INDAGINI SUL PENSIERO CONTEMPORANEO

SERGIO GALVAN

GÖDEL E IL MODELLO COMPUTAZIONALE DELLA MENTE

Il contributo si articola in tre parti. La prima espone i tratti essenziali del dibattito sui teoremi di Gödel e l'IA iniziato da Lucas e portato avanti da Penrose. Gli aspetti rilevanti del dibattito sono due: da una parte le argomentazioni di entrambi gli autori sono segnate da passaggi poco chiari, inconcludenti e talvolta scorretti: noi ci soffermeremo su alcuni di questi. D'altra parte, però, la tesi difesa da entrambi gli autori ha dalla sua parte elementi di forte intuitività, che la loro argomentazione non riesce a mettere a fuoco, ma che sono tuttavia importanti per rendersi conto dell'impatto negativo che i teoremi di Gödel hanno sul modello computazionale della mente. Si tratta del fatto – fortemente avvertito da tutti quelli che fanno ricerca effettiva nel campo logico e matematico – che la mente umana non funziona nella sua attività ordinaria come una macchina computazionale. Ciò è illustrato nella seconda parte della relazione attraverso un confronto serrato tra alcuni principi che presiedono al funzionamento dei sistemi formali (logica della dimostrabilità) ed altri che presiedono al funzionamento del pensiero (logica epistemica dell'evidenza). Tali principi mettono in evi-

---

* I contenuti del presente articolo sono stati presentati al Convegno "Neurofisiologia e Teorie della Mente", che ha avuto luogo a Milano dal 4 al 5 ottobre 2002 ed è stato organizzato dal Centro di Bioetica dell'Istituto Auxologico Italiano. Si ringrazia il Centro di Bioetica dell'Istituto per aver autorizzato l'autore a pubblicare i contenuti della relazione tenuta al Convegno anche sulla «Rivista di Filosofia Neo-scolastica».

denza una diversità profonda. Come è possibile spiegare tale divaricazione? Nella terza parte saranno svolte alcune considerazioni su tale problema, alla luce di alcune riflessioni svolte da Gödel sui suoi teoremi. Si tratta in particolare delle riflessioni presenti nella Gibbs Conferenza del '51 e in un successivo scritto del '53, contenuti entrambi nel terzo volume dell'*Opera Omnia* (Collected Works) di Gödel pubblicato nel 1995.

1. *Aspetti del dibattito su Gödel e IA*

Dal momento che l'argomento centrale del presente contributo riguarda il nesso tra i teoremi di Gödel e il modello computazionale della mente occorre richiamare in sede preliminare l'enunciato dei due teoremi. Naturalmente i presupposti teorici di entrambi i teoremi sono dati per scontati e noi ci soffermeremo soltanto su quegli aspetti che presentano un interesse particolare per il nostro discorso. Con la sigla G1 indichiamo il primo teorema di Gödel e con G2 il secondo.

G1: PRA sia la teoria formale dell'aritmetica ricorsiva primitiva. Allora esiste una proposizione G(PRA) (detta la gödeliana di PRA) che, sotto la condizione che PRA sia consistente, è intuitivamente vera (vera rispetto al modello standard dell'aritmetica) e, tuttavia, non derivabile in PRA.

$$\text{Cons PRA} \Rightarrow \text{PRA-} \mid\!\!+ \text{G(PRA)}$$

Innanzitutto vale la pena di notare che l'enunciato può essere generalizzato. Al posto di PRA può essere scelta una qualsiasi teoria formale estensione di PRA, di modo che il teorema viene a dichiarare che supposta la consistenza di una qualsiasi teoria formale T estensione di PRA, né in T né *a fortiori* in PRA è derivabile la gödeliana di T G(T). In secondo luogo, PRA formalizza l'idea di procedura matematica finitista. Ciò significa che $\text{PRA-} \mid\!\!- \text{ A} \Leftrightarrow$ A è finitisticamente evidenziabile. Per questo G1 si può parafrasare anche nel modo seguente: esiste una proposizione G(PRA) che, sotto la condizione che PRA sia consistente, è intuitivamente vera (vera rispetto al modello standard dell'aritmetica) e, tuttavia, non finitisticamente evidenziabile.

G2: È un corollario di G1, in quanto deriva da G1 + l'equivalenza di G(PRA) con l'espressione formale della consistenza di PRA (Cons_{PRA}). In breve:

$$\text{Cons PRA} \Rightarrow \text{PRA-} \mid\!\!+ \text{Cons}_{\text{PRA}}$$

In base alle osservazioni precedenti, G2 si può parafrasare anche nel modo seguente: se PRA è consistente, allora, non è finitisticamente evidenziabile che lo sia.

Naturalmente i due teoremi sono importanti *in primis* per la filosofia della matematica e, in particolare, perché segnano il fallimento del programma hilbertiano di fondazione della matematica. Essi tuttavia hanno mostrato aspetti rilevanti anche in settori diversi dalla filosofia della matematica, come la teoria della conoscenza e la scienza dei processi cognitivi. È questo il contesto entro il quale si colloca il dibattito sul significato che G1 e G2 hanno per l'intelligenza artificiale e in particolare per le teorie dei processi cognitivi basate su un modello computazionale della mente. È nostra intenzione affrontare in modo diretto e completo solo questo aspetto della questione, senza entrare nel merito delle tematiche proprie di tali teorie.

Gli autori che hanno dato luogo al dibattito sono J.R. Lucas[1] e R. Penrose[2]. Essi partono dall'assunto che sostenere che la mente sia assimilabile ad una macchina (ovvero che ci sia un modello computazionale della mente) sia equivalente al dire che esiste un sistema formale T che: (i) rappresenta tale macchina e (ii) il conoscere A da parte della mente corrisponde al fatto che A è derivabile in T. Non è nostra intenzione commentare tale assunto, che del resto viene accettato in modo pressoché aproblematico entro il contesto funzionalistico della IA. Ci limitiamo solo a citare due passi assai significativi di Lucas. Egli, ad esempio, scrive nel suo articolo del 1961: "Il teorema di Gödel deve applicarsi alle macchine cibernetiche perché appartiene alla natura di una macchina l'essere una concreta realizzazione di una teoria formale" e poco più avanti: "Se fosse costruita una macchina siffatta da produrre i teoremi dell'aritmetica (per molti aspetti la parte più semplice della matematica), essa avrebbe solo un numero finito di componenti e così sarebbe finito il numero di operazioni che potrebbe compiere e finito il numero di assunzioni iniziali su cui potrebbe operare... Ma se c'è solo un numero finito di tipi di operazioni e di assunzioni iniziali immesse nel sistema, queste possono essere tutte rappresentate mediante opportuni simboli grafici. Tali operazioni si possono far corrispondere a

[1] Cfr. LUCAS (1961). Il contenuto di questo articolo era stato presentato alla Oxford Philosophical Society già nel 1959 e discusso con Putnam a Princeton ancora prima nel 1957. Successivamente LUCAS approfondisce e difende la sua tesi in LUCAS (1968) e in LUCAS (1970). Per un approfondimento delle fonti storiche di questo dibattito e del suo sviluppo cfr. BUTTI (2002). Si veda anche BUTTI (2001).

[2] PENROSE (1989), (1994), (1996).

delle regole (regole di inferenza o schemi d'assioma), che consentono di andare da una a più formule (o anche da nessuna formula) ad altre e le assunzioni iniziali (se ve ne sono) si possono far corrispondere a un insieme iniziale di formule (proposizioni primitive, postulati o assiomi)… Le conclusioni che la macchina sarà capace di produrre come vere, corrisponderanno quindi ai teoremi che si possono dimostrare nel corrispondente sistema formale"[3]. Intento poi dei due autori è di mostrare che i teoremi di incompletezza di Gödel provano che la mente umana non può essere una macchina, dal momento che essa non presenta i limiti che i teoremi attribuiscono ai sistemi formali, vale a dire perché essa si dimostra capace di prestazioni superiori a qualsiasi macchina computazionale pensabile. Sia Lucas che Penrose fanno uso a questo fine di entrambi i teoremi e si muovono nell'arco delle loro riflessioni all'interno del medesimo contesto di fondo. Tuttavia è possibile distinguere nella molteplicità delle loro argomentazioni uno sviluppo che li porta da una sostanziale convergenza in quello che può essere chiamato il 1° argomento (di Lucas in Lucas (1961) e di Penrose in Penrose (1989)) alla formulazione (diversa nei due autori) di un 2° argomento (di Lucas in Lucas (1968) e di Penrose in Penrose (1994)). Cerchiamo di mettere a fuoco l'essenziale dei due argomenti.

Il 1° argomento si basa, indistintamente per Lucas e per Penrose, sul fatto che la mente umana è superiore alla macchina pensante perché mentre la macchina è soggetta a G1 e a G2, la mente umana no: mentre in T non è derivabile G(T) (o equivalentemente $Cons_T$) G(T) (o $Cons_T$) è conosciuta come vera dalla mente. Come abbiamo sopra ricordato, G1 e G2 affermano rispettivamente che, supposta la coerenza di T, in T non è derivabile né G(T) né $Cons_T$. Dal momento poi che alla derivabilità in T corrisponde, in base all'ipotesi di corrispondenza tra meccanismo computazionale da un lato e mente dall'altro, la conoscenza da parte della mente, alla non derivabiltà corrisponderà la mancanza di conoscenza. La mente umana tuttavia ha, secondo i due autori, la capacità di cogliere la verità di G(T) (ovvero della consistenza di T), per cui la superiorità (e quindi la diversità) della mente sulla macchina risulterebbe un dato consequenziale innegabile. A questo argomento si contrappone l'obiezione classica di H. Putnam, che l'autore aveva già esposto in un articolo del 1960 scritto in seguito alla discussione avuta in precedenza con Lucas. Putnam contesta all'argomento di Lucas il fatto che la mente sappia che G(T) è vera. In realtà, la mente umana sa che G(T) è vera sotto la condizione di sapere che Cons T. Rispetto alla conoscenza della consistenza di T la mente umana si trova, però, nella stessa

[3] LUCAS (1961), pp. 113-115.

situazione in cui si trova a proposito della verità di G(T), per cui – non tenendo conto di eventuali ragioni specifiche possedute dalla mente a sostegno della consistenza di T e non condivise dalla macchina – la mente non si trova in una situazione di superiorità rispetto alla macchina per il fatto di conoscere il valore di verità di G(T)[4]. All'obiezione classica di Putnam non è facile replicare. Non è possibile, ad esempio, controbattere dicendo che, per quanto la mente umana si trovi impotente nei confronti del valore di verità di G(T), possa tuttavia cogliere, differentemente dalla macchina, il nesso tra la verità di G(T) e la consistenza di T. Non si può dire ciò, in quanto il nesso tra G(T) e Cons T è colto anche dalla macchina. Infatti, la traduzione formale dell'implicazione Cons T $\Rightarrow$ G(T) è derivabile anche in T, vale cioè T- $\vdash$ $\mathrm{Cons}_T \rightarrow$ G(T)[5]. Una replica più seria consiste nel sostenere che esistono delle ragioni conosciute dalla mente e non dalla macchina per affermare la consistenza di T. Il problema di fondo connesso con questa controobiezione sta però nel fatto che non esiste nessuna giustificazione dell'esistenza di tali ragioni. Non appena, infatti, si prova a chiarire in modo sufficientemente univoco di quali ragioni si potrebbe trattare, si incorre subito nella difficoltà di dover dire che si tratta sì di ragioni collocate a un livello formale superiore (o diverso da) a quello di T, ma in ogni caso trattabili formalmente, in modo tale da assicurare l'accessibilità anche alla macchina, nella misura in cui questa è concepita dinamicamente quale macchina capace, per così dire, di crescere su se stessa, interagendo con l'esterno e acquisendo via via informazioni nuove. Si noti che l'ipotesi di una possibile struttura aperta della macchina computazionale è strettamente connessa con una seconda grave obiezione al 1° argomento di Lucas e Penrose. Si tratta dell'obiezione, sviluppata lucidamente in data recente da Feferman, ma radicata nella teoria assiomatica della verità tarskiana, secondo la quale la verità coincide con la derivabilità in una teoria superiore. Le parole di Feferman sono molto eloquenti al riguardo: "Sulla base delle considerazioni precedenti si può affermare che il pensiero matematico non è prodotto in modo meccanico; sono perciò d'accordo con Penrose riguardo a ciò, ovvero riguardo al fatto che una certa forma di *comprensione* risulta essenziale e che proprio questo aspetto del pensiero matematico non possa essere condiviso con noi dalle macchine. Penrose va però oltre e cerca di rafforzare questa convinzione mostrando che, per il primo teorema di Gödel, il pensiero matematico non può essere *ripresentato* in termini meccanici. A mio parere, invece di rafforzare la convinzione di cui si diceva, tale sforzo fa sorgere più problemi di quanti ne riesca a risolvere e si riduce ad uno sforzo di

[4] PUTNAM (1960), p. 366.

[5] In effetti la derivazione è eseguibile già in PRA: PRA- $\vdash$ $\mathrm{Cons}_T \rightarrow$ G(T).

dialettica senza alcun sbocco"[6]. E più avanti: "L'argomento di Penrose procede più o meno nel modo seguente: come potremmo sapere che F è consistente, se non comprendessimo a che cosa si riferisce – il suo modello inteso – e se quindi non vedessimo che tutti i suoi assiomi sono resi veri da questo modello e che tutte le sue regole *preservano la verità*? È in questo modo, prosegue il ragionamento, che noi possiamo riconoscere la consistenza dei nostri sistemi formali, da PA alla teoria degli insiemi ZF e oltre. E una volta riconosciuta la consistenza di F e accettata come parte dei principi su cui facciamo affidamento, noi vediamo che G(F) è vera e siamo costretti ad accettare per il teorema di Gödel qualcosa che va al di là delle capacità di F … Ci possono però essere altri modi per riconoscere la verità di G(F) oltre a quello di disporre di una nozione generale di verità per il sistema F… Il primo consiste nel ricorso alla dimostrazione in una teoria più potente. Se è vero che tutti ormai riconoscono che il programma hilbertiano mirante a stabilire la consistenza di sistemi formali sempre più potenti su basi finitarie è reso velleitario dal secondo teorema di Gödel, tuttavia una certa forma relativizzata di tale programma si è rivelata adeguata … In questo senso la prova di consistenza di un sistema F può essere scaricata sulla consistenza di un altro sistema F' e si verifica un progresso quando si hanno a disposizione ragioni più forti per accettare F', di quante se ne avevano inizialmente per accettare F… Se da un lato i matematici possono *comprendere* in termini platonici ciò di cui stanno parlando, questi risultati mostrano che tale concezione *non è necessaria* per assicurare fiducia nel corpo della pratica matematica"[7]. Della interpretazione della verità come derivabilità in una teoria superiore e del suo significato per la teoria della mente si parlerà anche nella terza parte del presente scritto. Per ora vale la pena di notare che essa mette chiaramente in crisi la tesi lucasiana della superiorità della mente umana sulla macchina, dal momento che là ove giunge la mente umana può giungere anche la macchina computazionale, almeno nel senso che il risultato colto dalla mente può essere a sua volta ripresentato in forma algoritmica leggibile meccanicamente.

Sono certamente le difficoltà relative al 1° argomento che hanno spinto entrambi gli autori a escogitare un 2° argomento non affetto dalle stesse difficoltà. Gli argomenti forniti da Lucas e da Penrose in questa prospettiva sono diversi. Essi sono tuttavia caratterizzati da una stessa nota. In entrambi non si parte dall'obiettivo di mostrare che la mente è superiore alla macchina perché capace di dimostrare ciò che alla macchina è preclu-

[6] FEFERMAN (1995), p. 8.
[7] FEFERMAN (1995), pp. 8-9.

so, ma da quello di mostrare che la mente non può essere una macchina perché mentre la mente sa o almeno crede d'essere consistente, la macchina non lo può sapere né credere, perché sapere o credere per la macchina significa derivare e ciò è precluso ad essa da G2. Vediamo innanzitutto il 2° argomento proposto da Lucas [8]. Esso si articola in due parti. Nella prima parte si dice semplicemente che se la teoria T è inconsistente allora è diversa da me. Ciò deriva dalla ferma convinzione di Lucas che noi ci riteniamo consistenti. In effetti Lucas manifesta questa convinzione già nell'articolo del 1961[9], ma la sfrutta in modo sistematico solo nel secondo argomento del 1968. Data questa assunzione si può tranquillamente ritenere che:

1. Se T è inconsistente allora T è diversa da me

A questo punto inizia il secondo ramo dell'argomentazione che si conclude con l'affermazione che:

2. Se T è consistente allora T è diversa da me

Da 1. e 2. segue facilmente per esaustione che noi non siamo macchine.

Interessante è dunque vedere come Lucas costruisce il secondo ramo della

---

[8] L'argomento è esposto in LUCAS (1968), p. 152. Noi lo esponiamo tuttavia secondo l'analisi suggerita a Lucas da Putnam, che ne mostra in modo particolarmente lucido la struttura. Per questi particolari si veda l'articolo citato della Butti. Dai lavori della Butti sono presi, con qualche modifica, alcuni dei passi d'autore da noi citati in questo scritto.

[9] Cfr. LUCAS (1961): "Putnam ha suggerito che noi potremmo essere macchine, ma macchine inconsistenti... Il fatto che a volte tutti noi siamo inconsistenti non può essere negato, ma da ciò non segue che siamo equivalenti a sistemi contraddittori. Le nostre inconsistenze corrispondono ad errori piuttosto che a qualcosa di connaturato e stabilito per la nostra natura… Se fossimo davvero macchine inconsistenti, rimarremmo soddisfatti con le nostre inconsistenze e affermeremmo felicemente entrambi i rami di una contraddizione. Inoltre, saremmo disposti a starcene completamente zitti, cosa che invece non siamo. È noto che in un sistema formale contraddittorio si può dimostrare qualsiasi cosa e richiedere che la teoria sia consistente significa appunto richiedere che non si possa dimostrare qualsiasi cosa al suo interno… Questa è sicuramente la caratteristica delle operazioni mentali degli esseri umani: essi sono selettivi: discriminano tra proposizioni vere, che accolgono con favore, e proposizioni false che invece rifiutano: quando una persona è preparata a dire qualsiasi cosa e a contraddirsi, senza alcuna ripugnanza per ciò, allora si dice che ha "perso la testa". Gli esseri umani, anche se non perfettamente consistenti, sono più fallibili che inconsistenti ", pp. 120-121. E più avanti: "Sembra adeguato e ragionevole per la mente asserire la propria consistenza… Non soltanto noi possiamo dire semplicemente che *sappiamo* di essere consistenti, non considerando ovviamente i nostri limiti, ma in ogni caso dobbiamo *assumere* di esserlo, se il pensiero deve essere possibile…", p. 124.

sua argomentazione. Egli naturalmente esordisce con la ripresa di G2: (a) Se T è consistente, allora G(T) è vera. Quindi per introduzione dell'operatore epistemico "vedo", ovvero l'inserimento del condizionale aletico entro un contesto epistemico, si ottiene (b) Vedo che se T è consistente, allora G(T) è vera. Da (b) Lucas deriva poi (c) Se T è consistente, allora vedo che G(T) è vera. D'altra parte, vale anche (d) Se T è consistente, allora T non vede (perché in forza di G2 non può derivare) che G(T) è vera. Dalla combinazione di (c) e di (d) segue infine (e) Se T è consistente allora T è diversa da me.

Il 2° argomento di Lucas presenta una struttura molto solida e senza dubbio più convincente del primo. Tuttavia esso poggia su un assunto iniziale discutibile e contiene inoltre un passaggio scorretto. Innanzitutto è attaccabile l'affermazione che la mente umana sia consistente. Al massimo possiamo essere convinti d'essere *attualmente* consistenti, nel senso che non appena c'accorgiamo d'entrare in contraddizione con qualcosa che per altro riteniamo vero ci affrettiamo a rimuovere le radici della incoerenza che si è appena palesata, ma non nel senso di poter escludere che dal corpo delle nostre credenze possa derivare qualche contraddizione attualmente sconosciuta. Se, però, la convinzione di autocoerenza è così indebolita non possiamo più affermare con categoricità d'essere diversi da una macchina. La distanza, infatti, tra noi e una macchina che agisce secondo le regole d'un formalismo computazionale capace di crescere su di sé e, dunque, di autocorreggersi, non risulta alla fin fine così grande da legittimarne la diversità radicale affermata da Lucas. Ma l'obiezione pù forte all'argomento viene dalla scorrettezza – osservata dallo stesso Putnam – del passaggio da (b) a (c). Da (b) Vedo che se T è consistente, allora G(T) è vera segue, per distribuzione dell'operatore epistemico, (c)' Se vedo che T è consistente, allora vedo che G(T) è vera, ma non (c) Se T è consistente, allora vedo che G(T) è vera.

Anche il 2° argomento proposto in Penrose (1994) e successivamente in Penrose (1996) parte dall'assunto che noi sappiamo (o almeno riteniamo) d'essere consistenti, per cui il formalismo che per ipotesi ci corrisponde dovrebbe essere esso stesso consistente. In questo senso l'argomento è affetto dallo stesso primo limite inerente al 2° argomento di Lucas. Seguiamo nell'esposizione dell'argomento la formulazione dovuta a Per Lindström[10], molto più chiara di quella originale. Con la sigla Sd(T) si intenda esprimere la correttezza del sistema formale T. Lindström non precisa ulteriormente il concetto di correttezza, dichiarando tuttavia esplicita-

[10] LINDSTRÖM (2001).

mente il fatto che la correttezza implica la consistenza. In ogni caso il significato più probabile di un sistema T che è Sd è quello di un sistema che preserva la verità. Con HC(T) si intenda affermare che il sistema T contiene tutti i metodi di dimostrazione matematica accessibili alla mente umana: HC(T) significhi cioè che il sistema T è umanamente completo. Allora l'argomento si sviluppa nel modo seguente:

1. $Sd(T) \Rightarrow Sd(T \wedge SdT)$
   (Se T è corretta allora è corretta pure la teoria che si ottiene aggiungendo a T l'affermazione della sua correttezza)

2. $Sd(T \wedge SdT) \Rightarrow G(T \wedge SdT)$
   (Se una teoria è corretta allora è consistente, ma se è consistente allora è vera la sua gödeliana)

3. $Sd(T \wedge SdT) \Rightarrow T \wedge SdT \mid\!+ G(T \wedge SdT)$ per G1

4. $SdT \Rightarrow G(T \wedge SdT)$ per concatenazione

5. $SdT \Rightarrow T \wedge SdT \mid\!+ G(T \wedge SdT)$ per concatenazione

6. $SdT \Rightarrow T \mid\!+ SdT \rightarrow G(T \wedge SdT)$ per Introduzione di $\rightarrow$

7. $HC(T) \Rightarrow T \mid\!- SdT \rightarrow G(T \wedge SdT)$ per def. di HC e passo 4.

8. $\neg(SdT \wedge HC(T))$ per refutazione

9. Io non sono T perché io sono consistente, mentre T non è tale.

Possiamo dire che così rigorizzato il 2° argomento di Penrose è corretto e almeno condizionatamente (cioè sotto la condizione della consistenza della mente umana) conclusivo? Lindström solleva innanzitutto la questione della correttezza del primo passo. Egli afferma che se la correttezza si intende, ad esempio, come correttezza rispetto alle $\prod_1$-formule, allora essa è equivalente alla consistenza. Ma allora, dal momento che Cons(T $\wedge$ non Cons(T)) e tuttavia non Cons[(T $\wedge$ non Cons(T)) $\wedge$ Cons(T $\wedge$ non Cons(T))] non può valere che $Sd(T) \Rightarrow Sd(T \wedge SdT)$[11]. Naturalmente, per evitare que-

[11] LINDSTRÖM (2001), p. 246.

sto esito, si può intendere la proprietà di correttezza in un senso più forte, per esempio come predicato definibile nel linguaggio di T che vale di T se T è vera. In tal caso però si incorre in una serie di altre difficoltà, evitabili solo se Sd è inteso nel senso pieno del predicato di verità. Questo, tuttavia, per il teorema di indefinibilità di Tarski non può essere definito nel linguaggio di T, il che mina dalle fondamenta la possibilità di formulare correttamente passaggi cruciali dell'argomento come il passo 7[12]. Pare in ogni caso che la conclusione dell'argomento raccolta nel passo ottavo sia legittimabile già intuitivamente a partire dal significato generale dei teoremi di limitazione. Proprio per il fatto che i teoremi di limitazione mettono chiaramente in luce la struttura aperta del nostro procedere dimostrativo, è impossibile che esista un sistema *formale* al contempo *corretto* e umanamente *completo*.

L'analisi del 2° argomento di Penrose appena svolta ci consente, a questo punto, di tirare un po' le fila della prima parte del presente contributo. L'argomento nella formalizzazione di Lindström costituisce la versione più avanzata e rigorosa degli argomenti che Lucas e Penrose propongono a confutazione della teoria meccanicistica della mente a partire dai teoremi di Gödel. Eppure anche l'argomento di Lindström/Penrose presta il fianco a obiezioni decisive. Quale può essere la ragione di questa situazione? La fine dell'articolo di Lindström è molto eloquente al riguardo: "Se uno o l'altro di questi argomenti possa essere modificato, in modo da generare un argomento conclusivo rimane, almeno in teoria, una questione aperta… In effetti può essere che la questione della meccanizzabilità del ragionamento matematico sia posta malamente e che così posta non possa avere una risposta ben definita". Siamo inclini a ritenere che il problema di fondo stia proprio nel fatto che la domanda stessa circa il rapporto tra la mente umana e la macchina sia mal posta. Non pare, infatti, sensato porre la questione nei termini di un argomento che metta in luce la superiorità della prima sulla seconda, in quanto, come dice Feferman, non si danno contenuti di conoscenza matematica intuitiva che non possano essere *ripresentati* come le conseguenze interne (i teoremi) di una teoria formale. La conoscenza matematica si manifesta sempre attraverso qualche sistema formale, nel senso che la verità rispetto al modello va sempre di pari passo con la derivabilità in qualche teoria superiore. Ci si può tuttavia chiedere: Per quanto la diversità tra mente e macchina non si possa determinare nella forma della superiorità dell'una sull'altra, non esiste, purtuttavia, qualcosa di vero nelle tesi dei non computazionalisti, consistente nel fatto – fortemente avvertito

[12] LINDSTRÖM (2001), p. 246.

da tutti quelli che fanno ricerca effettiva nel campo logico e matematico – che la mente umana non funziona nella sua attività immediata come una macchina computazionale? Anche se, per ipotesi, gli obiettivi raggiungibili dalla mente umana potessero essere conseguiti da una macchina, non sarebbe forse importante il fatto che il modo di procedere della mente è diverso da quello della macchina? Il pensare matematico appare in effetti, almeno a prima vista, come un vedere e non come un computare. Anche se la verità che si vede si può esplicitare come qualcosa di dimostrabile, non è facile convincersi che alla fin fine il vedere sia una *illusione*, generata dal fatto che, in ultima istanza, il pensare è una sorta di *computazione contratta*. Ebbene, se le cose stanno in questi termini, allora l'uso dei teoremi di Gödel ai fini dello stabilire l'eventuale diversità tra mente e macchina non è quello di servirsene per dimostrare la superiorità della mente rispetto alla macchina, ma è quello di vedere se attraverso di essi è possibile mettere effettivamente in evidenza una *diversità irriducibile di funzionamento* della mente rispetto al computer. Ma come è possibile servirsene in questa prospettiva? A questo fine, non basta sapere come funziona un computer ma occorre conoscere, per lo meno in modo altrettanto soddisfacente, come funziona la mente quando fa matematica. Ora, i principi fondamentali che presiedono al funzionamento di una macchina formale sono naturalmente oggetto di studio della metamatematica formale (ovvero della logica della dimostrabilità) e i teoremi di Gödel ne sono tra i più rappresentativi. Ma al giorno d'oggi ha fatto qualche passo in avanti anche un particolare settore della logica intensionale, la cosiddetta logica epistemica[13] (di cui fa parte anche la logica della dimostrabilità), la quale ha in oggetto le leggi relative alle modalità tipiche del conoscere umano. Della logica epistemica fa, poi, parte anche la logica dell'evidenza[14] che è particolarmente importante nel confronto che intendiamo impostare tra funzionamento della mente (logica dell'evidenza) e funzionamento della macchina (logica della dimostrabilità). Intento del prossimo capitolo è quello di stabilire un confronto serrato tra quattro principi della logica dell'evidenza e i corrispondenti principi della logica della dimostrabilità. L'analisi di questi principi metterà in evidenza la loro irriducibile diversità.

---

[13] Per quanto riguarda la logica della dimostrabilità cfr. SMORYNSKI (1977), BOOLOS (1993), GALVAN (1982), GALVAN (1992); per quanto riguarda la logica epistemica in generale cfr. HINTIKKA (1962), LENZEN (1980), KUTSCHERA (1982), GALVAN (1991), pp. 211-272, GIORDANI (2002).

[14] Cfr. GALVAN (1991), pp. 243-272, GALVAN (2001), pp. 81-97, GIORDANI (2000), GIORDANI (2002), pp. 105-122 e pp. 192-202.

2. *Logica epistemica e logica della dimostrabilità*

La logica della dimostrabilità studia le leggi dell'operatore di dimostrabilità Pr_T (è dimostrabile in T che…), mentre la logica dell'evidenza studia le leggi dell'operatore di evidenza *E* (è evidente che…). L'operatore di dimostrabilità è unico, in quanto coincide con il predicato di dimostrabilità definibile canonicamente entro il linguaggio aritmetico di PRA e quindi tale da soddisfare le cosiddette tre condizioni di derivabilità[15]. Per questo, i principi che regolano l'operatore Pr_T saranno presentati nella forma di teoremi di PRA. Poiché non esiste un unico concetto di evidenza, di operatori dell'evidenza ne esistono invece molti e questi sono caratterizzati da sistemi assiomatici diversi[16]. Conseguentemente certe leggi sono specifiche di certe accezioni dell'operatore di evidenza ed altre sono comuni. Nel seguito prenderemo in considerazione sia leggi comuni sia leggi specifiche.

(i) *Principio di consistenza*:

È un principio comune a tutti i sistemi di logica dell'evidenza (come del resto anche della logica della credenza) il principio di *E*-consistenza:

$$\vdash \neg E(\bot)$$

secondo il quale è logicamente vero che non si può portare ad evidenza la contraddizione. Al contrario, nella logica della dimostrabilità (della macchina) non vale, in forza di G2, lo stesso principio:

$$\text{PRA-} \ \vdash\!\!+ \neg \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner\bot\urcorner)$$

vale a dire, non è finitisticamente evidenziabile (vero) che non si possa dimostrare la contraddizione in PRA.

(ii) *Principio di riflessività 1*:

E sia l'operatore dell'evidenza incontrovertibile. Per evidenza incontrovertibile intendiamo una evidenza che non può essere illusoria o ingannevole e perciò è indubitabile. Si tratta di quella forma di evidenza che lascia apparire esattamente ciò che è (ciò che appare = ciò che è) e che pertanto viene negata da coloro che richiedono all'evidenza di soddisfare il requisito dell'informatività. Per questi, infatti, tutte le forme d'evidenza informativa

---

[15] Cfr., ad esempio, SMORINSKI (1977) e GALVAN (1992), pp. 127-155.
[16] Cfr. nota 13.

sono esposte al pericolo dell'illusione e perciò non possono essere incontrovertibili. In ogni caso il fenomeno dell'evidenza di qualcosa che si riduce a essere il contenuto dell'apparire si dà. È l'evidenza di "mi appare proprio questo che mi appare". Chiaramente si tratta di una forma d'evidenza non informativa e perciò tautologicamente incontrovertibile. Tuttavia in questa sede è rilevante il fatto che esista. In sintesi si ha il seguente principio di riflessività 1:

$$\vdash EA \rightarrow A$$

Al contrario, nella logica della dimostrabilità non vale il principio di riflessività 1:

$$\text{PRA-} \ \vdash + \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner A \urcorner) \rightarrow A,$$

anche se il principio è derivabile in teorie più potenti di PRA come PA.

(iii) *Principio di riflessività 2*:

Si tratta del principio di riflessività che è contenuto esso stesso di metaevidenza. Tale principio è comune a molti sistemi, dal momento che esprime l'affidabilità degli stati d'evidenza per il soggetto dell'evidenza (dunque un'affidabilità in prima persona), il che non è escluso neppure nei casi di evidenza illusoria: è sempre evidente a un soggetto che le sue evidenze sono affidabili (in prima persona), anche se in realtà tali evidenze possono essere (in terza persona) illusorie. Nella logica dell'evidenza il principio assume la seguente forma:

$$\vdash E(EA \rightarrow A)$$

Al contrario nella logica della dimostrabilità si ha:

$$\text{PRA-} \ \vdash + \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner A \urcorner) \rightarrow A \urcorner),^{17}$$

anche se $\text{PRA-} \ \vdash \mathrm{Pr}_{\mathrm{T}}(\ulcorner \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner A \urcorner) \rightarrow A \urcorner)$, ove T è una opportuna estensione di PRA.

(iv) *Principio di ω-completezza:*

E veniamo ora al principio più importante: il principio di ω-completezza. Questo principio ha a che fare naturalmente con linguaggi numerici, vale a

[17] In realtà si ha l'enunciato ancora più forte: $\text{PRA-} \ \vdash \neg \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner A \urcorner) \rightarrow A \urcorner)$.

dire con linguaggi in cui compaiono i nomi per i numeri naturali. Usualmente tali nomi sono le cifre 0,1,2,3 … oppure sono i termini numerali, cioè i termini che si ottengono per applicazione passo dopo passo della funzione di successione (usualmente indicata con il segno ') a partire dalla cifra 0. Così, ad esempio, si ha $\overline{1} =_{def} 0'$, $\overline{2} =_{def} 0''$, … L'n-esimo numerale è indicato attraverso il segno $\overline{n}$. Ebbene, in che cosa consiste il principio di ω-completezza per l'operatore d'evidenza *E*? Esso si può esprimere attraverso tre formulazioni diverse, di cui le prime due – solo notazionalmente diverse – ne rappresentano la versione semiformale non finitaria, mentre la terza formulazione rappresenta la versione formale finitaria:

1. $E(A(0)), E(A(\overline{1})), \ldots. \rightarrow E\forall xA(x)$

2. $\forall nEA(\overline{n}) \rightarrow E\forall xA(x)$

3. $\forall xEA(x) \rightarrow E\forall xA(x)$

È chiara la differenza tra le prime due formulazioni e la terza. Le prime due contengono un antecedente costituito da una formula infinita (l'antecedente della seconda formulazione è semplicemente un accorciamento dell'antecedente della prima formulazione). La terza, invece, contiene un antecedente costituito da una formula finita. Il significato delle tre versioni è però sempre lo stesso, in quanto il dominio di interpretazione della variabile x è l'insieme dei numeri naturali standard, vale a dire l'insieme degli oggetti a cui corrispondono esattamente i termini numerali. Data l'invarianza del dominio di interpretazione, le tre versioni sono dunque equivalenti. Così il principio afferma che se è evidente per ogni numero standard n (immagine del numerale $\overline{n}$) che A(n) allora è evidente che $\forall xA(x)$ (che tutti i numeri standard siano A). Si notino con attenzione i significati dell'antecedente e quello del conseguente del principio. Nell'antecedente si dice che esiste la possibilità di portare ad evidenza per ogni singolo numero naturale standard che di esso vale A. Naturalmente è da chiedersi quali siano i requisiti perché esista tale possibilità. Essendo escluso che questa sia data in quanto la mente è capace di passare in rassegna uno per uno tutti i casi – impossibilità legata al fatto che per la mente umana *infinitum non potest pertransiri* –, il requisito necessario e sufficiente consiste nel presupporre che sia attraverso la struttura di x (l'essere x un naturale standard, immagine di qualche numerale $\overline{n}$) che risulta evidente che x sia A. In definitiva non rimane altro modo di intendere l'antecedente se non nel senso che è evidente in forza della *struttura generale e astratta* di numero naturale standard (la sua intensione) che ogni singolo numero standard x è A. E qual è il significato del conseguente? Il significato del conseguente è immediato.

Esso significa che è evidente che tutti gli x, ovvero che tutti i numeri standard, sono A. Il contenuto di tale evidenza non è dunque dato dai singoli casi, per quanto l'evidenza dei singoli casi venga dall'evidenza della struttura, ma dallo stato di cose complessivo consistente nel fatto che la totalità dei numeri naturali standard è caratterizzata dal predicato A. Una volta determinato il significato univoco di antecedente e conseguente la giustificabilità del principio appare in modo immediato. Se per ogni singolo numero naturale x esiste la possibilità di portare ad evidenza che x è A, allora è evidente che tutti i numeri naturali standard sono A[18].

Per quanto sia immediato il principio di ω-completezza nella logica dell'evidenza, esso non vale nella logica della dimostrabilità. Anzi è un teorema della metamatematica post-gödeliana che l'ω-completezza è addirittura equivalente alla inconsistenza. Innanzitutto i due seguenti schemi sono gli equivalenti delle due prime versioni del principio di ω-completezza:

1. $\mathrm{Pr}_T(\ulcorner A(0)\urcorner) \wedge \mathrm{Pr}_T(\ulcorner A(\overline{1})\urcorner) \wedge \mathrm{Pr}_T(\ulcorner A(\overline{2})\urcorner), \ldots \rightarrow \mathrm{Pr}_T(\ulcorner \forall x A(x)\urcorner)$

2. $\forall n \mathrm{Pr}_T(\ulcorner A(\overline{n})\urcorner) \rightarrow \mathrm{Pr}_T(\ulcorner \forall x A(x)\urcorner)$

Sia ora $\forall x \mathrm{Pr}_T(\ulcorner A(\dot{x})\urcorner) =_{\mathrm{def}} \forall x \mathrm{Pr}_T(\mathrm{so}(\ulcorner A(x)\urcorner, \ulcorner x\urcorner, \mathrm{num}(x))$ l'espressione formale del fatto che è dimostrabile in T per ogni numerale $\overline{n}$ che $A(\overline{n})$. Allora l'equivalente della terza versione del principio di ω-completezza è:

3. $\forall x \mathrm{Pr}_T(\ulcorner A(\dot{x})\urcorner) \rightarrow \mathrm{Pr}_T(\ulcorner \forall x A(x)\urcorner)$

---

[18] L'immediatezza del principio di ω-completezza viene conservata anche se l'antecedente è inteso in un senso diverso rispetto al senso della lettura "intensionale" che noi ne abbiamo dato. Le interpretazioni alternative sono nella sostanza due. Innanzitutto vi è l'interpretazione fondata sulla concezione attualistica degli stati di evidenza, secondo la quale è possibile che gli stati di evidenza siano dati attualmente in quantità infinita. Si tratta di una concezione massimamente idealizzante della mente. Questa è considerata capace di considerare attualmente una infinità di casi presi uno ad uno. È ovvio che in tale ipotesi il conseguente del principio sia immediatamente implicato dall'antecedente. La seconda interpretazione è fondata – come quella intensionale – su una concezione potenziale della successione degli stati di coscienza. Tuttavia si differenzia da quest'ultima, in quanto, come la prima, parte da una idealizzazione molto forte della mente in un altro senso. L'idealizzazione ora riguarda la durata nel tempo. Si ipotizzi, ad esempio, che una mente esista da sempre. Ciò vuol dire che, nel corso della sua esistenza senza inizio quella mente ha avuto, per così dire, il tempo di attualizzare l'infinità degli stati d'evidenza presupposti dall'antecedente del principio. È ovvio così che anche nell'ipotesi di una durata infinita della mente il conseguente del principio segua immediatamente dall'antecedente. Entrambe le interpretazioni prese in considerazione nella presente nota sono tuttavia da escludere, in quanto poggiano su ipotesi eccessivamente idealizzanti della mente, che non tengono conto della finitezza propria della mente umana.

Ebbene, non solo le prime due versioni sono inderivabili entro PRA, ma la terza è addirittura equivalente in PRA con l'espressione formale della inconsistenza di T. Questo fatto è metamatematicamente molto significativo, perché la terza formulazione dello schema di di ω-completezza è l'espressione finitaria delle due precedenti versioni, caratterizzate entrambe da un antecedente che è una formula infinita. Questa circostanza non solo sbarra immediatamente la strada all'obiezione che l'inderivabilità delle due prime versioni in PRA è poco rilevante, perché si tratta di espressioni infinitarie e dunque finitisticamente insignificanti, ma consente di stabilire un perfetto parallelo con la terza versione del principio di ω-completezza della logica dell'evidenza. Là l'interpretazione "intensionale" ci consentiva di dare un significato plausibile alla possibilità di portare ad evidenza il fatto che A valga di tutti i numeri naturali standard presi uno ad uno. Tale possibilità sarebbe garantita dal fatto che l'evidenza di A(x) viene colta attraverso la struttura di x. Qui la situazione è perfettamente analoga. Che sia dimostrabile A per ogni numero che sia immagine di un numerale lo si può eventualmente stabilire in quanto si mostra $Pr_T(\ulcorner A(\dot{x})\urcorner)$, vale a dire mostrando che per ogni n vale $A(\overline{n})$ in forza della struttura di $\overline{n}$. A conclusione di tutto ciò, è particolarmente significativo, che, mentre il principio di ω-completezza sia valido per l'operatore epistemico dell'evidenza, esso risulti al contrario equivalente con l'inconsistenza di T per l'operatore di dimostrabilità Pr_T.

Osservazione 1. In realtà, la frattura tra antecedente e conseguente nel principio di ω-completezza per Pr_T è superabile se l'operatore di dimostrabilità nel conseguente è rafforzato, se, cioè, Pr_T diventa Pr_T', ove T' è una estensione essenziale e non conservativa di T. Per esempio, T' potrebbe essere l'estensione di T che si ottiene aggiungendo a T la definizione della verità per le formule di T e l'induzione estesa al nuovo predicato di verità, oppure T' potrebbe essere T + $Cons_T$. In entrambi i casi il principio $\forall x Pr_T(\ulcorner A(\dot{x})\urcorner) \rightarrow Pr_T'(\ulcorner \forall x A(x)\urcorner)$ sarebbe un teorema di PRA. Lungi dal colmare la distanza tra logica dell'evidenza e logica della dimostrabilità, questo fatto però ribadisce l'essenzialità della differenza tra l'ω-completezza per *E* e l'ω-completezza per Pr_T: mentre nel passaggio da antecedente a conseguente nello schema di ω-completezza per *E* non occorre nessun incremento d'informazione (d'evidenza) (il contenuto d'evidenza racchiuso nell'antecedente è sufficiente per garantire il contenuto d'evidenza racchiuso nel conseguente; ovvero, l'evidenza che $\forall x A(x)$ è conseguenza esclusiva dall'evidenza che A(x) per ogni x), nel passaggio dall'antecedente dello schema di ω-completezza per Pr_T al conseguente è necessario un incremento d'informazione (alla macchina per affermare $\forall x A(x)$ non basta sapere che $A(\overline{n})$ è ottenibile in T per ogni n, ma le occorre sapere qualcosa su T, cioè che T è corretta).

Osservazione 2. La diversità tra logica dell'evidenza e logica della dimostrabilità è una questione anche di contenuti. La diversità dei principi si riflette sui contenuti e la diversità dei contenuti si riflette sui principi. Si rifletta un po' sul contenuto delle evidenze coinvolte nel principio di ω-completezza per *E*. Nella formulazione del principio si è fatto costante riferimento al modello standard dei numeri naturali. Ebbene, la validità del principio nella logica dell'evidenza e il suo cadere nella logica della dimostrabilità è in fondo imputabile al fatto che l'evidenza in gioco è l'evidenza del modello standard dei numeri naturali e tale modello non può essere caratterizzato formalmente. In altre parole, che la circostanza che per ogni x sia evidente A(x) garantisca che sia evidente $\forall x A(x)$ è implicato proprio dal fatto che il concetto di numero naturale (ovvero di insieme dei naturali) presente alla mente è quello standard (ossia quello originato da 0 e dalla operazione ricorsiva di successione). Ora, non esiste nessuna teoria formale capace di definire formalmente il concetto di numero naturale. Vale a dire, non esiste nessuna teoria formale, per quanto questa possa essere potente, in cui sia presente una formula N(x) interpretabile categoricamente (cioè a meno di isomorfismi) sull'insieme dei numeri naturali. Il resto è consequenziale: se la teoria formale non ha presa sul concetto di numero, non può garantire il passaggio dalla dimostrabilità di $A(\overline{n})$ per ogni singolo numerale alla dimostrabilità della formula universale; non lo può fare perché la teoria formale non sa, non riuscendo a definirlo, che il concetto di naturale coincide con quello di immagine di un numerale, ovvero che l'insieme dei naturali è esaurito da quello delle immagini dei numerali.

Ma, si obietterà, il concetto di numero naturale è definibile al secondo ordine. Dunque, il limite messo in luce dalla logica della dimostrabilità a proposito di T vale solo per teorie formali del primo ordine. All'obiezione è facile replicare. Non esistono teorie *formali* al secondo ordine non riducibili al primo ordine. Dunque quello che formalmente è definibile al secondo ordine lo è altrettanto al primo ordine. Ma al primo ordine il concetto di naturale è formalmente indefinibile, pertanto è tale anche al secondo ordine. Ma allora non esiste nessuna differenza tra primo ordine e secondo ordine? La differenza c'è ed è profonda, ma non è una differenza tra teorie formali. È una differenza tra semantiche, ovvero tra modelli. Un modello aritmetico al secondo ordine (che sia modello di un linguaggio aritmetico al secondo ordine o di una teoria aritmetica al primo è indifferente) è un modello in cui esiste l'insieme di tutte le proprietà/relazioni dei numeri naturali (modello principale). Per questo, se si decide di associare a un opportuno linguaggio (del primo o del secondo ordine indifferentemente) modelli di questo tipo, allora è facile mostrare in base al teorema di Dedekind che tali modelli sono isomorfi. Ma decisivo è il fatto che, se una teoria è formale, essa ammette anche modelli non principali normali

(secondari). Ebbene, sono per l'appunto i modelli secondari ammessi dalle teorie in quanto formali responsabili del fatto che in tali teorie non sia definibile formalmente il concetto di numero naturale. In modo del tutto equivalente, come non è caratterizzabile formalmente il concetto di numero naturale, così non sono caratterizzabili formalmente i modelli principali. In altre parole, i modelli principali sono modelli *intesi*, esattamente come il modello standard dell'aritmetica al primo ordine, a cui si è fatto riferimento nel semantizzare il principio di ω-completezza dell'operatore di evidenza *E*. In conclusione, l'appello alle teorie del secondo ordine, lungi dal costituire lo strumento per colmare il divario tra logica dell'evidenza e logica della dimostrabilità, viene, al contrario, a confermare la diversità tra di esse.

Osservazione 3. L'applicazione dello schema di ω-completezza nella logica dell'evidenza alla formula aritmetica esprimente la nozione di consistenza consente di chiarire con un alto grado di precisione in che senso l'evidenza (o intuizione) costituisca un superamento delle prestazioni di cui è capace una macchina, senza che questo significhi che la mente è capace di cogliere la verità di proposizioni inaccessibili alla macchina. Attraverso tale applicazione diventa così possibile intravvedere, negli argomenti di Lucas e di Penrose, un'anima di verità in mezzo a un corpo caduco. $\neg Prov_{PRA}(x,\ulcorner\bot\urcorner)$ sia l'espressione aritmetica della relazione secondo la quale x non è dimostrazione in PRA della contraddizione $\bot$. $\forall x\neg Prov_{PRA}(x,\ulcorner\bot\urcorner)$ significa allora che non esiste dimostrazione della contraddizione in PRA, ovvero che PRA è consistente. $\forall x\neg Prov_{PRA}(x,\ulcorner\bot\urcorner)$ sta dunque per $Cons_{PRA}$. Pr_{PRA} sia il solito operatore della dimostrabilità. $Pr_{PRA}(\ulcorner A\urcorner)$ stia dunque per $\exists xProv_{PRA}(x,\ulcorner A\urcorner)$. Ora, in base al cosiddetto lemma di Feferman, è dimostrabile in PRA[19] la proposizione:

$$(\alpha) \qquad \forall xPr_{PRA}(\ulcorner\neg Prov_{PRA}(\dot{x},\ulcorner\bot\urcorner)\urcorner),$$

cioè l'espressione dichiarante che è un teorema di PRA che non esiste nessun numero standard (immagine di un termine numerale) che sia il codice di una dimostrazione della contraddizione in PRA. Inoltre, all'inizio del presente scritto abbiamo detto che la dimostrabilità in PRA si può considerare criterio per l'evidenza finitista: ciò che è dimostrabile in PRA è finitisticamente evidente e ciò che è finitisticamente evidente è dimostrabile in PRA. L'evidenza finitista è però una particolare forma di evidenza. Da (α) si può dunque ottenere

[19] Cfr. GALVAN (1992), p. 176.

(β) $\forall x E(\neg \mathrm{Prov}_{\mathrm{PRA}}(x, \ulcorner \bot \urcorner))$

Ebbene. In base alla seguente istanza dello schema di ω-completezza: $\forall x E(\neg \mathrm{Prov}_{\mathrm{PRA}}(x, \ulcorner \bot \urcorner)) \rightarrow E(\forall x \neg \mathrm{Prov}_{\mathrm{PRA}}(x, \ulcorner \bot \urcorner))$, si ha anche:

(γ) $E(\forall x \neg \mathrm{Prov}_{\mathrm{PRA}}(x, \ulcorner \bot \urcorner))$

ovvero, in base alle considerazioni svolte sopra,

$$E(\mathrm{Cons}_{\mathrm{PRA}})$$

A questo punto il lettore affermerà. Ma allora la mente riesce effettivamente a dimostrare $\mathrm{Cons}_{\mathrm{PRA}}$ e dunque è superiore alla macchina. La conclusione è tuttavia affrettata. Noi abbiamo ottenuto che è evidente la consistenza di PRA, ma non abbiamo ottenuto al contempo che la consistenza di PRA è vera. Per ottenere la verità della consistenza occorre impegnarsi ad accettare che l'evidenza di $E(\mathrm{Cons}_{\mathrm{PRA}})$ – che non è una evidenza finitista – è tuttavia riflessiva, ossia tale da implicare la verità dei suoi contenuti. Questa conclusione è altamente giustificata – perché prende le mosse da evidenze finitiste, sfrutta il principio di ω–completezza per E, a sua volta fondato sulla evidenza del modello standard –, ma non si tratta in ogni caso di conclusione giustificata in modo incondizionato, perché non si tratta di proposizione oggetto di evidenza puramente finitista.

In conclusione, l'ultima osservazione ribadisce e riprende il motivo centrale di tutta questa sezione. La comparazione tra alcuni significativi principi della logica dell'evidenza e altri della logica della dimostrabilità manifesta una profonda diversità di funzionamento della mente rispetto alla macchina. Questo fatto non costituisce di per sé una prova di superiorità della mente sulla macchina, in quanto ciò non implica che determinate verità possono essere conosciute dalla mente e risultino tuttavia inaccessibili alla macchina. Ci dice soltanto che la mente segue *strade proprie nella individuazione* di tali verità e – cosa ancora più importante – che essa segue *modalità proprie nella giustificazione* di esse, perché in questa impresa non può limitarsi a fare uso delle sole evidenze finitiste. In ogni caso questa discontinuità pone la seguente serie di problemi: "Come può essere spiegata la diversità di funzionamento di una mente rispetto a una macchina?" Esiste una spiegazione di tale dualismo operativo e, se esiste, qual è? Si tratta di una dicotomia apparente oppure è radicata nella realtà delle cose? E se è così, quali ne sono le conseguenze?" Fino a che non si trovano risposte plausibili a tali domande, le difficoltà che la prospettiva metamatematica adossa al modello computazionale della mente risultano decisamente gravi e non facilmente eludibili. Nella prossima sezione avanzeremo qual-

che ipotesi in ordine alla soluzione di tale problematica e la discuteremo alla luce delle riflessioni gödeliane sopra richiamate.

3. *Logica epistemica, logica della computabilità e Gödel*

Gödel prende le mosse dal significato limitativo dei suoi teoremi. Questo consiste nel fatto che l'edificio del sapere matematico non si può racchiudere entro l'ambito della matematica finitista, vale a dire non può essere fondato su un insieme di procedure assolutamente indubitabili. Ciò è all'origine dell'alternativa seguente caratteristica dell'intera riflessione gödeliana sul problema. O alla mente umana è consentito l'accesso alle verità matematiche che non possono essere trattate mediante procedure finitarie e in questo caso la mente umana non può essere una macchina, oppure la mente è uguale ad una macchina e, in tal caso, le verità matematiche che trascendono l'orizzonte delle evidenze finitarie risultano inconoscibili alla mente umana[20]. Da queste due alternative si possono ricavare due linee di risposta all'interrogativo posto sopra[21].

---

[20] Cfr. (CW, III, 51, p. 310): "Così è inevitabile la seguente disgiunzione: o la matematica è incompleta, nel senso che i suoi assiomi evidenti non possono mai essere compresi in una regola finita, cioè la mente umana (persino entro l'ambito della matematica) sorpassa infinitamente la potenza di qualsiasi macchina finita, oppure esistono problemi diofantei del tipo specificato assolutamente indecidibili (ove non è escluso il caso che entrambi i termini della disgiunzione siano veri, cosicchè le alternative siano strettamente parlando tre)". Anche nelle note successive i testi gödeliani saranno citati attraverso la sigla formata da CW che sta per "Collected Works", III che sta per il terzo volume, anno di pubblicazione del testo, numerazione delle pagine.

[21] Cfr. (CW, 51, pp. 311-312): "In modo corrispondente alla forma disgiuntiva del teorema principale sull'incompletabilità della matematica, le implicazioni filosofiche sono *prima facie* disgiuntive esse stesse: tuttavia tutte queste implicazioni sono decisamente in contrasto con una filosofia materialistica. Infatti, se vale la prima alternativa, questo sembra implicare che l'attività della mente non possa essere ridotta all'attività del cervello, il quale ha tutte le sembianze di una macchina finita con un numero finito di parti, vale a dire i neuroni e le loro connessioni. Così si è presumibilmente indotti ad accettare qualche punto di vista vitalistico. D'altra parte, la seconda alternativa, secondo la quale esistono proposizioni matematiche assolutamente indecidibili, sembra invalidare la concezione secondo la quale la matematica è solo una nostra creazione; infatti il creatore conosce necessariamente tutte le proprietà delle sue creature, perché queste non possono averne altre da quelle ricevute. Così questa alternativa sembra implicare che gli oggetti e fatti matematici (o almeno qualcosa in essi) esistono oggettivamente e indipendentemente dai nostri atti e decisioni mentali, vale a dire, sembra implicare qualche forma di Platonismo o 'realismo'nei confronti degli oggetti matematici. Infatti, l'interpretazione empirica della matematica, cioè la visione che i fatti matematici sono un tipo speciale di fatti fisici o

Secondo la prima, la diversità di funzionamento dipende dal fatto che la mente ha delle risorse che non sono disponibili alla macchina. Si tratta della capacità di cogliere attraverso forme di evidenza più impegnative rispetto a quelle in gioco in PRA contenuti astratti e non finiti che non sono intuibili attraverso forme d'evidenza finitista[22]. In tal caso si arriva alla conclusione che mente ≠ macchina. Gödel dà una versione fortemente fenomenologica di questa spiegazione, verso la quale nutre una netta preferenza rispetto a quella alternativa[23]. Tale preferenza si basa, tra il resto, sul

---

psicologici, è troppo assurda per essere seriamente sostenuta. Non si sa se vale la prima alternativa, ma ad ogni buon conto essa è in perfetto accordo con le opinioni di alcuni importanti studiosi del cervello e della fisiologia del sistema nervoso, i quali decisamente negano la possibilità di una spiegazione puramente meccanicistica dei processi psichici e nervosi. Per quanto riguarda la seconda alternativa, si può obiettare che non necessariamente il creatore deve conoscere ogni proprietà di ciò che egli costruisce. Per esempio, noi costruiamo macchine e non possiamo prevedere il loro comportamento in tutti i suoi dettagli. Ma questa obiezione è veramente debole. Infatti noi non costruiamo le macchine dal nulla, ma le costruiamo a partire da qualche materiale. Se la situazione fosse simile in matematica, allora questo materiale o base per le nostre costruzioni sarebbe qualcosa di oggettivo, il che ci forzerebbe all'accettazione di una concezione realistica anche se certi altri ingredienti della matematica fossero una nostra creazione. Lo stesso sarebbe vero se nelle nostre creazioni noi avessimo da usare qualche organo presente in noi stessi ma diverso dal nostro ego (tale sarebbe la 'ragione' interpretata come qualcosa di simile ad una macchina pensante). Infatti i fatti matematici esprimerebbero (almeno in parte) proprietà di questo organo, che avrebbe un'esistenza oggettiva".

[22] Cfr. (CW, III, 51, p. 318): "Comunque, quello che segue con certezza pratica è questo: per dimostrare la consistenza della teoria classica dei numeri (ed *a fortiori* di tutti i sistemi più forti) devono essere usati concetti astratti (e gli assiomi evidenti che si riferiscono ad essi), ove 'astratto' è un concetto che non si riferisce ad oggetti sensibili, di cui i simboli sono un tipo speciale. (Nota 27: esempi di tali concetti astratti sono 'insieme', 'funzione di interi', 'dimostrabile, [nel senso formalistico di 'conoscibile come vero'], 'derivabile', etc., o infine 'esiste' riferito a tutte le combinazioni possibili di simboli)".

[23] Cfr. (CW, III, 51, pp. 320-321): "Infatti, è corretto che una proposizione matematica non dice niente sulla realtà fisica o psichica esistente nello spazio e nel tempo, in quanto è vera già in virtù del significato dei termini che occorrono in essa, indipendentemente dal mondo delle cose reali. Ciò che è falso, tuttavia, è asserire che il significato dei termini (cioè, i concetti che essi denotano) sia qualcosa di posto dall'uomo (man-made) e consista soltanto in convenzioni semantiche. La verità, io credo, è che questi concetti appartengono ad una realtà che è loro propria, che noi non possiamo creare o cambiare, ma solo percepire e descrivere. Perciò una proposizione matematica, per quanto essa non dica niente sulla realtà spazio-temporale, può tuttavia avere un contenuto del tutto correttamente oggettivo, nella misura in cui essa dice qualcosa sulle relazioni tra concetti. L'esistenza di relazioni non 'tautologiche' tra i concetti della matematica appare soprattutto nel fatto che a proposito dei termini matematici primitivi devono essere assunti degli assiomi, che per niente sono tautologici (nel senso d'essere riducibili all'espressione a = a), ma derivano piuttosto dal significato dei termini primitivi sotto considerazione. Per esempio, l'assioma base o

cosiddetto "ottimismo razionalistico gödeliano", concezione secondo la quale non esiste, in linea di principio, nessuna proposizione matematica vera inconoscibile (T = K)[24].

---

piuttosto lo schema d'assioma per il concetto di insieme di interi dice che, data una proprietà ben definita di interi (cioè una espressione proposizionale F(n) con una variabile per interi n), allora esiste l'insieme M degli interi provvisti di tale proprietà. Ora, considerando la circostanza che F può essa stessa contenere il termine 'insieme di interi', noi abbiamo qui una serie di assiomi piuttosto complicati sul concetto di insieme. Questi assiomi, però, (come i risultati precedenti mostrano) non possono essere ridotti a qualcosa di sostanzialmente più semplice, men che meno a tautologie. È vero che questi assiomi sono validi in forza del significato del termine 'insieme' – si può persino dire che essi esprimono il vero significato del termine 'insieme' – e perciò essi potrebbero essere chiamati analitici; in ogni caso, il termine 'tautologico', cioè vuoto di contenuto, è del tutto fuori luogo per questo tipo di proposizione, dal momento che l'asserzione dell'esistenza di un concetto d'insieme soddisfacente questi assiomi (o la consistenza degli stessi assiomi) è ben lungi dall'essere vuota, tanto che essa non può essere provata senza già far uso del concetto stesso d'insieme, o di qualche altro concetto di natura simile. Naturalmente, questo argomento particolare è indirizzato solo ai matematici che ammettono il concetto generale d'insieme nell'ambito della matematica propria. Per i finitisti, tuttavia, letteralmente lo stesso argomento potrebbe essere dichiarato per il concetto di intero e l'assioma d'induzione completa. Infatti, se nella matematica propria non è ammesso il concetto generale d'insieme, allora deve essere assunto come assioma l'induzione completa. Intendo ripetere che 'analitico' qui non significa 'vero in forza della definizione', ma piuttosto 'vero in forza della natura dei concetti che vi occorrono', il che a sua volta è distinto da 'vero in forza delle proprietà e del comportamento delle cose'. Questo concetto di analitico è così distante dal significare 'vuoto di contenuto' che è perfettamente possibile che una proposizione analitica sia indecidibile (o decidibile solo con una certa probabilità). Infatti la nostra conoscenza del mondo dei concetti può essere limitata e incompleta come quella del mondo delle cose. Non si può certamente negare che questa conoscenza, in certi casi, non solo è incompleta, ma persino indistinta. Questo accade nei paradossi della teoria degli insiemi, che sono frequentemente addotti come una prova contro il platonismo, ma, io penso, molto ingiustamente. La nostra percezione visiva talvolta contraddice la nostra percezione tattile, per esempio, nel caso del bastone immerso nell'acqua, ma nessuno si sentirebbe ragionevolmente giustificato a concludere da ciò che non esiste il mondo esterno".

Cfr. anche (CW, III, 51, p. 323): "Perciò io sostengo la concezione che la matematica descrive una realtà non-sensibile, che esiste indipendentemente sia dagli atti sia dalle disposizioni della mente umana e che è solo percepita, probabilmente in modo molto incompleto, dalla mente umana".

[24] Cfr. WANG (1974), p. 341: "Se essa [la seconda alternativa] fosse vera significherebbe che la ragione umana è assolutamente irrazionale a porre questioni cui non può rispondere, proprio mentre asserisce enfaticamente che solo la ragione può trovarne la risposta. La ragione umana sarebbe allora davvero imperfetta e in un certo senso perfino incoerente, in contraddizione stridente col fatto che quelle parti della matematica che sono state sviluppate sistematicamente e completamente… mostrano un grado stupefacente di bellezza e perfezione … Questi fatti sembrano giustificare quello che si potrebbe chiamare "ottimismo razionalistico".

La seconda linea interpretativa è più complessa ed articolata. Si tratta, poi, della modalità di spiegazione decisamente preferita dai commentatori di Gödel inclini ad una visione naturalistica della conoscenza matematica[25]. L'idea di fondo consiste nell'ipotizzare che: (i) non si danno forme di evidenza ulteriori rispetto alle evidenze finitiste, di modo che non si può dire che la mente sia dotata di risorse ulteriori rispetto alla macchina e si giunge pertanto ad affermare che (ii) mente = macchina. Non si può più, conseguentemente, affermare che G(PRA) risulti vera in base a qualche forma di *insight* non finitista, ma si può solo dichiarare di assumere la verità di G(PRA). Dunque (iii) si danno delle verità matematiche (almeno presunte tali) inconoscibili ($T \neq K$) e tra queste c'è proprio la proposizione che esprime la correttezza di PRA. Perciò (iv) la mente è una macchina incapace di intendere completamente il suo funzionamento, il che è alla radice del fenomeno stesso della incompletezza[26].

Posta nei termini in cui la presenta Gödel, questa linea interpretativa non consente di dare una risposta scontata alla domanda di cui sopra, in quanto, se vale l'identità di mente e macchina il funzionamento dovrebbe essere lo stesso. Tuttavia una risposta esiste: essa consiste nel mostrare che la diversità è semplicemente apparente. Questa è in effetti la strada seguita dai computazionalisti che accettano la seconda alternativa della disgiunzione gödeliana. Essi innanzitutto si rifiutano di ritenere che

---

[25] Si vedano, ad esempio, tra gli autori più recenti BENACERRAF (1967), DENNETT (1995), GILLIES (1998), GAIFMAN (2000), CASTI/DE PAULI (2001).

[26] Cfr. (CW, III, 51, pp. 309-310): "Comunque, per quanto riguarda la matematica soggettiva, non è precluso che possa esistere una regola finita capace di produrre tutti gli assiomi evidenti. Tuttavia, se tale regola esiste, noi non potremmo mai conoscere con certezza umana che questa è tale, cioè noi non potremmo mai conoscere con certezza matematica che tutte le proposizioni prodotte dalla regola siano corrette; o, in altri termini, noi potremmo percepire come vera solo una proposizione dopo l'altra, per un numero finito di esse. L'asserzione che esse sono tutte vere potrebbe al massimo essere conosciuta con certezza empirica, sulla base di un numero sufficiente di esempi o attraverso altre inferenze induttive. Se le cose stessero così, ciò significherebbe che la mente umana (nell'ambito della pura matematica) è equivalente a una macchina finita che, però, è incapace di intendere completamente il suo funzionamento. Questa incapacità dell'uomo di intendere se stesso gli apparirebbe erroneamente come apertura o inesauribilità della sua mente. Ma si noti, per favore, che se le cose stessero così, non si derogherebbe in nessuna maniera dall'incompletezza della matematica oggettiva. Al contrario, questa sarebbe resa particolarmente incisiva. Infatti se la mente umana fosse equivalente ad una macchina finita, allora la matematica oggettiva non solo sarebbe incompleta nel senso di non essere contenuta in nessun sistema assiomatico ben-definito, ma inoltre esisterebbero problemi diofantei del tipo sopra esposto assolutamente indecidibili, ove la locuzione 'assolutamente' significa che essi risulterebbero indecidibili, non già rispetto a qualche sistema formale, ma rispetto a qualsiasi dimostrazione matematica che la mente umana possa concepire".

la seconda alternativa porti necessariamente alla accettazione di un regno platonico di verità inconoscibili. Sulla scorta della teoria formale della verità, essi affermano che ogni verità matematica, per quanto ideale, è sempre presentabile entro gli schemi formali di qualche teoria come teorema, il che, a loro parere, non richiede alcuna forma di evidenza non finitaria. Secondo tali autori la verità matematica è in realtà qualcosa di costruito dalla mente umana, in quanto quest'ultima – pur essendo identica ad una macchina – è capace di oltrepassare sempre se stessa – senza modificare la propria struttura finitaria –, esattamente come un sistema formale può essere via via esteso attraverso l'aggiunta di nuovi assiomi. Naturalmente, i medesimi autori riconoscono che Gödel abbia ragione nel dire che, se la mente è una macchina, essa non può avere accesso a verità matematiche nella misura in cui queste si presentano dotate di un contenuto infinitario. Però, se tali verità si presentano alla mente in veste finitaria – il che accade quando diventano teoremi formali di teorie superiori – le cose cambiano. In questa prospettiva, la mente può accedere a tali verità, non in quanto sia dotata di capacità intuitive superiori, ma in quanto tali contenuti sono stati trasformati in contenuti sintattici di teorie superiori e perciò dominabili mediante procedure finitiste (Vero A $\Rightarrow$ (exT)(T- $\vdash$ A) $\Rightarrow$ PRA- $\vdash \mathrm{Pr}_{T}(\ulcorner A \urcorner)$). Una volta convertito il concetto di *verità* in quello di *derivabilità in una teoria superiore*, è plausibile ritenere che sia convertibile anche il concetto di evidenza. Come è plausibile intendere la verità quale derivabilità in una teoria superiore, così è plausibile ritenere che l'*evidenza* sia una forma di *derivabilità contratta*. In altre parole, ciò che ci appare come evidente ad un certo livello può apparire come derivabile – o meglio, viene smascherato nel suo carattere di derivabilità –, se ci si colloca al livello della teoria superiore. Dal punto di vista di questi autori, il programma di Feferman basato sulle progressioni transfinite di teorie si può esattamente leggere in questo modo.

Tuttavia, tale punto di vista è difficilmente sostenibile. La ragione fondamentale è costituita dal fatto che la tesi circa l'identificazione di verità con derivabilità in una teoria superiore è legittima solo condizionatamente, vale a dire sotto la condizione che la teoria superiore sia essa stessa corretta, ossia sia tale da condurre a conseguenze almeno aritmeticamente vere. È vero, infatti, che la proposizione esprimente la consistenza di PRA si può ottenere nella teoria superiore PA; è vero in conseguenza di ciò che la nozione (infinitaria) di verità della proposizione Cons PRA può esere riespressa nella forma di nesso (finitario) di derivabilità in PA, ma ciò non consente di dire che Cons PRA cessi d'essere vera nel senso semantico genuino, nel senso cioè che essa esprime *un reale stato di cose* relativo alla struttura della teoria PRA e che tale stato di cose consiste in una relazione

non linguistica (infinitaria perché non finitista)[27]. Tale aspetto non linguistico (infinitario) delle proposizioni matematiche vere non appartenenti a PRA (o a qualche sua estensione conservativa) è poi ineliminabile, se ci si colloca nella prospettiva della loro *giustificazione*. Infatti se ci si chiede qual è la giustificazione di Cons PRA, non basta rispondere che questa sta nel fatto che Cons PRA è derivabile in PA, a meno che non si aggiunga che gli assiomi di PA sono a loro volta veri. Si noti che giustificare Cons PRA significa giustificare la credenza che Cons PRA sia vera (nel senso genuino del termine sopra richiamato) e ciò non è possibile se non si presuppone che la teoria superiore sia a sua volta corretta. Ma che cosa significa affermare la correttezza di PA se non richiedere che gli assiomi stessi di PA siano a loro volta giustificati? Ora il fatto che gli assiomi siano una semplice estensione di PRA non è sufficiente a giustificarli, occorrono ragioni più sostanziali. Ma quale tipo di ragioni? Non possono essere *ragioni d'evidenza* perché il contesto della seconda alternativa lo vieta. Devono per forza essere *ragioni induttive* o *selezionistiche*. Nel primo caso, se ad esempio l'uso di una teoria non ha mai portato alla contraddizione, allora è plausibile che l'estensione avvenga nella direzione di aggiungere come ulteriore assioma l'espressione della sua consistenza. Nel secondo, viene selezionato l'assioma (o teoria) che assicura un adattamento migliore. Entrambe le ipotesi presentano tuttavia serie difficoltà.

La soluzione induttiva è affetta da quattro ordini di difficoltà: (i) l'induzione presuppone forme d'evidenza astratta, di tipo intensionale, simili a quelle in gioco nel sapere matematico: non è possibile, ad esempio, operare induzioni corrette se non si ammette la capacità di cogliere un insieme adeguato di predicati proiettabili. Tali forme d'evidenza sono tuttavia da escludere trattandosi di forme d'evidenza non finitista. La matematica, poi, è presupposta all'induzione[28]; (ii) l'induzione fornisce connessioni di tipo

---

[27] Cfr. (CW, III, 53, p. 346): "L'intento del programma sintattico di rimpiazzare l'intuizione matematica attraverso le regole d'uso dei simboli fallisce perché questo rimpiazzamento priva di giustificazione ogni aspettativa di consistenza, che è vitale sia per la matematica pura che per quella applicata, e questo perché la prova di consistenza o richiede una intuizione matematica di pari potenza per discernere la verità degli assiomi matematici o una conoscenza dei fatti empirici coinvolgenti un contenuto matematico equivalente".

[28] Cfr. (CW, III, 51, pp. 348-349): "… le leggi di natura senza la matematica sono esattamente "vuote di contenuto" (in questo senso) come la matematica senza le leggi di natura. Il fatto è che solo leggi di natura più matematica (o logica) hanno conseguenze controllabili empiricamente. È, pertanto, arbitrario collocare tutto il contenuto nelle leggi di natura. … Ciò che la matematica aggiunge alle leggi fisiche, non sono nuove proprietà della realtà fisica, ma piuttosto proprietà dei concetti che si riferiscono alla realtà fisica, per essere più esatti, ai concetti che si riferiscono alla combinazione delle cose. Ma tali proprietà sono qualcosa di oggettivo come le proprietà della realtà fisica…".

empirico (dunque legate alla fattualità dei mondi sperimentati e perciò contingenti) e non necessarie di tipo apriorico[29]; (iii) l'induzione ha normalmente un valore statistico, che non si addice all'ambito matematico. Supponiamo, infatti, che un numero molto alto di numeri naturali testati non contenga il codice di nessuna dimostrazione in T della contraddizione. In conseguenza di ciò si concluda induttivamente che G(T) sia vera e la si aggiunga a T. In realtà però G(T) sia falsa. È difficile ritenere che l'edificio matematico sia nato in un modo così poco affidabile; (iv) allo scopo di evitare il caso di (iii) si può intendere l'induzione come un metodo che consente il passaggio da un caso a tutti (come in fisica il rilevamento di una misura a meno di errore). Ebbene in tale interpretazione l'induzione si basa sull'intuizione dell'oggetto in questione e coincide con la generalizzazione a tutti gli oggetti di una proprietà o relazione rilevata in un oggetto singolo, ma con una procedura estendibile (in base alla conoscenza della natura dell'oggetto) a tutti gli oggetti dello stesso tipo. È questo il modo, ad esempio, attraverso il quale, si può giungere al fatto che le procedure formalizzate in PRA sono del tutto affidabili ($Pr_{PRA}(\ulcorner A \urcorner) \rightarrow A$ e, quindi, come conseguenza anche a $Cons_{PRA}$). Questo si basa sulla conoscenza della struttura di PRA e degli oggetti concreti e finiti del suo modello naturale[30].

---

[29] Cfr. (CW, III, 53, p. 354): "Io credo che, almeno per una matematica finitaria e alcune parti della matematica intuizionistica, ciascuno sia praticamente d'accordo che la dimostrazione della consistenza basata sull'intuizione matematica sia incomparabilmente più convincente [di una fondazione basata sull'induzione empirica] e certamente non per la ragione che siamo convinti che l'intuizione matematica derivi da una induzione inconscia o dall' adattamento darwiniano".

Cfr. anche (CW, III, 53, p. 351): "Si può mostrare che il ragionamento che conduce alla conclusione che non esiste nessun fatto matematico non è niente altro che una petitio principii, in quanto 'fatto' fin dall'inizio è identificato con 'fatto empirico' cioè 'fatto sintetico concernente le sensazioni'. In questo senso si può ammettere che la matematica sia priva di contenuto, ma questo cessa di essere rilevante per le questioni sollevate all'inizio, dal momento che anche i platonisti sarebbero d'accordo sul fatto che la matematica non ha contenuto di questo tipo. Infatti, per i platonisti, il suo contenuto consiste in relazioni tra concetti o altri oggetti astratti che sussistono indipendentemente dalle nostre sensazioni, per quanto queste siano percepite con un tipo speciale d'esperienza e da loro, congiuntamente a certe leggi di natura universalmente accettate, seguano conseguenze verificabili attraverso la percezione sensibile".

[30] Cfr. (CW, III, 53, pp. 346-347): "Per questi assiomi non esiste altra fondazione che non consista o nel fatto che essi (o proposizioni che li implicano) possono essere percepiti come veri (in forza del significato dei termini o per intuizione degli oggetti che cadono sotto di essi) o nel fatto che essi sono assunti (come le ipotesi fisiche) sulla base di argomenti induttivi, cioè in base al loro successo nelle applicazioni. (Nota 33: Per 'successo', nella matematica pura, di qualche assioma matematico non evidente, io intendo il fatto che molte

La soluzione selezionistica va incontro a due difficoltà particolari: (i) Innazitutto come è possibile dire che il meccanismo di selezione privilegia le verità rispetto alle falsità? Si può rispondere: perché la verità matematica è più adattiva. Tuttavia questo vale solo per certe verità matematiche molto elementari, non per altre astratte e ideali; (ii) il meccanismo di selezione assicura verità empiriche e non matematiche (vedi la difficoltà (ii) del caso precedente). Supponiamo che l'esperienza evolutiva abbia sempre riscontrato che 2 oggetti più 3 oggetti = 4 oggetti. Nella mente sarebbe allora fissata la legge 2 + 3 = 4, la quale dovrebbe apparire evidente altrettanto quanto 2 + 3 = 5. Questo è però del tutto implausibile. Le verità matematiche riguardano oggetti astratti modellati secondo considerazione di possibilità e non di mera attualità[31].

Conclusione

In conclusione, l'ipotesi consistente nel ritenere che la diversità di funzionamento della mente rispetto ad una macchina computazionale sia meramente illusoria va incontro a serie difficoltà. D'altra parte che la mente proceda secondo una logica della scoperta e una logica della giustificazione che chiami in causa modalità e procedure non soggette alle leggi della computabilità è un fatto d'esperienza matematica e di riflessione metamatematica altrettanto probante quanto i dati positivi su cui poggia l'immagine scientifica della mente e delle sue capacità cognitive. Ne segue che allo stato attuale della ricerca non è possibile pronunciarsi con un grado sufficiente di certezza sulla attendibilità di questo o quest'altro modello globa-

---

delle sue conseguenze possono essere verificate sulla base di assiomi evidenti, ove tuttavia le dimostrazioni [di tali conseguenze a partire da questi] sono maggiormente difficoltose, e il fatto ulteriore che esso risolve problemi importanti irresolubili senzo di esso. Il primo caso sembrerebbe valere almeno per qualche assioma matematico come il modus ponens e l'induzione completa. Nell'ultimo caso, a dispetto della loro fondazione induttiva, gli assiomi manifestano il loro contenuto matematico dal fatto che essi hanno conseguenze in quella parte della matematica a cui si addicono gli esempi del primo caso, cioè ove i termini hanno un chiaro significato immediatamente comprensibile (come gli assiomi dell'infinito hanno delle conseguenze sulla teoria dei numeri)".

[31] Cfr. (CW, III, 51, p. 312): "Nota 18: … 1. Le proposizioni matematiche, propriamente analizzate, non risultano asserire alcunché su ciò che è attuale nel mondo spazio-temporale. … 2. Gli oggetti matematici si conoscono con precisione e le leggi generali si possono riconoscere con certezza, cioè, attraverso inferenze deduttive e non induttive. 3. Esse si possono conoscere (in linea di principio) senza l'uso dei sensi (cioè, per mezzo della sola ragione) per questa semplice ragione, che esse non riguardano stati di cose attuali dei quali ci informano i sensi (compreso il senso interno), ma possibilità e impossibilità".

le della mente. Alla luce, poi, della riflessione metamatematica non ci sono ragioni sufficienti per ritenere giustificato, neppure a lungo termine, un modello computazionale della mente (nel suo complesso) fondato sulla analogia tra mente e programma di funzionamento (sistema formale) del cervello. Anzi le difficoltà messe in luce da una corretta considerazione dei teoremi di limitazione sembrano puntare in direzioni diverse.

Riferimenti bibliografici

Agazzi, E. (1967), *Alcune osservazioni sul problema dell'intelligenza artificiale*, «Rivista di Filosofia Neo-scolastica», 59, pp. 1-34.

Arbib, M.A. (1987), *Brains, Machines, and Mathematics*, Springer-Verlag, New York.

Arrigoni, T. (2000), *Il realismo in filosofia della matematica*, «Rivista di Filosofia Neo-scolastica», 92, pp. 627-646.

Arrigoni, T. (2002), *Il platonismo di K. Gödel alla luce della filosofia di E. Husserl. Una breve analisi*, «Epistemologia», 25, pp. 281-310.

Benacerraf, P. (1967), *God, the Devil, and Gödel*, «The Monist», 51, pp. 9-32.

Boolos, G. (1993), *The Logic of Provability*, Cambridge University Press, Cambridge.

Butti, A. (2002), *Lucas, Gödel e l'intelligenza meccanica*, «Rivista di Filosofia Neo-scolastica», 94, pp. 637-674.

Butti, A. (2001), *Gödel e l'intelligenza artificiale*, tesi di laurea in filosofia (a.a. 2000/2001), Università Cattolica, Milano.

Casti, J.L. – De Pauli, W. (2001), *Gödel. L'eccentrica vita di un genio*, Cortina, Milano.

Cellucci, C. (2002), *Filosofia e matematica*, Laterza, Bari.

Chalmers, D.J. (1995), *Minds, Machines, and Mathematics. A review of Shadows of the Mind by Roger Penrose*, «Psyche», 2(9), (http://psyche.cs.monash.edu.au/v2/psyche-2-09-chalmers.html).

Dennett, D.C. (1995), *Darwin's dangerous idea. Evolution and the meanings of life*, Simon and Shuster, New York; trad. it., *L'idea pericolosa di Darwin. L'evoluzione e i significati della vita*, Bollati Boringhieri, Torino 1997.

Feferman, S. (1962), *Transfinite Recursive Progressions of Axiomatic Theories*, «The Journal of Symbolic Logic», 27, pp. 383-390.

Feferman, S. (1995), *"Penrose's Gödelian Argument". A Review of Shadows of the Mind by Roger Penrose*, Psyche, 2(7), (http://psyche.cs.monash.edu.au/v2/psyche-2-07-feferman.html).

GAIFMAN, H. (2000), *What Gödel's Incompleteness Result does and does not show*, «The Journal of Philosophy», pp. 462-470.

GALVAN, S. (1982), *Teoria formale dei numeri naturali*, Franco Angeli, Milano.

GALVAN, S. (1991), *Logiche intensionali. Sistemi proposizionali di logica modale, deontica, epistemica*, Franco Angeli, Milano.

GALVAN, S. (1992), *Introduzione ai teoremi di incompletezza*, Franco Angeli, Milano.

GALVAN, S. (2001), *Ricerche di logica epistemica*, ISU, Milano.

GILLIES, D. (1998), *Intelligenza artificiale e metodo scientifico*, Cortina, Milano.

GIORDANI, A. (2000), *La logica dell'evidenza*, «Rivista di Filosofia Neo-scolastica», 92, pp. 582-626.

GIORDANI, A. (2002), *Teoria della fondazione epistemica*, Franco Angeli, Milano.

GÖDEL, K. (CW 1995), *Collected Works*, vol. III, Unpublished Essays and Lectures (a cura di S. FEFERMAN, J.W. DAWSON, Jr., W. GOLDFARB, C. PARSONS, R.M. SOLOVAY), Oxford University Press, New York - Oxford.

HINTIKKA, (1962), *Knowledge and Belief*, Cornell University Press, Ithaca N.J.

KUTSCHERA, (1982), *Grundfragen der Erkenntnistheorie*, de Gruyter, Berlin-New York.

LENZEN, (1980), *Glauben, Wissen und Wahrscheinlichkeit. Systeme der epistemischen Logik*, Springer Verlag, Wien-New York.

LINDSTRÖM, P. (2001), *Penrose's new Argument*, «Journal of Philosophical Logic», 3, pp. 240-250.

LUCAS, J.R. (1961), Minds, Machines and Gödel, «Philosophy», 36, pp. 112-127.

LUCAS, J.R. (1968), *Satan Stultified: a Rejoinder to Paul Benacerraf*, «The Monist», 52, pp. 145-158.

LUCAS, J.R. (1970), *The Freedom of the Will,* Clarendon Press, Oxford.

PENROSE, R. (1989), *The Emperor's New Mind*, Oxford University Press, Oxford; trad. it., *La mente nuova dell'imperatore. La mente, i computer e le leggi della fisica*, Sansoni, Milano 1998[2].

PENROSE, R. (1994), *Shadows of the Mind*, Oxford University Press, Oxford; trad. it., *Ombre della mente. Alla ricerca della coscienza*, Rizzoli, Milano 1996.

PENROSE, R. (1996), *Beyond the Doubting of a Shadow. A Reply to Commentaries on Shadows of the Mind*, «Psyche», 2(23), (http://psyche.cs.monash.edu.au/v2/psyche-2-23-penrose.html).

PUTNAM, H. (1960), *Minds and Machines*, in *Dimensions of Mind*, ed. Sidney

Hook, New York, tr. it., *Menti e Macchine*, in *Mente Linguaggio e Realtà,* Adelphi, Milano, 1993.

SHAPIRO, S. (1998), *Incompleteness, Mechanism, and Optimism*, «The Bulletin of Symbolic Logic», 4, pp. 273-302.

SMORYNSKI, C. (1977), *The Incompleteness Theorems*, in *Handbook of Mathematical Logic*, ed. J. BARWISE, North Holland, Amsterdam, pp. 821-1142.

STORRS MCCALL (2001), *On 'seeing' the Truth of the Gödel Sentence*, «Facta Philosophica», 3, pp. 25-29.

TAMBURRINI, G. (2002), *I matematici e le macchine intelligenti*, Bruno Mondadori, Milano.

TIESZEN, R. (1998), *Gödel's Path from the Incompleteness Theorems (1931) to Phenomenology (1961)*, «The Bulletin of Symbolic Logic», 4, pp. 181-203.

TURING, A.M. (1950), *Computing, Machinery and Intelligence*, «Mind», 59, pp. 433-460.

TURING, A.M. (1994), *Intelligenza Meccanica*, Boringhieri, Torino.

WANG, H. (1974), *From Mathematics to Philosophy*, Routledge & Kegan Paul, Londra, trad. it., *Dalla matematica alla filosofia*, Boringhieri, Torino 1984.

WANG, H. (1996), *A Logical Journey. From Gödel to Philosophy*, The MIT Press, Cambridge, Massachusetts, London, England.