esperimento. Un tale problema è connesso da vicino a quelli in precedenza indagati: la riproducibilità degli esperimenti è collegata infatti al requisito della condivisibilità intersoggettiva dei risultati e delle procedure della ricerca scientifica. Ma

⁵¹ V. Schiaffonati, *Computer, Robot ed esperimenti*, Meltemi, Roma, 2020; B.G. Buchanan, *Artificial Intelligence as an Experimental Science*, in J.H. Fetzer (a cura di), *Aspects of Artificial Intelligence. Studies in Cognitive Systems, vol. 1*, Springer, Dordrecht, 1988. https://doi.org/10.1007/978-94-009-2699-8_8; O.E. Gundersen, *State of the Art: Reproducibility in Artificial Intelligence. Proceeding of the AAAI Conference on Artificial Intelligence*, 32/1 (2018). <https://doi.org/10.1609/aaai.v32i1.11503>; A. Schwarz, W. Krohn, *Experimenting with the Concept of Experiment: Probing the Epochal Break*, in A. Nordmann, H. Radder, G. Schiemann (a cura di), *Science Transformed?* University of Pittsburgh Press, Pittsburgh, 2011, pp. 119-134; F. Amigoni, M. Reggiani, V. Schiaffonati, *An Insightful Comparison Between Experiments in mobile Robotics and in Science*, in: «Autonomous Robots», 27 (2009), art. 313. <https://doi.org/10.1007/s10514-009-9137-8>; C. Andujar, V. Schiaffonati, F. Schreiber *et al.*, *The Role and Relevance of Experimentation in Informatics*, in *Proceeding 8th European Computer Science Summit (ECSS 2012)*, Barcellona, 2012.

⁵² M. Boden, *Creativity and Artificial Intelligence*, in «Artificial Intelligence», 103 (1998), pp. 347-356; R.A. Beghetto, *There is no Creativity Without Uncertainty: Dubito Ergo Creo*, in: «Journal of Creativity», 31 (2021), art. 100005.

come dimostrare che un esperimento ha avuto successo nella sua prima realizzazione e nelle sue repliche successive, se ci si deve aspettare che vengano prodotti risultati in genere diversi? Da un punto di vista pratico e strettamente ingegneristico il problema può essere forse facilmente aggirato: si può definire «which are the quantities used for evaluation and, therefore, to state when a new experiment has been able to replicate the original results. These quantities are emerging from the current practice»⁵³. Ci si accontenta cioè di un margine di indeterminazione dei risultati, stabilendo delle soglie che operativamente definiscono le condizioni per parlare di “successo” dell’esperimento. Sul piano filosofico però le conseguenze che se ne possono derivare sono assai rilevanti. In molti ambiti, specie nelle scienze umane, si parla di “esperimenti” (in particolare *fuori* dai laboratori) di cui si ammette che non siano strettamente riproducibili. Ma nel caso dell’I.A., se si ammette la possibilità di esiti *non* previsti, come controllare le possibili conseguenze che vengono ad essere innescate, specialmente sul lungo termine, attraverso decisioni che hanno ricadute tecnologiche e pratiche così estese?

È stato proposto di utilizzare per l’A.I. un modello di esperimento di tipo particolare, quello “esplorativo”⁵⁴, che servirebbe a valutare le potenzialità di un dispositivo per esempio in condizioni che in generale non consentono previsioni abbastanza precise sui risultati – dato anche che spesso gli strumenti informatici servono proprio a potenziare le capacità di calcolo e previsione degli esseri umani⁵⁵. Rimane però il fatto che in questo modo il requisito comunemente accettato in filosofia della scienza della riproducibilità degli esperimenti – e quindi anche della condivisibilità intersoggettiva dei risultati e delle procedure della ricerca – dovrebbe allora essere fortemente indebolito.

In generale è lo stesso test di Turing che dovrebbe in effetti soddisfare un doppio standard: in quanto prodotto ingegneristico, un computer in grado di superarlo dovrebbe essere omologato a requisiti rigorosi e univoci, ma in quanto macchina in grado di “pensare” dovrebbe essere originale e “creativa”. Si può sostenere però che la contraddizione sia ancora più profonda. Si pensi all’originalità richiesta da una creazione artistica oppure alle risposte richieste a uno psicologo in grado di guidare un paziente verso la guarigione. Che uno psicolo-

⁵³ F. Amigoni, M. Reggiani, V. Schiaffonati, *An Insightful Comparison Between Experiments in mobile Robotics and in Science*, in: «Autonomous Robots», 27 (2009), art. 313. <https://doi.org/10.1007/s10514-009-9137-8>.

⁵⁴ V. Schiaffonati, *Experimenting with Computing and in Computing: Stretching the Traditional Notion of Experimentation*, in C. Knibbe (a cura di), *Proceedings of the ECAL 2017*, Lyon, France, 4-8 September 2017, MIT Press, Cambridge, Ma, 2017; V. Schiaffonati, *Philosophy of Engineering and the Quest for a Novel Notion of Experimentation*, in «The Future of Engineering», Springer, Cham, 2018, pp. 89-103; V. Schiaffonati, *Explorative Experiments and Digital Humanism: Adding an Epistemic Dimension to the Ethical Debate*, in H. Werthner, E. Prem, E.A. Lee, C. Ghezzi (a cura di), *Perspectives on Digital Humanism*, Springer, Dordrecht, 2022, pp. 77-82; V. Schiaffonati, *Explorative Experiments: A Paradigm Shift to Deal with Severe Uncertainty in Autonomous Robotics*, in «Perspectives on Science», 30/2 (2022), pp. 284-304.

⁵⁵ S.H. Clearwater, *Artificial Intelligence and Large Scale Scientific Experiments*, in *Proceedings of AAAI Technical Report FS*, 1992, pp. 92-101.

go-macchina possa essere scambiato da un suo paziente per un vero psicologo non implica minimamente che si sia comportato in un modo che giudicheremmo accettabile e soddisfacente per uno psicologo, per esempio riuscendo davvero a portare il paziente verso la guarigione. Nel caso della composizione musicale⁵⁶, artisti come Beethoven o Chopin hanno sicuramente espresso qualcosa, un contenuto, nelle loro opere, che li rappresentava. Ma che un computer possa creare brani musicali “nello stile” di un compositore (come è stato fatto per Stravinsky) *non* significa che sia originale, mentre se un computer fosse in grado di creare musica in grado di esprimere una ricchezza e una originalità di significato come si trovano nei brani di Beethoven o di Chopin (*senza* essere solo una copia del loro particolare stile), ciò *non* mi direbbe che quella musica è espressione di uno stato interiore dell’artista. Se associo la bellezza della Ballata n. 4 di Chopin a un clima emotivo di incanto, di magia e di sospensione della coscienza in uno spazio di tipo fantastico, è legittimo pensare che – se sapessi per assurdo che il pezzo è stato composto da un computer – non mi sentirei più portato a percepire quel brano nello stesso modo⁵⁷. Questo discorso rimanda a un tipo di questioni sollevate nel dibattito tra Dale Jacquette⁵⁸ e Robert French⁵⁹: il test di Turing può essere accusato di essere implicitamente “sciovinista”? L’attribuzione di intelligenza rimanderebbe in effetti alla “forma di vita” che è quella abituale *per il giudice* che deve valutare se essa è legittima oppure no. In direzione opposta, David Chalmers⁶⁰ ha proposto una teoria della mente che ha suscitato un accesso

⁵⁶ C. Ariza, *The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems*, in: «Computer Music Journal», 33/2 (2009), pp. 48-70.

⁵⁷ Per questa via si può contrastare anche la proposta più prudente di Roger Vergauwen e Rodrigo González, che rifacendosi a Putnam (H. Putnam, *The Project of Artificial Intelligence*, in H. Putnam, *Renewing Philosophy*, Harvard University Press, Cambridge, Ma, 1992) sostengono che la verosomiglianza delle simulazioni ottenute da una macchina potrebbe considerarsi adeguata senza arrivare, come ritiene Turing, a una vera «duplication» (R. Vergauwen, R. Gonzalez, *On the Verosimilitude of Artificial Intelligence*, in: «Logique et Analyse», 48/189-192 [2005], pp. 323-350, p. 323) dell’intelligenza umana, ma anche solo «*if and only if they are able to epistemologically persuade human beings that intelligence has been duplicated*» (*ibidem*): non servirebbe la “duplicazione” (ontologicamente e metafisicamente impossibile), ma basterebbe la *persuasione* di una raggiunta duplicazione. Le mie osservazioni vorrebbero mostrare che si può criticare anche questo tipo più circoscritto di richiesta: per la persona comunicare significa essere in contatto con i contenuti di un’esperienza altrui, non con una immagine che soltanto la persuade che vi siano dei contenuti nuovi ottenuti attraverso la comunicazione.

⁵⁸ D. Jacquette, *Who’s Afraid of the Turing Test?* In: «Behavior and Philosophy», 20/2-21/1 (1993), pp. 63-74.

⁵⁹ R.M. French, *Subcognition and the Limits of the Turing Test*, in S. Shieber (a cura di), *The Turing Test. Verbal Behavior as the Hallmark of Intelligence*, MIT Press, Cambridge, Ma, 2004, pp. 183-197. Cfr anche S. Woolgar, *A Coffeehouse Conversation on the Possibility of Mechanizing Discovery and Its Sociological Analysis*, in: «Social Studies of Science», 19/4 (1989), pp. 658-668; P. Slezak, *Scientific Discovery by Computer as Empirical Refutation of the Strong Programme*, in: «Social Studies of Science», 19/4 (1989), pp. 53-600. Cfr anche B. Edmonds, *The Social Embedding of Intelligence. Towards Producing a Machine that Could Pass the Turing Test*, in R. Epstein, G. Roberts, G. Beber (a cura di), *Parsing the Turing Test. Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Springer, Dordrecht, 2008, pp. 211-235.

⁶⁰ D. Chalmers, *The Conscious Mind*, Oxford University Press, Oxford, 1996.

e aspro dibattito, dato che sembrerebbe condurre paradossalmente a identificare come “mente” anche un semplice termostato, in quanto portatore anche esso in qualche senso di “informazione”.

Il *frame* teorico di una logica non-monotona spinge a guardare a queste questioni da un’angolazione particolare, combinando inoltre la questione che può essere definita dello “sciovinismo” (pertinente anche per il dibattito sui processi culturali che accompagnano la diffusione delle tecnologie digitali) con quella vista subito prima della “creatività”. In ogni atto di comunicazione si può assumere che l’intenzione che anima il parlante sia quella di acquisire informazione *nuova*: nei termini del *Tractatus logico-philosophicus*, «È nell’essenza della proposizione la possibilità di comunicarci un senso *nuovo*»⁶¹. Come lo psicoterapeuta è visto dal suo paziente come qualcuno da cui attendersi – in modi che presumibilmente non si sente in grado di prevedere con precisione – che lo aiuterà ad alleviare i propri problemi psicologici, e non come qualcuno che deve solo agire in modo conforme a un copione già noto, in modo da essere “verosimile” e “credibile” come psicoterapeuta, allo stesso modo nel parlare non voglio che il mio interlocutore si limiti a comportarsi in modo conforme a degli *script*, ma desidero che mi dica quello che *lui* pensa su ciò di cui discutiamo. Si noti che se davvero per orientarci nell’esistenza fossero sufficienti gli indizi che compongono le nostre conoscenze sui copioni e sugli *script* che costellano le nostre vite, ne seguirebbe che – dato che questi copioni in quanto tali ci sono per definizione già noti – non sarebbe più necessario (per assurdo) *parlare*. Se il paziente già sapesse che cosa lo psicoterapeuta gli dirà per farlo stare meglio, questo vorrebbe dire probabilmente che quella persona non ha bisogno di uno psicoterapeuta. L’equilibrio tra i contenuti *nuovi* che cerchiamo di ottenere e l’esistenza di codici e prassi già note, che ci servono per poter *capire* i nuovi messaggi, è l’effetto del processo sopra descritto: l’intersoggettività è costruita in modo da includere nuove persone e nuovi punti di vista, ma partendo da esperienze pregresse e già note che ci rendano comprensibili i messaggi che in una tale sfera dell’intersoggettività riceviamo. Se i messaggi fossero *troppo* nuovi, non li capiremmo. Se fossero troppo simili a quelli già noti, non ci servirebbero. Il criterio per valutare quando un tale equilibrio sia raggiunto, ha le sue radici nell’esperienza già disponibile. Si vede quindi come il rischio dello “sciovinismo” sia in effetti presente – nel caso delle macchine, così come purtroppo anche nel rapporto con gli altri esseri umani. Originalità e creatività eccessive da un lato e rischio dello sciovinismo dall’altro sono i due poli entro cui ci si deve necessariamente muovere.

È però necessario considerare la dialettica tra forme di ragionamento non-monotono (individuali) e monotono (in misura variabile più condivise) al fine di potersi orientare in modo non maldestro nel tentativo di trovare volta per volta un punto di equilibrio soddisfacente. Secondo questa prospettiva teorica, chi rifiuta la partecipazione leale e trasparente a un livello più esteso di condivi-

⁶¹ L. Wittgenstein, *Tractatus logico-philosophicus*, Kegan Paul, Trench and Trübner, London, 1922; ed. it. *Tractatus logico-philosophicus e Quaderni*, Einaudi, Torino, 1995, prop. 4.027.

sione sociale (si pensi a fenomeni come quello degli *haters*), lo fa perché le sue esperienze (veridiche o no che siano le generalizzazioni su di esse condotte) lo distolgono dall'aderirvi. Un modello di cultura che voglia affrontare con profitto questi problemi, dovrebbe evitare di limitarsi a ribadire le ragioni di un processo normativo di socializzazione dei risultati del pensare, che in questi casi *ex hypothesi* non si è ancora verificato. Con questo non si vuole dare l'impressione che si ritenga possibile dare indicazioni schematiche che possano essere di una qualsiasi utilità operativa: l'abbozzo di teoria che si è cercato di delineare spinge in direzioni troppo ampie, generali e impegnative per potersi procedere in deduzioni troppo rapide e mirate.

Si ritiene però che servirebbe prima di tutto sviluppare una visione dei rapporti tra individuo e sapere condiviso che sia in grado di tener conto sul piano esplicativo dei molti motivi di tensione che sono operanti nel contesto sociale in cui ci muoviamo. Difendere il significato e il valore della capacità di criticare l'esistente, in particolare di fronte a quella che oggi si configura come una invasione delle nostre vite da parte delle tecnologie della comunicazione digitale, presuppone la definizione di criteri per giudicare una critica come pertinente e plausibile. Ma questo a sua volta – in un quadro strutturalmente complesso e nuovo come quello odierno – presuppone l'adozione di strumenti concettuali che per essere all'altezza dei compiti imposti dal nostro momento storico dovrebbero forse essere oggetto di una radicale revisione. Secondo l'argomentazione che qui ho esposto, lo stesso "test" di Turing, una delle pietre miliari nella filosofia e nella storia dell'informatica, dovrebbe essere visto come espressione di un modello epistemologico – contro il quale possono rivolgersi obiezioni assai importanti – che ha poi ricadute anche nel modo di interpretare i processi culturali e sociali innescati dalle applicazioni dell'informatica e delle ricerche nell'ambito dell'I.A. La scelta teorica di considerare importanti i segni esterni dell'intelligenza, e non i suoi contenuti, corrisponde a un modello culturale per il quale è l'esterno – il "pubblico" – il luogo in cui ciascuno di noi trova e conosce la propria essenza. È però proprio questa concezione che deve essere forse profondamente rivista.