Università degli Studi di Firenze
Tesi di Dottorato - XVI^o ciclo
anno 2004

Riflessioni sull'incompletezza
I teoremi di Gödel tra logica e filosofia

Riccardo Bruni
Dipartimento di Filosofia

Indice

Introduzione

"Mi pare che il teorema di Gödel dimostri che il Meccanicismo è falso". Questa tesi, avanzata dal filosofo inglese J. R. Lucas[1] e che potrebbe apparire bizzarra tanto al logico avvezzo alle formulazioni standard dei teoremi di incompletezza, quanto al filosofo poco abituato alla valutazione delle prese di posizione nel proprio campo alla luce di un risultato matematico, ha suscitato, nel corso della seconda metà del '900, un fervente dibattito tra il gruppo (poco numeroso, in realtà) di coloro i quali hanno creduto di riconoscere una posizione corretta dietro l'affermazione di Lucas e l'argomentazione portata in suo favore, e quanti (in numero decisamente maggiore) vi hanno visto un'inferenza indebita, tanto da un punto di vista logico che concettuale[2].

Che il tema di fondo, ossia il problema se la mente umana abbia o meno una natura puramente meccanica, appassioni, è perfettamente comprensibile. Che per dirimere una simile questione si possa chiamare in causa un risultato il quale, per quanto celebre, appartiene ad un ambito non banalmente connesso con quello del problema menzionato, appare meno ovvio.

Per spiegare le motivazioni iniziali del presente lavoro, occorre allora menzionare una circostanza ulteriore che è apparentemente indipendente da quanto detto fin qui: il fatto che, a partire dal 1995, sia iniziata un'onerosa, ma quanto mai opportuna opera di pubblicazione ragionata di parte del materiale inedito contenuto nell'ingente *Nachlass* di Kurt Gödel, prima nella forma di un volume di scritti non pubblicati[3], e, più di recente[4], di due tomi dedicati alla corrispondenza. L'occasione, ha rappresentato la spinta per una

[1] In [Luc61].

[2] Celebre ed efficace, al riguardo, è la caratterizzazione che della tesi di Lucas ha dato Douglas Hofstadter nel suo [Hof79], e secondo il quale "[a]lcuni guardano ad essa alla stregua di una dimostrazione religiosa dell'esistenza dell'anima, mentre altri, sorridendo, la rigettano come qualcosa di immeritevole di alcun commento" (*loc. cit.*, p. 472).

[3] Si tratta di [Göd95].

[4] [Göd03a], [Göd03b].

ricostruzione puntuale, che seguisse cioè il più da vicino possibile il contenuto degli scritti di Gödel, dell'analisi gödeliana del significato dei teoremi di incompletezza.

Il frutto di quest'idea, è la prima parte della tesi che porta il significativo titolo di "Gödel secondo Gödel". Lo sviluppo dell'intenzione iniziale è avvenuto secondo due diverse impostazioni, più di tipo storico l'una, di stampo maggiormente analitico la seconda, le quali sono all'origine dei due capitoli che compongono questa parte del lavoro.

La prima esigenza che si è ritenuto di dover soddisfare infatti, è stata la necessità di delineare con precisione il quadro delle prime reazioni alla scoperta di Gödel negli anni immediatamente successivi la pubblicazione della memoria del 1931, e di valutare, contemporaneamente, il peso ricoperto, per l'accettazione del risultato nella sua forma più generale, da certe, ben note implicazioni concettuali legate allo sviluppo della teoria della ricorsività. L'elemento di novità, rispetto ai numerosi lavori già presenti al proposito in letteratura, ci pare consistere proprio nella volontà di assumere, partendo da un'analisi di materiale riconducibile a Gödel, l'ottica del diretto interessato.

Con riferimento alla corrispondenza gödeliana dei primi anni '30, ad esempio, è possibile allora tanto farsi un'idea precisa del tipo di fraintendimenti ai quali la novità del risultato di Gödel ebbe modo di dare adito, con la ricostruzione delle controversie con Zermelo e Finsler, quanto comprendere, attraverso gli scambi epistolari con Church e Post, la difficoltà concettuale insita nella formulazione generale (cioè rispetto ad *ogni* sistema formale che soddisfi certi requisiti) dei teoremi di incompletezza (§1.1.1).

Il fatto che, soprattutto nell'andamento, si sia scelto per il primo capitolo l'aderenza ad un criterio di sviluppo di tipo cronologico, non significa che manchi il tentativo di fare il punto, seppur parziale, dell'analisi gödeliana: valutato l'impatto decisivo del contributo di Turing per fugare ogni dubbio sull'assoluta generalità del risultato originario (§1.2), un simile bilancio è affidato all'analisi di un testo inedito, presumibilmente appartenente alla fine degli anni '30, che contiene la lettura da parte di Gödel del fenomeno dell'indecidibilità algoritmica della matematica (§1.3), del quale i teoremi di incompletezza così come essi vengono ordinariamente formulati sono un corollario pressoché immediato.

Con il secondo capitolo ci si propone, come si diceva, uno scopo diverso: la disamina di alcune tra le implicazioni concettuali del risultato ottenuto da Gödel. Tra esse, la questione del rapporto tra l'incompletezza sintattica di certi sistemi formali ed il risultato di Tarski sull'indefinibilità di un predicato di verità per il loro linguaggio (§2.1), e quella del riflesso della scoperta gödeliana sul programma fondazionale di David Hilbert (§2.2), rappresentano due tematiche tradizionalmente legate ai risultati di Gödel. L'occasione, dunque, data la scelta delle fonti di cui si è detto, è quella rappresentata dalla possibilità di valutare, con sostegno documentale, l'incidenza che tali questioni hanno avuto sull'evoluzione del pensiero di Gödel e, con esso, del significato da lui stesso attribuito alla propria scoperta.

Ed è la possibilità di legare le riflessioni di Gödel al progressivo esplicitarsi della propria concezione filosofica della matematica, l'aspetto che viene approfondito nelle due parti successive.

Nella prima (§2.3), si tratta di misurare le suggestioni individuate in merito alle considerazioni di Gödel relative ad un altro risultato significativo della propria carriera, la dimostrazione di consistenza con gli assiomi della teoria degli insiemi dell'Ipotesi Generalizzata del Continuo di Cantor, problema che rivela legami forse inattesi con la questione dell'incompletezza, e che, soprattutto, fornisce preziose indicazioni per la precisazione della portata complessiva della riflessione gödeliana sulla matematica.

Nel caso dell'altro paragrafo (§2.4), si vuole ritornare sul problema della valutazione da parte di Gödel del significato dei teoremi di incompletezza alla luce di un'ormai celebre conferenza del 1951 (pubblicata in [Göd95]), nella quale il logico austriaco affronta, a proprio modo, il tema delle eventuali implicazioni sul confronto tra le capacità di un matematico e quelle di una macchina di Turing. Un tema, quello della natura della mente umana, che, come si cerca di argomentare, sembra integrarsi perfettamente, seppur forse rappresentando una questione di secondo piano rispetto ad altre, nel quadro della posizione di Gödel nel suo complesso.

La panoramica sulle articolate ramificazioni della riflessione gödeliana che anche la scelta apparentemente restrittiva della chiave di lettura dei teoremi di incompletezza riesce a fare emergere, vuole costituire la base per la discussione della seconda parte del lavoro la quale, come si è anticipato, intende affrontare in modo più sistematico quella lettura della scoperta di Gödel che tende ad evidenziarne l'applicabilità alla filosofia della mente.

La questione comporta interrogativi di carattere formale (Ci sono le condizioni *logiche* affinché una simile operazione sia pienamente giustificata?) e di tipo concettuale (Cosa significa *davvero* sostenere la natura computazionale/non computazionale della mente?).

Il capitolo dedicato alla figura di Alan Turing che apre questa parte del lavoro (§3), si spiega anche pensando a questo tipo di implicazioni: il matematico britannico è stato apparentemente il primo a valutare il potenziale impatto dei risultati limitativi della logica, risultati che si riferiscono essenzialmente al modello astratto di macchina da lui introdotto in un celeberrimo contributo del 1937, sul progetto di dare vita ad un dispositivo meccanico che, nella pratica, riproduca le potenzialità di quel modello.

Si è così deciso di ricostruire il percorso di Turing dall'indagine sulle Logiche Ordinali (§3.1), che è direttamente ispirata ad una certa lettura dei teoremi di Gödel, agli scritti più tardi dell'immediato secondo dopo-guerra, dove Turing si occupa in modo sistematico dei problemi di natura pratica e teorica del progetto di una macchina che esibisca un comportamento intelligente (si vedano, in particolare, §§3.2.1, 3.3), e nei quali le riflessioni sui risultati di indecidibilità ricoprono un ruolo comunque importante. Ciò fornisce peraltro lo spunto per un confronto diretto tra Gödel, anti-meccanicista dichiarato, e Turing, meccanicista presunto (§3.4).

L'esame degli scritti di Turing, costituisce così un punto di partenza ideale per un'analisi più attenta tanto della tesi di Lucas ricordata in apertura (§4.1), che, come lo stesso filosofo inglese ha riconosciuto, fu inizialmente concepita proprio come una risposta al contributo di Turing comparso su "Mind" nel 1950[5], quanto della riformulazione di essa da parte del matematico e fisico inglese Roger Penrose (§§4.3, 4.4). Delle due proposte, e di quella di Penrose in modo particolare, viene valutata la fondatezza logica alla luce di certe recenti critiche riconducibili a Per Lindström e Pavel Pudlàk (§4.4.2), ma anche la coerenza concettuale proprio in relazione alle posizioni di Turing, per Lucas (§4.2), di Gödel, per ciò che riguarda Penrose (§4.4.1).

Tra le questioni affrontate, una viene considerata di particolare rilievo: il sostenitore della tesi meccanicista contro cui argomenta chi ricorre ai teoremi di incompletezza, è un propugnatore dell'idea che, in un qualche senso, la mente umana equivalga ad una procedura effettiva di computo, ossia ad una macchina di Turing. Dai teoremi di incompletezza si desume, per contro, il fatto che, in certe circostanze da taluni considerate determinanti per discriminare la natura del suo procedere, un matematico umano acquisisca nuove verità per una via che è essenzialmente informale, e dunque non formalizzabile. Si pone allora il seguente problema: anche supposto che un argomento del genere sia sufficiente a mostrare l'inattendibilità del modello rappresentato dalle macchine di Turing, cosa ne sarebbe dell'efficacia di quello stesso argomento se riproposto in relazione ad un modello che, seppur

[5] [Tur50].

matematicamente preciso, sia ottenuto dal precedente per mezzo di una sorta di adeguamento ad un criterio di maggiore ‘affinità euristica’ con il procedere di un matematico umano idealizzato?

A quest’affascinante prospettiva è dedicato il capitolo conclusivo del lavoro (§5), sulla scorta di due contributi del matematico Roberto Magari della metà degli anni ’70 (e che fanno parte di un più ampio novero di articoli, ad opera anche di altri studiosi, dedicati alle cosiddette “Logiche Sperimentali”). La disamina della “Questione Euristica”, come al problema succitato ci riferiamo nel testo, passa attraverso la definizione dei sistemi alternativi agli usuali sistemi formali (§5.2) e delle proprietà che determinano le differenze principali tra le due nozioni, la riformulazione, per essi, dei risultati limitativi (§5.2.2) e la valutazione delle ricadute concettuali di questi ultimi (§5.3).

Parte della ricerca è stata finanziata nell’anno 2002 con i fondi dell’Università degli Studi di Firenze relativi al “Progetto Giovani Ricercatori”.

Ringraziamenti. L’autore desidera ringraziare il Prof. A. Cantini per il prezioso supporto durante le varie fasi di ‘gestazione’ del lavoro, nonché per la puntuale revisione di parti di esso (e delle rispettive versioni) che ha contribuito enormemente al loro arricchimento. A lui, unitamente al Prof. E. Casari e al Prof. P. Minari, l’autore intende esprimere la propria riconoscenza anche per una sempre stimolante ‘opera di formazione’ che dura ormai da molti anni. Un ringraziamento particolare va al Prof. W. Sieg, il quale è riuscito a trasformare una breve visita alla Carnegie Mellon University di Pittsburgh in una tappa fondamentale per gli sviluppi della ricerca.

Un grazie anche a Marco Galvagni e Giacomo Sillari: al primo, per aver contribuito a rendere piacevoli anche i periodi più faticosi del lavoro, pur attraverso mai banali pause caffé; al secondo, ospite impeccabile e punto di riferimento insostituibile durante il soggiorno americano (e oltre).

Al di fuori dell’ambito strettamente accademico, la lista dei ‘creditori’ si fa inevitabilmente più lunga e più complicata da gestire. Tuttavia, è doveroso citare almeno quanti, Sonia, mamma e babbo, sono stati costretti ad un gravoso compito di sopportazione imposto loro dalla convivenza con l’autore. Per tutti gli altri, valga il principio che anche colui che ritenga di non aver avuto alcun ruolo ai fini della presente ricerca, vi ha probabilmente contribuito, in effetti, ben oltre ciò che egli crede.

Nessuna delle persone citate, tranne l’autore, è ovviamente responsabile di qualsiasi errore o omissione che il presente lavoro dovesse contenere.

PARTE I

GÖDEL SECONDO GÖDEL

Capitolo 1

Gli anni '30: da Gödel a Turing

Il primo annuncio pubblico della scoperta dell'esistenza di proposizioni formalmente indecidibili nel sistema dei *Principia Mathematica*, risale al 7 settembre del 1930, giorno della tavola rotonda di chiusura della seconda Conferenza sull'Epistemologia e le Scienze Esatte organizzata a Königsberg, dal 5 al 7 settembre di quell'anno, dalla *Gesellschaft für empirische Philosophie* (Società per la Filosofia Empirica). L'argomento della discussione era costituito dal problema dei fondamenti della matematica, e vedeva la partecipazione di alcune indubbie personalità della materia, come John von Neumann, Arend Heyting e Rudolf Carnap.

Fu in quell'occasione che il giovane (all'epoca ventiquattrenne) e promettente Kurt Gödel, per sostenere la propria avversione a guardare hilbertianamente alla consistenza come ad un criterio per l'adeguatezza delle teorie formali, accennò al fatto di poter offrire esempi di proposizioni che, sebbene contenutisticamente vere, risultavano essere indecidibili, né dimostrabili né refutabili, nel formalismo della matematica classica.

Il carattere apparentemente casuale dell'osservazione, stride con la natura rivoluzionaria del risultato annunciato che, come tale, fu peraltro percepito dai presenti al punto da spingere gli editori della rivista *Erkenntnis*, presso la quale furono pubblicati gli atti della tavola rotonda, a chedere a Gödel di scrivere un poscritto esplicitamente dedicato ad una più diffusa spiegazione della scoperta[6].

[6] La riproduzione della parte degli atti contenenti il riferimento all'intervento di Gödel e il poscritto, si trovano ripubblicati e tradotti in [Göd86a], pp. 200-204, come [Göd31a]. Alcune notizie più dettagliate sulle circostanze storiche relative all'evento a cui si è accennato brevemente nel testo, possono essere trovate nella nota introduttiva di John Dawson.

Tuttavia, se la congettura che il materiale inedito ritrovato tra le carte del *Nachlass* gödeliano e pubblicato nel terzo volume delle opere complete di Gödel come [Göd30b] è corretta, ossia se è vero, come accredita Warren Goldfarb nella nota introduttiva, che il testo in questione contenga quanto detto da Gödel nell'intervento di venti minuti, tenuto il 6 settembre della tre giorni di Königsberg e finalizzato alla presentazione del teorema di completezza semantica per la logica dei predicati del prim'ordine, parrebbe doveroso retrodatare a questa occasione l'annuncio pubblico dell'incompletezza sintattica di certi sistemi formali.

Il testo si conclude infatti con quella che Gödel chiama "un'applicazione alla teoria generale dei sistemi d'assiomi" di quanto fino a quel punto esposto, ossia del teorema di completezza semantica. L'applicazione intende svelare il legame che sussiste tra la categoricità di un sistema formale d'assiomi, la proprietà di possedere modelli a due a due isomorfi, e la completezza sintattica. Un legame che, nota Gödel, consiste, data la completezza semantica, nel fatto che una teoria assiomatica categorica è anche sintatticamente completa, fatto facilmente verificabile mediante la seguente inferenza (che Gödel, tuttavia, non indica nel testo): se S fosse sintatticamente incompleto ed esistesse quindi un enunciato ϕ tale che $\mathsf{S} \nvdash \phi$ e $\mathsf{S} \nvdash \neg\phi$, seguirebbe che $\mathsf{S}' := \mathsf{S} \cup \{\phi\}$ e $\mathsf{S}'' := \mathsf{S} \cup \{\neg\phi\}$ sono due teorie consistenti. Quindi, per una delle conseguenze del teorema di completezza semantica, esisterebbero due modelli $\mathcal{M}'$ e $\mathcal{M}''$ di S' e S'' rispettivamente, che sarebbero, per la definizione dei due sistemi, anche modelli di S; tuttavia, contrariamente all'ipotesi di partenza, essi risulterebbero non isomorfi verificando l'uno un enunciato del quale il secondo verifica la negazione.

Gödel prende spunto da quest'osservazione per sottolineare come, quindi, se il teorema di completezza semantico potesse venir esteso così com'è, alle logiche di ordine superiore, si potrebbe anche concludere che una teoria categorica per l'aritmetica, come risultano essere gli assiomi di Peano, originariamente formulati al secondo ordine, è anche sintatticamente completa. Una tale teoria garantirebbe quindi la risolubilità di ogni problema formulabile nel proprio linguaggio. Tuttavia, riferendosi ad una sua scoperta recente, Gödel annuncia l'impossibilità di una simile estensione: "[i]l sistema degli assiomi di Peano con la logica dei *Principia mathematica* aggiunta come sovrastruttura, non è sintatticamente completo"[7].

[7][Göd30b], p. 29. *NdA*: il riferimento per le citazioni tratte dai lavori di Gödel è la ripubblicazione dei medesimi nei volumi dei *Collected Works*. Laddove il testo vi compaia

Al di là del fatto di costituire, appunto, il primo riferimento negli scritti di Gödel ai teoremi di incompletezza, l'osservazione merita alcune considerazioni ulteriori.

Per quanto non si ritrovi qui quell'enfasi che mancava già all'annuncio dato durante la tavola rotonda di Königsberg e che sarebbe stato lecito aspettarsi, il passo in questione infatti lega tra loro le due scoperte, la completezza semantica e l'incompletezza sintattica, che garantirono a Gödel la notorietà in campo logico tra la fine degli anni '20 e i primi anni '30 del '900.

Nel merito, l'osservazione di Gödel, per quanto appaia oggi datata, acquisisce rilevanza proprio da un punto di vista storico. Da un lato, perché fornisce un esempio di come l'uso di trattare informalmente certe nozioni (quelle semantiche in primo luogo) aveva alimentato convinzioni rispetto alle quali i risultati ottenuti da Gödel si trovarono a svolgere un ruolo essenziale, confermandole in certi casi (come nel caso del teorema di adeguatezza semantica) ma, soprattutto, rivelandone l'assoluta infondatezza in altri (incompletezza sintattica e, in misura minore, categoricità). Dall'altro, in quanto vi si può ravvisare il preludio degli approfondimenti della ricerca successivi, in gran parte derivanti dall'aprirsi di nuovi campi dell'indagine metamatematica come nel caso dello sviluppo sistematico della teoria dei modelli.

L'interesse per l'osservazione di Gödel in relazione alle finalità del presente lavoro, deriva però anche da quel riferimento alla possibilità di utilizzare certi risultati dell'indagine metamatematica per una 'teoria generale dei sistemi d'assiomi'.

Con il presente capitolo, adottando un taglio più storico di quanto non si farà successivamente nell'analizzare le considerazioni di Gödel relative alle implicazioni (soprattutto di natura concettuale) dei propri risultati, si vuole ripercorrere le tappe che hanno condotto alla formulazione dei teoremi di incompletezza nella loro generalità, così come queste tappe si manifestano attraverso gli stessi scritti gödeliani.

Un percorso che sembra particolarmente rilevante proprio al fine di poterne valutare l'impatto sulla 'teoria generale' dei sistemi formali d'assiomi.

---

in traduzione inglese con il testo tedesco a fronte, la citazione viene riferita al passo originale. La traduzione dei passi in lingua è opera dell'autore del presente lavoro, che se ne assume ovviamente l'intera responsabilità.

1.1 Una versione generale dei teoremi di incompletezza (1): 1931-1934

Al termine del celeberrimo lavoro che contiene la prima esposizione dei teoremi di incompletezza, si legge:

> Nel presente articolo si è ristretto del tutto la nostra attenzione al sistema P, e si sono solamente indicate le applicazioni agli altri sistemi. I risultati verranno enunciati e dimostrati in tutta la loro generalità in un seguito da pubblicarsi presto. In quell'occasione verranno dati anche i dettagli della dimostrazione del Teorema XI [relativo all'indecidibilità della proposizione che esprime la consistenza di un sistema d'assiomi], che qui è stata soltanto sommariamente indicata.[8]

L'osservazione di Gödel merita attenzione per due motivi: da un lato in quanto segnala una differenza importante tra come il risultato fu comunicato originariamente e assimilato (in relazione cioè ad un certo sistema formale P), e la forma assolutamente generale che esso ha acquisito col tempo e con la quale viene oggi trasmesso; dall'altro, perché contiene il riferimento al risultato relativo all'indecidibilità della proposizione che esprime la consistenza di un sistema formale e che rappresenta, della scoperta gödeliana nel suo complesso, l'aspetto forse più tecnicamente delicato, come ne conferma la sorte 'travagliata'[9].

Sotto questo riguardo, si comprende allora perché possano suscitare curiosità tanto il passo succitato, quanto la nota editoriale ad esso secondo la

[8][Göd31b], p. 195.

[9]Forse anche a causa della rinuncia successiva da parte di Gödel a pubblicare la seconda parte della memoria del 1931, la prima dimostrazione dettagliata del secondo teorema di incompletezza risale al 1939 ([HB39], pp. 285-328). È in quest'opera che si trova la prima formulazione delle cosiddette "condizioni di derivabilità", le proprietà relative al predicato $Teor$ che un sistema formale deve soddisfare affinché ad esso si applichino i teoremi di incompletezza. A partire dagli anni '50 il tema del secondo teorema di Gödel fu riaperto, da un lato, grazie all'opera di di M. H. Löb ([Löb50]) il quale propose una semplificazione delle condizioni di derivabilità di Hilbert e Bernays, che diverranno in tale versione un modo standard ed estremamente elegante per esporre il risultato di Gödel in tutta generalità; dall'altro a seguito di una serie di lavori (in particolare, [Tak55], [Fef60], [Kre65a]) dedicati a sistemi formali che, a vario titolo, forniscono esempi di teorie che sono in grado di dimostrare la propria consistenza. La problematica ha ricevuto così definitiva chiarificazione dal momento che Solomon Feferman ha mostrato come l'indecidibilità della consistenza dipenda in modo particolare dai dettagli della 'presentazione' sintattica di un sistema formale, nel senso che, a certe condizioni, è possibile formulare un controesempio al risultato di Gödel (Cfr. [Fef60], pp. 68-69).

quale fu la "pronta accettazione dei suoi risultati" a far rinunciare a Gödel al proposito di pubblicare il seguito nel volume successivo dei *Monatschefte für Mathematik und Physik* (la stessa rivista che aveva ospitato la memoria del 1931)[10].

Da un punto di vista cronologico, si può persino essere più precisi rispetto a questo mutamento di intenti, grazie alle indicazioni provenienti della corrispondenza gödeliana: rispondendo ad una sollecitazione di Carnap al termine di una lettera datata 23 febbraio 1932 che lo invitava ad optare, nel progettato seguito di [Göd31b], per "una presentazione più dettagliata e comprensibile"[11], Gödel sottolinea, l'11 settembre di quell'anno, come nella seconda parte del lavoro egli intenda dare "una definizione di 'verità'" per un linguaggio formale[12], segno evidente del fatto che che egli ne stava considerando ancora la pubblicazione. Tuttavia, già dalla lettera successiva del 28 novembre 1932, in risposta ad un suggerimento di Carnap del tutto analogo al precedente, si apprende come Gödel considerasse ormai tramontata l'ipotesi (egli dice infatti: "la seconda parte del mio lavoro esiste solo nel regno delle idee"[13]).

Nel suo [Daw85], John Dawson, che può essere certamente annoverato tra gli esegeti più informati sull'opera gödeliana essendo egli stato tra i primi ad analizzare in modo dettagliato l'ingente quantità di materiale proveniente dal *Nachlass*, si propone proprio di verificare quanto l' "impressione soggettiva di Gödel" relativamente all'accettazione del risultato, alla quale allude la nota 68a di [Göd31b] citata in precedenza, "riflettesse circostanze oggettive"[14]. La conclusione dello studioso americano è ben più cauta di quanto la motivazione per la decisione relativa alla mancata pubblicazione della seconda parte di [Göd31b] lasci pensare: se infatti l'impressione gödeliana viene giudicata corretta alla luce delle reazioni manifestate dai "formalisti", ossia da coloro ai quali un interesse verso la teoria della dimostrazione di stampo hilbertiano

[10]La nota in questione, la 68a di [Göd31b], era stata pensata per la traduzione del lavoro apparsa in [vH67]. Si tratta di una delle numerose chiose al testo originale aggiunte su suggerimento dello stesso Gödel (parte della corrispondenza tra quest'ultimo e van Heijenoort relativa alla ripubblicazione in questione, si trova riprodotta in [Göd03b], pp. 307-325).

[11][Göd03a], p. 345.

[12]Cfr. *Ibidem*, p. 347. Per una tematizzazione di questo, e di altri accenni dello stesso tenore che si possono trovare in alcune lettere di Gödel, si veda §2.1.

[13]*Ibid.*, p. 355.

[14][Daw85], p 75 (*NdA*: il riferimento per questa ed altre citazioni tratte da [Daw85], è la riedizione del medesimo in [Sha88] come risulta dalla voce bibliografica).

rendeva più facilmente accessibile il taglio logico del lavoro di Gödel, essa appare invece meno giustificata se misurata rispetto ad una *audience* più vasta, considerato soprattutto che "furono sollevate obiezioni sia tecniche che filosofiche" e che "prevaleva l'idea che i risulati di Gödel fossero di natura antinomica o di generalità limitata"[15]. La spiegazione di Dawson consiste allora nell'ipotizzare come l'affermazione relativa alla "pronta accettazione" dei propri risultati, nasconda l'adozione di un criterio selettivo da parte di Gödel e che in realtà, quindi, ciò sia vero rispetto a "*coloro la cui opinione fosse per lui di reale interesse*"[16].

Trattandosi di questioni non irrilevanti per il tema che si intende affrontare, la recezione della memoria di Gödel rimarrà al centro anche dei prossimi paragrafi secondo il seguente andamento: in primo luogo ci si richiamerà ai personaggi la cui reazione alla scoperta gödeliana permette di far emergere quei passaggi fondamentali attraverso i quali si è giunti alla formulazione ed all'analisi dei teoremi di incompletezza nella loro generalità (che rappresenta lo scopo specifico di questo primo capitolo).

Per ciò che riguarda invece la lettura da parte di coloro che erano più vicini alla cerchia di Hilbert o che comunque avevano sviluppato interesse per il programma fondazionale del matematico tedesco, la disamina del materiale relativo all'accoglienza riservata da questi al risultato di Gödel avverrà nel secondo capitolo, all'interno di un paragrafo (§2.2) esplicitamente dedicato al finitsmo.

1.1.1 Le prime reazioni al risultato nella corrispondenza gödeliana

La memoria del 1931 era stata concepita da Gödel con un paragrafo iniziale che, in termini intuitivi e con pochi dettagli tecnici, doveva costituire una introduzione a certi aspetti rilevanti del teorema.

Le pagine in questione contengono in effetti un'indicazione dei pilastri su cui poggia la dimostrazione vera e propria del risultato ed è, come si vedrà poi in modo più approfondito[17], una fonte preziosa di informazioni relative alle genesi del medesimo. Data l'attinenza almeno con parte del materiale relativo alla recezione del lavoro di Gödel che si intende esaminare nel seguito,

[15] *Ibid.*, p. 91.
[16] *Ibid.*, p. 91.
[17] Cfr., in particolare, §2.1.

ci pare opportuno indicare a fini meramente espositivi i passaggi salienti della presentazione gödeliana.

Tra i momenti cruciali della dimostrazione dei teoremi di incompletezza, ricopre un ruolo fondamentale la distinzione tra linguaggio oggetto e metalinguaggio. In questo quadro si inserisce, in primo luogo, il riferimento alla necessità di precisare la *sintassi* di un sistema d'assiomi:

> Le formule di un sistema formale (restringiamo la nostra attenzione in questo caso al sistema PM [Principia Mathematica]) appaiono come successioni finite di simboli primitivi (variabili, costanti logiche e parentesi o segni di interpunzione), ed è semplice stabilire con precisione assoluta *quali* successioni di simboli primitivi siano formule ben formate e quali no. In modo del tutto simile, le dimostrazioni, da un punto di vista formale, non sono altro che successioni finite di formule (con certe proprietà specifiche).[18]

Gödel prosegue poi sottolineando la necessità di fissare l'ambito entro il quale svolgere la *metateoria* del sistema in questione:

> Naturalmente, ai fini delle considerazioni metamatematiche non ha alcuna importanza quali oggetti vengono scelti come simboli primitivi e si assegneranno i numeri naturali per un tale scopo.[19]

Si diceva in precedenza della volontà dichiarata da parte di Gödel di evitare i dettagli tecnici. Tra i passaggi nei quali egli ricorre a delle perifrasi per esprimere il contenuto di risultati che vengono invece illustrati con dovizia nel corpo dell'articolo, il primo è quello dedicato alla questione della corrispondenza tra l'ambito meta-teorico e quello intra-teorico (ossia tra le relazioni sussistenti tra i codici numerici delle configurazioni dei simboli del linguaggio, e le formule di quest'ultimo):

> Le nozioni (proposizioni) metamatematiche diventano quindi nozioni (proposizioni) su numeri naturali e loro successioni; quindi esse possono (almeno in parte) essere espresse mediante i simboli del sistema PM stesso. In particolare, si può dimostrare che le nozioni di "formula", "dimostrazione" e "teorema" possono essere definite nel sistema PM; ciò significa che è possibile, ad esempio, trovare una formula $F(v)$ di PM contenente una variabile libera v [...] tale che $F(v)$, interpretata in accordo col significato dei termini di PM dice: v è un teorema.[20]

[18][Göd31b], p. 147.
[19]*Ibid.*, p. 147.
[20]*Ibid.*, p. 147.

Il fondamento implicito delle osservazioni 'semantiche' di Gödel, relative cioè all''interpretazione' di certe formule del linguaggio del sistema PM, altro non è che il teorema di rappresentabilità delle funzioni e relazioni ricorsive primitive, il quale dice che per ogni relazione primitiva ricorsiva P, esiste una formula ϕ_P tale che, per ogni $n_1, \ldots, n_k \in \mathbf{N}$:

$$\begin{array}{lcl} P(n_1, \ldots, n_k) & \Rightarrow & \vdash \phi_P(\overline{n}_1, \ldots, \overline{n}_k) \\ \overline{P}(n_1, \ldots, n_k) & \Rightarrow & \vdash \neg\phi_P(\overline{n}_1, \ldots, \overline{n}_k) \end{array} \tag{1.1}$$

(dove $\vdash \phi$ sta per $PM \vdash \phi$, e $\overline{n}$ per il numerale corrispondente a $n \in \mathbf{N}$). Tuttavia, è proprio la resa di questo fatto da parte di Gödel a costituire una delle fonti di interesse di questa introduzione al testo[21].

Passando poi alla descrizione del metodo di prova, Gödel sottolinea come, indicando a un tempo una delle ipotesi fondamentali su cui il teorema si basa e la possibilità di una generalizzazione sganciata dal formalismo dei *Principia mathematica*, esso si applichi ad ogni sistema formale che "in primo luogo, una volta che venga interpretato come un sistema di nozioni e proposizioni, abbia mezzi espressivi sufficienti per definire le nozioni occorrenti [nell'argomento] (in particolare quella di 'teorema')", e che sia tale che "ogni formula dimostrabile sia vera sotto l'interpretazione considerata"[22].

Nel merito, l'argomento si fonda su due pilastri ulteriori: (i) l'esistenza di una funzione primitiva ricorsiva $R(x)$ che enumera tutti i "segni di classe" (cioè le formule del linguaggio di PM contenenti una sola variabile libera); (ii) una funzione (anch'essa primitiva ricorsiva) di sostituzione, indicata con $[x; y]$, del termine y al posto della variabile libera della formula x[23].

La prova procede quindi lungo i binari ben noti. Sia dunque:

$$\phi_0(x), \phi_1(x), \phi_2(x), \ldots, \phi_n(x), \ldots$$

[21] Come è noto, l'osservazione di Gödel relativa all' 'interpretazione' di certe formule può essere anche resa vera in senso letterale: è possibile cioè definire un'interpretazione delle formule di PM in $\mathbf{N}$, data la quale il teorema (1.1) citato nel testo garantisce l'adeguatezza semantica della teoria rispetto alle funzioni e relazioni primitive ricorsive.

[22] *Ibid.*, p. 151.

[23] Si tratta di funzioni metamatematiche. Dunque, considerato che i numeri naturali sono "i simboli primitivi", per riprendere l'espressione di Gödel, per considerazioni di tipo metateorico, per ogni $n, m \in \mathbf{N}$, $R(n)$ e $[R(n); m]$ indicano rispettivamente, il codice numerico dell'ennesimo 'segno di classe', ed il codice della formula ottenuta sostituendo il termine con codice m alla variabile di $R(n)$.

l'enumerazione generata dalla funzione R. Si definisce allora, nella notazione scelta da Gödel, il seguente insieme:

$$K := \left\{ n \in \mathbf{N} \mid \overline{Bew}([R(n); n]) \right\} \tag{1.2}$$

dove Bew (da *Beweisbar*) indica la proprietà metateorica di 'essere dimostrabile in PM'.

Come applicazione di (1.1), ossia del fatto che certe proprietà metateoriche sono "esprimibili" in PM mediante formule che, quindi, "interpretate in accordo al significato dei termini", hanno quelle stesse proprietà come propria 'realizzazione', Gödel sottolinea che esiste un $q \in \mathbf{N}$ tale che la formula $[R(q); n]$ "dice" che n appartiene all'insieme K.

È chiaro come a partire da questi elementi si possa costruire l'argomento che mostra l'indecidibilità della formula $[R(q); q]$.

Tuttavia, è ugualmente meritevole di citazione il passo relativo di [Göd31b], che permette di comprendere appieno la scelta stilistica di Gödel nell'esporre introduttivamente il risultato:

> Dimostriamo ora che la proposizione $[R(q); q]$ è indecidibile in PM. Infatti supponiamo che la proposizione $[R(q); q]$ sia dimostrabile; allora [poiché per ipotesi ogni teorema è vero "nell'intepretazione considerata"] sarebbe anche vera. Ma allora, secondo la definizione data prima, q apparterrebbe a K, quindi, per [(1.2)], si avrebbe $\overline{Bew}([R(q); q])$, che contraddice l'assunzione. Se, d'altro canto, la negazione di $[R(q); q]$ fosse dimostrabile, allora $\overline{q \in K}$, cioè varrebbe $Bew[R(q); q]$. Ma allora $[R(q); q]$, così come la sua negazione, sarebbero dimostrabili, che è ancora una volta impossibile.[24]

È chiaro che per quanto una simile presentazione eviti i dettagli dell'aritmetizzazione della sintassi e della rappresentabilità formale delle relazioni primitive ricorsive, anch'essa nasconde le sue insidie: presuppone infatti, in particolare, una comprensione chiara della distinzione tra linguaggio oggetto e metalinguaggio e del problema della corrispondenza tra i due livelli (che sta dietro il riferimento, intuitivamente chiaro ma notoriamente rischioso, alla 'verità' delle formule "sotto l'interpretazione considerata").

Senza una simile consapevolezza, poteva essere facile fraintendere i commenti conclusivi di Gödel:

> L'analogia tra questo argomento e l'antinomia di Richard salta agli occhi. Esso risulta legato anche al "mentitore", dal momento che la proposizione

[24] *Ibid.*, p. 148.

indecidibile $[R(q); q]$ asserisce che q appartiene a K, cioè, per [(1.2)], che $[R(q); q]$ non è dimostrabile. Si ha dunque davanti a noi una proposizione che dice di se stessa di non essere dimostrabile [...]. Da [questa] osservazione, segue immediatamente che $[R(q); q]$ è vera, dal momento che $[R(q); q]$ *è* in effetti indimostrabile (in quanto indecidibile). Dunque, la proposizione che è indecidibile *nel sistema PM* è stata decisa già per mezzo di considerazioni di tipo metamatematico. L'analisi precisa di questa situazione conduce a risultati sorprendenti relativi alle prove di consistenza per i sistemi formali, risultati che verranno discusssi più in dettaglio nella Sezione 4.[25]

Le due lettere di Zermelo

Gödel poteva forse non immaginare che i termini della questione così come egli aveva deciso di esporli nel paragrafo introduttivo di [Göd31b], avrebbero creato difficoltà ad Ernst Zermelo, uno dei padri della teoria assiomatica degli insiemi, con il quale intrattenne, tra il settembre e l'ottobre del 1931, un breve e, a quanto pare, infruttuoso scambio epistolare.

Nella prima lettera, datata 21 settembre 1931, Zermelo intendeva informare Gödel come, a suo dire, la dimostrazione da lui fornita dell'esistenza di una proposizione indecidibile nel sistema PM contenesse "una *lacuna* sostanziale" nella definizione dell'insieme K riportata poc'anzi. La spiegazione che Zermelo dà per la propria caustica analisi, denota una non chiara comprensione di alcuni fatti basilari insiti nel discorso di Gödel:

> Per costruire una proposizione "indecidibile", Lei definisce a pagina 175[26] un "simbolo di classe" (una funzione proposizionale di *una* variabile) $S = R(q)$, e mostra che né $[R(q); q] = A$ né la sua negazione $\overline{A}$ sarebbero "dimostrabili". Ma è effettivamente vero che
>
> $$S = \overline{\text{Bew}}[R(n); n]$$
>
> appartiene al Suo "sistema", ed è Lei giustificato ad identificare questa funzione con $R(q)$, solo perché si tratta di un "simbolo di classe"? So che più avanti segue una dettagliata teoria dei "simboli di classe", ma per una critica è sufficiente la seguente osservazione: nella Sua formula [(1.2)], si ometta il simbolo "Bew" e si scriva invece
>
> $$n \in K^* \equiv \overline{[R(n); n]} = S^*$$
>
> Ponendo nuovamente $S^* = R(q^*)$, segue che la proposizione
>
> $$A^* = R(q^*; q^*)$$

[25] *Ibid.*, pp. 149-150.

[26] *NdT*: Zermelo si riferisce ovviamente alla numerazione delle pagine nell'articolo originario di Gödel.

non può essere *né* vera *né* falsa; ossia, la Sua assunzione conduce ad una *contraddizione* analoga all'antinomia di Russell. Come nel caso dei paradossi di Richard e di Skolem, l'errore sta nella (fallace) assunzione che ogni nozione matematica definibile sia esprimibile mediante una "combinazione finita di simboli" (secondo un sistema *fissato*)– cosa che chiamo il "pregiudizio finitista". In realtà la situazione è diversa, e solo una volta che questo pregiudizio venga superato (un compito che è divenuto un mio specifico dovere) sarà possibile dare vita ad una "metamatematica" ragionevole. Correttamente intesa, proprio il Suo tipo di dimostrazione potrebbe contribuire in gran misura a questo e potrebbe quindi rendere un servigio sostanziale alla causa della verità. Ma la Sua "dimostrazione" così come è concepita al momento, non mi pare possa essere considerata stringente.[27]

Appare chiaro, da questo stralcio, come il fraintendimento di Zermelo sia pressoché totale: egli non solo sembra ignorare che l'affermazione di Gödel circa l'esistenza di una formula del linguaggio di PM che "rappresenta" l'insieme K nel sistema (la S del passo succitato) si basi su di un *teorema*[28], ma, con il rilievo mosso a Gödel, Zermelo denota anche il travisamento della distizione cruciale tra linguaggio oggetto e metalinguaggio (come altrimenti spiegare l'identificazione della formula indecidibile A con il suo codice, $[R(q);q]$, o, nella definizione di K^*, l'apposizione del connettivo di negazione di fronte al *numero* $[R(n);n]$?). La supposta contraddizione "alla Russell" che Zermelo segnala è quindi gravemente inficiata dagli effetti di una simile interpretazione erronea.

La risposta di Gödel del 12 ottobre 1931, infatti, punta l'indice proprio su quest'aspetto, che mina alla base l'attendibilità dell'osservazione di Zermelo.

Tuttavia, fatto rilevante, essa non ne destituisce completamente l'interesse come poteva essere lecito attendersi, dal momento che, *correttamente intesa*, essa fornisce a Gödel lo spunto per chiarire ulteriormente un fatto rimasto implicito nell'esposizione informale del paragrafo introduttivo di [Göd31b]:

> Lei definisce la classe K^* mediante la specificazione: "n appartiene a K^* se $[R(n);n]$ non è corretta", laddove io definisco una classe K mediante: "n appartiene a K se $[R(n);n]$ non è *dimostrabile*". L'assunzione che K^* sia esprimibile mediante un simbolo di classe del sistema dato, porta quindi a contraddizione (tuttavia si tratta non della mia assunzione, ma della Sua). [...]
>
> La definizione di K^* tuttavia, deve essere propriamente scritta come
>
> $$n \in K^* = \overline{W}[R(n);n]$$

[27] [Göd03b], pp. 420-422.

[28] Come si ricordava in una nota precedente, si tratta del Teorema V, p. 171 di [Göd31b].

> dove $W(x)$ è supposto significare: "x è una formula corretta", o, più precisamente, "x è una formula che esprime una asserzione vera". Risulta chiaro adesso che nella definizione per K^* compare un nuovo concetto, quello di "formula corretta", o, rispettivamente, di classe delle formule corrette. Questo concetto però non dovrebbe essere in alcun modo ricondotto ad una proprietà combinatoria delle formule (ma si fonda piuttosto sul significato dei simboli), e dunque non dovrebbe essere ricondotto nella metamatematica aritmetizzata a concetti aritmetici semplici; oppure, in altre parole: La classe delle formule corrette *non* è esprimibile mediante un simbolo di classe del sistema dato (e dunque non lo è la classe K^* definita a partire da essa).[29]

Il rilievo di Gödel è ineccepibile: affinché l'ipotesi dell'esistenza di una formula che rappresenti la collezione K^* sia ammissibile, occorre assumere che, in primo luogo, la proprietà metateorica sulla base della quale essa è definita sia rappresentabile, nel senso in cui lo sono altre proprietà del genere, nel sistema PM.

Ma questo, lo si è visto, significa assumere l'esistenza di un $p \in \mathbf{N}$ tale che:

$$\begin{array}{rcl} W(n) & \Rightarrow & \vdash \phi_p(\overline{n}) \\ \overline{W}(n) & \Rightarrow & \vdash \neg\phi_p(\overline{n}) \end{array}$$

Ed è questa assunzione che, come è noto, si può mostrare essere contraddittoria (assumendo la consistenza del formalismo scelto) considerando, per parafrasare la costruzione dell'enunciato indecidibile, l'enunciato che 'dice di se stesso di essere falso' (cioè formalizzando letteralmente la cosiddetta 'antinomia del mentitore').

Se agli occhi del logico di oggi le osservazioni di Gödel possono apparire scontate e banali, lo sono meno in una prospettiva storica, proprio perché, alludendo ad un'anticipazione del fondamento del teorema attribuito comunemente ad Alfred Tarski, le parole di Gödel svelano, tra l'altro, un senso possibile (quello letterale) delle riflessioni relative alla verità della proposizione indecidibile contenute nell'introduzione della memoria del 1931.

Queste stesse considerazioni portano Gödel a segnalare a Zermelo la possibilità di vedere alternativamente il risultato relativo all'esistenza di proposizioni formalmente indecidibili, come la contrapposizione tra il concetto di 'verità' di una formula di un certo linguaggio e quello di 'dimostrabilità entro un sistema formale':

[29] [Göd03b], pp. 424-425.

> In relazione a ciò che si è detto, si potrebbe effettuare la mia dimostrazione come segue: La classe W delle formule vere *non è in nessun caso* coestesa con un simbolo di classe del sistema stesso (perché l'assunzione che lo sia porta a contraddizione). La classe B delle formule dimostrabili *è* coestesa con un simbolo di classe del sistema in questione (così come può essere dimostrato in dettaglio); di conseguenza B e W non possono essere coestesi l'uno rispetto all'altro. Siccome $B \subseteq W$, vale $B \subset W$, ossia, c'è una formula A vera che non è dimostrabile. Essendo A vera, non-A [è falsa e quindi] non è dimostrabile anch'essa, quindi, A è indecidibile. Questa prova tuttavia ha lo svantaggio che non fornisce alcuna costruzione dell'enunciato indecidibile e non risulta quindi essere intuizionisticamente ineccepibile.[30]

Si vedrà successivamente (§2.1) in che modo queste considerazioni di Gödel si inseriscano in un quadro più generale di un richiamo a tematiche di tipo tarskiano nell'opera gödeliana.

Oltre ai rilievi sulla supposta contraddittorietà della prova di Gödel, c'erano poi le osservazioni di Zermelo relative all'opposizione del risultato al "pregiudizio finitista". A tale proposito Gödel sottolinea inoltre[31] come il teorema di incompletezza si riveli importante non tanto nel mostrare che "non si può catturare tutta la matematica in un sistema formale" (un fatto questo che, nota Gödel, "segue già in accordo alla procedura diagonale di Cantor"), bensì che neanche "certi sottosistemi della matematica" possano essere formalizzati in modo (sintatticamente) completo.

Nonostante le nuove indicazioni, il risultato di Gödel era destinato ad essere non del tutto compreso da Zermelo, come la lettera di quest'ultimo del 29 ottobre 1931 dimostra.

A proposito delle spiegazioni gödeliane, il matematico tedesco scrive infatti:

> Dunque la Sua "restrizione finitista" (come la chiamo io) si applica solo rispetto alle formule *dimostrabili* del Suo "PM-sistema", e non per i suoi *enunciati in generale*, laddove per gli *enunciati* del sistema Lei permette in effetti nuove formazioni *libere* alla maniera della procedura diagonale di Cantor. Così lei ottiene naturalmente un insieme *più che numerabile* di enunciati possibili, rispetto ai quali solo un sottoinsieme *numerabile* risulta "dimostrabile" e quindi devono certamente esisterne di indecidibili"[32]

Con ancora più veemenza che nella prima lettera inviata a Gödel, Zermelo persisteva a leggere nel risultato una conferma delle proprie posizioni 'antifinitiste':

[30] *Ibid.*, pp. 426-428.
[31] *Ibid.*, p. 428.
[32] *Ibid.*, p. 430.

> Lei ammette che questi enunciati "indecidibili" divengono nonostante tutto "decidibili" in "sistemi superiori". Ma questi sistemi "superiori" si distinguono dall'originario non per l'ammissione di *nuovi enunciati*, come si potrebbe pensare dalla Sua formulazione, ma semplicemente per nuovi *strumenti di dimostrazione*, e tutto ciò che Lei dimostra nell'articolo si riconduce a ciò che anche *io* ho sempre sostenuto, che uno schema di prova "finitisticamente ristretto" *non sia sufficiente* per "decidere" gli enunciati di un sistema matematico più che numerabile.[33]

Non è dato sapere cosa Gödel pensasse di questi commenti ulteriori dal momento che non fu inviata alcuna risposta a quest'ultima missiva. Certo, immaginiamo che possano averlo indotto a ritenere quantomeno che lui e Zermelo dovevano avere idee irrimediabilmente diverse su ciò che potesse giovare alla "causa della verità", alla quale il matematico tedesco aveva alluso nella prima lettera.

La controversia con Finsler

Zermelo non fu certo l'unica personalità che, intrigata dai teoremi di incompletezza pur non avendone colto tutti gli aspetti essenziali, fu spinta a cercare un contatto diretto con Gödel. Poco più di un anno dopo rispetto alla corrispondenza con il matematico tedesco infatti, nella primavera del 1933 Gödel si trovò ad intrattenere un carteggio di tenore analogo con il matematico svizzero Paul Finsler.

Rispetto alla corrispondenza con Zermelo, i documenti che ne sono testimonianza permettono di individuarne una certa specificità, dovuta al fatto che Finsler credeva di avere fondati motivi (più di quanto Zermelo fosse giustificato a fare vedendo nei teoremi di incompletezza una conferma della propria critica al "finitismo"), per considerare la scoperta di Gödel come il raffinamento di un risultato da lui presentato in un articolo pubblicato nel 1926 sulla rivista *Mathematische Zeitschrift*, e dal titolo "Dimostrazioni formali e indecidibilità".

Con il lavoro in questione, Finsler intendeva mostrare l'esistenza di una proposizione falsa e "formalmente" indecidibile, sfruttando l'idea alla base del paradosso di Richard sulla definibilità finita.

Finsler assume come dato, infatti, un sistema di simboli S che contenga "tutti i simboli utilizzati a fini matematici", e, più in generale, "tutti i simboli usati fino ad oggi (o persino in futuro) per la scrittura o la stampa", e al quale

[33] *Ibid.*, p. 430.

sia associato un "dizionario" B "contenente una 'grammatica' che specifichi in modo inequivocabile il significato associato a certe combinazioni finite di questi simboli, le 'parole' "[34].

Con l'accezione "dimostrazione formale", Finsler intende indicare "una combinazione finita di simboli del sistema S tale che questa combinazione abbia un significato, da appurare mediante l'uso di B, che costituisca una dimostrazione logicamente ineccepibile"[35].

Considerato quanti lontani siano i termini di una simile definizione dalla specificazione della sintassi di un sistema logico, il virgolettato appare un obbligo nell'affermare che le clausole di Finsler isolino un 'formalismo'.

Per ciò che riguarda la prova dell'esistenza di una proposizione falsa e 'formalmente' indecidibile, nel senso appena definito del termine, questo lo si può brevemente riassumere come segue: Finsler considera un'ordinamento lessicografico di tutte le 'dimostrazioni formali' del fatto che, in una certa successione composta dai numeri 0 e 1, il simbolo 0 compaia infinite volte o del fatto che, viceversa, non vi compaia un numero infinito di volte.

Tale enumerazione induce un'enumerazione delle successioni binarie stesse, potendo noi associare ad ogni dimostrazione la successione binaria a cui quella prova si riferisce.

Se si considera a questo punto la successione antidiagonale, che cioè, per ogni n, differisce nell'n-esimo termine dall'n-esima successione dell'enumerazione indotta, segue dal fatto che questa successione non può rientrare nell'ordinamento delle successioni binarie (poiché differisce in almeno un elemento da ognuna di esse), che nessuna 'dimostrazione formale' relativa ad essa può far parte del novero di quelle considerate. Dunque, in particolare, la proposizione "Nella successione antidiagonale il numero 0 non occorre infinite volte" risulta essere 'indecidibile'.

Questa stessa proposizione peraltro, osserva Finsler, è anche falsa: ci sono un numero infinito di 'dimostrazioni formali', di successioni finite di parole logicamente ineccepibili, del fatto che nella successione $1, 1, \ldots, 1, \ldots$, lo 0 non occorre infinite volte; dunque, questa successione compare infinite volte nell'enumerazione indotta e, ad ogni occorrenza di essa, corrisponde, per come è stata definita, uno 0 nella successione antidiagonale.

Qui si rivela allora il paradosso che lega l'argomento di Finsler a quello

[34][Fin26], pp. 441-442 (*NdA*: i riferimenti a [Fin26], sono relativi alla ripubblicazione del lavoro di Finsler in [vH67], come indicato alla voce bibliografica).
[35]*Ibid.*, p. 443.

di Richard: perché in questo modo si è dopo tutto data una succesione finita di parole che apparentemente costituisce una 'dimostrazione formale' (nella specifico una confutazione), nel senso adottato da Finsler, per la proposizione risultata per altra via indecidibile.

Tuttavia, la contraddizione è per Finsler solo apparente e deriva dal fatto che la combinazione di parole in questione non può far parte dell'enumerazione di partenza di tutte le dimostrazioni valide, in quanto si riferisce ad una successione, l'antidiagonale, che non compare nell'ordinamento indotto. Per tale motivo, essendo partiti da un ordinamento di tutte le dimostrazioni "logicamente ineccepibili", conclude Finsler, essa non può essere tale.

La proposizione in questione è quindi 'formalmente' indecidibile e falsa, ed il paradosso, rivelatosi solo apparente se trasferito dall'ambito della 'definibilità finita' a quello della 'dimostrabilità formale', è così risolto.

Quello che è importante notare rispetto all'argomento appena descritto, è il fatto che esso fosse stato elaborato da Finsler con un intento esplicitamente fondazionale. Il paragrafo di apertura del lavoro del 1926 è infatti dedicato al programma di matrice hilbertiana basato sulle dimostrazioni finitarie di consistenza e all'osservazione di come, dalla prova dell'esistenza di una proposizione falsa e formalmente indecidibile, discenda un argomento che va contro l'idea che la non contraddittorietà di un sistema d'assiomi, ottenuta mediante la restrizione a metodi di prova finitari, costituisca una garanzia per la correttezza degli assiomi in questione (che segue, dato un simile enunciato, dall'ovvia possibilità di dare vita ad un sistema, ottenuto da quello di partenza per aggiunta della proposizione falsa, consistente ma scorretto).

Ciò costituisce un punto di contatto tra le posizioni con cui Finsler e Zermelo si presentarono a Gödel all'inizio degli anni '30: una comune avversione per la prospettiva finitista di riduzione della matematica ad un sistema finito di simboli.

Desta dunque curiosità che nonostante questo comune intento, la critica del progetto hilbertiano, animasse entrambi i personaggi, esso si associ nella discussione con Gödel sui teoremi di incompletezza ad un'evidente incomprensione degli aspetti squisitamente logici e metamatematici della questione.

Alla luce della breve ricostruzione del contenuto dell'articolo di Finsler del 1926, che permette di apprezzare, accanto alle analogie superficiali, le differenze decisive con l'argomento gödeliano, è possibile soltanto immaginare quale reazione possa aver avuto Gödel nell'apprendere, dalla lettera inviatagli l'11 marzo del 1933, la convinzione del matematico svizzero che la scoperta gödeliana altro non fosse se non la riproposizione, con differenze trascurabili, del proprio risultato:

> Sotto questo aspetto, Lei prende come punto di riferimento un formalismo più ristretto e dunque più preciso, mentre io, al fine di abbreviare la dimostrazione e mettere in luce solo l'essenziale, ho assunto un formalismo generale. Ovviamente, è comunque utile portare a termine queste idee entro una cornice formale particolare; nonostante ciò, ho evitato di farlo dal momento che il risultato mi sembrava essere stato stabilito e per questo motivo non mi è stato possibile sviluppare interesse sufficiente per il formalismo stesso.
>
> In più mi pareva che il mio risultato fosse di maggiore rilevanza relativamente alle sue applicazioni alla dottrina hilbertiana.[36]

La risposta di Gödel, della quale due *draft* preparatori sono stati ritrovati tra le carte del *Nachlass* e che quindi, come questi, doveva essere datata 25 marzo 1933, esclude, come è ovvio, in modo fermo ed inequivocabile qualsiasi legame tra l'articolo di Finsler ed i teoremi di incompletezza, mettendo in rilievo come l'ambiguità delle definizioni adottate da Finsler, non tanto permettono di "abbreviare la dimostrazione" e di "mettere in luce l'essenziale", quanto piuttosto rendono vano (e, a rigor di termini, fallace), ogni tentativo di vedervi un'anticipazione del lavoro di Gödel.

Riferendosi in particolare alla definizione data da Finsler del sistema S e delle nozioni ad esso connesse, Gödel nota infatti:

> Il sistema [S] con cui Lei lavora non è definito affatto in realtà, dal momento che, come è ben noto, alla questione di cosa sia una "dimostrazione logicamente ineccepibile" viene data una risposta diversa da matematici diversi. Ma [...] se anche si cercasse di applicare la Sua prova ad un sistema formale genuino P, questa risulterebbe essere sbagliata. Infatti la successione antidiagonale che Lei definisce e quindi anche l'enunciato formalmente indecidibile non è *mai* contenuto nello stesso sistema formale P di partenza.[37]

Gödel, come si evince dagli accenni contenuti nella bozza, cita il fatto che se così non fosse, infatti, ipotizzando cioè la definibilità della successione antidiagonale di Finsler nel sistema P e assunto quest'ultimo come non contraddittorio, si potrebbe definire in esso anche la successione nella quale lo 0 occupa ogni posto dispari e che, per ogni n, al $2n$-esimo posto presenti un numero che è diverso da quello che occupa il $2n$-esimo posto dell'n-esima

[36] [Göd03a], p. 406.
[37] *Ibid.*, p. 408.

successione nella lista delle successioni binarie per le quali esiste una prova del fatto che vi occorrono infiniti 0. Ma se così fosse, ci sarebbe una contraddizione (una formula falsa) dimostrabile già in P, dal momento che, dimostrabilmente in P, la successione in questione avrebbe infiniti 0, e quindi farebbe parte della collezione delle successioni da ognuna delle quali, per definizione, differisce in almeno un elemento.

Come nel caso di Zermelo, a poco servirono le spiegazioni chiarificatrici di Gödel. A queste osservazioni, anzi, Finsler rispose abbandonando il tono colloquiale della prima missiva e rendendo, nella collera, ancor più lampanti i propri limiti nella comprensione della memoria gödeliana:

> Vorrei replicare alle Sue osservazioni sul mio lavoro come segue: Se si intende fare delle asserzioni relative al sistema [S], non è affatto necessario che questo sia presentato in forma definita; è sufficiente che *lo si accetti come dato* e che si conoscano solo quelle poche proprietà dalle quali le conseguenze desiderate possono essere ottenute. Di questo si tratta nel caso in questione, e allora la presunta "vaghezza" del sistema non costituisce affatto un'obiezione. Con maggior magnanimità potrei caratterizzare il Suo articolo [...] come "completamente inutilizzabile" nel migliore dei casi, dal momento che non siete in grado di dimostrare la consistenza assoluta degli assiomi di cui fate uso in esso e in particolare degli assiomi di Peano. [...]
>
> Come potete poi concludere che la successione antidiagonale da me definita non sia *mai* formalmente rappresentabile nello stesso sistema formale di partenza, se il sistema [S] fornisce un controesempio? (Oppure, come si definisce in generale un sistema formale "genuino"?) Che per certi sistemi più ristretti si debba ricorrere a dimostrazioni differenti è comprensibile, ma questo non cambia la situazione nel suo insieme. Non riesco francamente a vedere una distinzione *di principio* tra la Sua prova e la mia.[38]

Seppur con le dovute differenze, i casi di Zermelo e Finsler possono essere avvicinati se non altro perché, come si diceva in precedenza, essi concordavano nel vedere dietro la concezione hilbertiana un'indebita riduzione della matematica e del problema del fondamento della disciplina, a questioni relative a sistemi di simboli costruiti su un linguaggio formale.

Da questo punto di vista, seppure da una prospettiva erronea, essi implicitamente ponevano la questione fondata del significato della scoperta di Gödel in relazione all'approccio finitista.

A tale proposito, soprattutto nel caso di Finsler, si credeva per di più di aver individuato un punto di vista più generale dal quale poter costruire una critica del programma di Hilbert. In altre parole, Zermelo (in misura

[38] *Ibid.*, p. 412.

minore) e Finsler credevano di poter mostrare, meglio di quanto non potesse fare Gödel, che *ogni* sistema logico è di fatto insufficiente per contenere tutto il discorso matematico, e, quindi, che esistono proposizioni formalmente indecidibili *in senso assoluto*, laddove Gödel aveva invece esposto nei dettagli una dimostrazione di questo fatto *relativamente ad un certo formalismo*.

In questo senso, pur rimanendo indubbiamente difficile passare sopra i limiti evidenti delle rispettive rivendicazioni, Zermelo e Finsler alludevano ad un problema *reale* che con la memoria di Gödel si era in effetti aperto e che era all'epoca all'attenzione della comunità dei logici: sgombrare il campo dal sospetto che i teoremi di incompletezza potessero dipendere dal formalismo dei *Principia mathematica* adottato, e confermare così le affermazioni di Gödel circa l'assoluta generalità di quanto da lui dimostrato.

È un tema questo, che emerge in modo migliore (e con una maggiore 'empatia' tra i protagonisti) dai contatti che Gödel ebbe con due personalità riconosciute in campo logico: Alonzo Church ed Emil Post.

Dai dubbi di Church alle certezze di Post

In [Göd31b] l'affermazione dell'indipendenza del risultato dal sistema scelto per la sua esposizione, era affidata al seguente passo, immediatamente successivo alla prova del primo teorema di Gödel (il Teorema VI dell'articolo):

> Nella dimostrazione del Teorema VI non si è fatto uso di alcuna proprietà del sistema P a parte le seguenti:
>
> 1. La classe degli assiomi e quella delle regole di inferenza (cioè, la relazione di "conseguenza immedita") sono ricorsivamente definibili[39] (fin tanto che si rimpiazzano in qualche modo i simboli primitivi mediante i numeri naturali).
> 2. Ogni relazione [primitiva] ricorsiva è definibile (nel senso del Teorema V[40]) nel sistema P.
>
> Quindi in ogni sistema formale che soddisfi le ipotesi 1 e 2 e che sia ω-consistente, esistono proposizioni indecidibili della forma $(x)F(x)$, dove F è una proprietà dei numeri naturali definita in modo [primitivo] ricorsivo, e ciò vale anche nelle estensioni di tali sistemi ottenute mediante una classe di assiomi ricorsivamente definibile e ω-consistente.[41]

[39] *NdT*: Come è noto, ciò che Gödel chiama "ricorsivo" in [Göd31b], viene oggi indicato con il termine "primitivo ricorsivo".

[40] *NdT*: Cioè nel senso della proprietà (1.1) da noi ricordata nell'esame del paragrafo introduttivo a [Göd31b].

[41] [Göd31b], p. 180.

Il fondamento dell'incompletezza, e persino quello dell'incompletabilità dei formalismi 'adeguati' per la matematica, era dunque già presente, come si poteva evincere anche dall'osservazione ulteriore di Gödel secondo la quale, tra i sistemi soddisfacenti le ipotesi 1 e 2 indicate, vi si trovavano tutti quelli noti fino ad allora a cominciare dai "sistemi d'assiomi per la teoria degli insiemi di Zermelo-Fraenkel e di von Neumann" (rispetto ai quali, come Gödel sottolinea nella nota 47 al testo, "la prova dell'assunzione 1 è persino più semplice"), per finire con "il sistema d'assiomi per l'aritmetica che consiste negli assiomi di Peano, le definizioni [primitive] ricorsive [...] e le regole della logica"[42].

Nonostante questi commenti, non mancò chi, pur avendo un'indubbia familiarità con i linguaggi formali ed i sistemi d'assiomi (e quindi rappresentando evidentemente una categoria ben diversa di quella dei casi di Zermelo e Finsler), rimaneva ancora convinto che il risultato di Gödel potesse dipendere in effetti dalle specificità della teoria dei tipi. Tra questi, vi era appunto Alonzo Church.

A dire il vero, la questione è solo marginalmente presente nei documenti che sono testimonianza di uno scambio di vedute diretto con Gödel, seguito alla pubblicazione di [Göd31b].

Si trova in essi solo un breve accenno, infatti, in chiusura di una lettera datata 17 giugno 1932 spedita da Church in risposta ad un messaggio da parte di Gödel il quale, dietro suggerimento di Oswald Veblen, aveva deciso di porre al diretto interessato due dubbi relativi al sistema presentato da Church nel suo [Chu32] e basato sul formalismo del λ-calcolo. Le due questioni riguardavano in particolare la possibilità di dimostrare nel sistema proposizioni esistenziali come l'assioma dell'infinito, e l'eventualità di poterne interpretare gli assiomi in un sistema di teoria dei tipi o di teoria degli insiemi al fine di dimostrarne la consistenza relativa.

La seconda osservazione di Gödel è particolarmente indicativa dal momento che, come è ben noto, il sistema di Church risulterà essere inconsistente[43].

Nel rispondere ai due quesiti il 27 luglio dello stesso anno, Church approfittò proprio del riferimento di Gödel al problema della consistenza per metterlo a conoscenza dei propri dubbi circa l'applicabilità al tipo di teoria formale da lui introdotta, del Teorema XI contenuto nel §4 della memoria

[42] *Ibid.*, p. 180.

[43] Questo verrà dimostrato da Kleene e Rosser (in [KR35]) solo più tardi rispetto alle date dello scambio epistolare tra Gödel e Church.

gödeliana (relativo, appunto, all'indecidibilità della proposizione che esprime la consistenza di P, nel sistema stesso).

Il passo della lettera in questione recita:

> Non sono stato in grado di verificare tuttavia, che la Sua conclusione nel §4 di [[Göd31b]] si applichi al mio sistema. È possibile che il Suo argomento possa venire modificato affinché ciò si realizzi, ma non sono riuscito a trovare una simile riformulazione della Vostra dimostrazione.[44]

Anche in un articolo presentato il 30 dicembre del 1933 ad un incontro congiunto della Mathematical Association of America e dell'American Mathematical Society a Cambridge (Massachussetts), Church avrebbe peraltro espresso gli stessi dubbi, a latere della presentazione di un sistema d'assiomi analogo a quello al centro delle questioni postegli da Gödel:

> Un sistema di questo genere [con differenti livelli di implicazioni] non solo sfugge al nostro indesiderato teorema secondo il quale esso deve risultare essere o insufficiente o ridondante, ma credo che sfugga anche all'ugualmente indesiderato risultato di Kurt Gödel secondo il quale, nel caso di un sistema logico che abbia pretese di adeguatezza, è impossibile dimostrarne la non contraddittorietà secondo le modalità pensate da Hilbert. Questo teorema di Gödel risulta essere infatti legato alle considerazioni precedenti più di quanto non appaia da ciò che è stato detto.[45]

Certo Church non poteva immaginare che le proprie difficoltà con la prova di Gödel derivassero, in particolare, dal fatto che il risultato in questione vale solo per sistemi consistenti. E comunque, pensando al fatto che la dimostrazione del secondo teorema di incompletezza data da Gödel era stata soltanto accennata[46], il dubbio da lui espresso poteva essere considerato, per l'epoca, in parte giustificabile.

Sappiamo tuttavia che le osservazioni di Church vanno inquadrate in uno scettismo complessivo (rivolto cioè non solo al secondo teorema di Gödel) e fondato su rilievi di merito (e quindi non dipendente in modo esclusivo dal fatto contingente che la critica veniva mossa in relazione ad un formalismo rivelatosi poi contraddittorio) sull'effettiva portata dei teoremi di incompletezza.

[44] [Göd03a], pp. 368-369.

[45] [Chu34], p. 360.

[46] A proposito delle difficoltà soggiacenti a questa parte del risultato di Gödel, si veda quanto ricordato alla nota 9.

È quanto emerge molto chiaramente infatti, da una lettera nota, scritta da Church a John Dawson il 25 luglio 1983[47]:

> [...] [E]ro tra coloro i quali ritenevano come il teorema di incompletezza di Gödel potesse risultare dipendere dalle specificità della teoria dei tipi (o, come avrei potuto aggiungere più tardi, della teoria degli insiemi) in un modo tale da rivelarne un carattere meno universale di quanto egli affermasse. C'era una ragione storica per questo, e cioè che anche prima che i risultati di Gödel venissero pubblicati, stavo lavorando ad una formulazione della logica radicalmente diversa che sarebbe sfuggita (così la vedevo io all'epoca) ad alcune restrizioni indesiderate della teoria dei tipi. In un certo senso stavo cercando di fare esattamente quello che Gödel dimostrò essere impossibile, ed è ovviamente un peccato che fui lento a riconoscere come il fatto che la prima dimostrazione di Gödel non si applicasse in modo esatto alla formulazione della logica che avevo in mente non fosse una cosa di grande valore.[48]

Alla reazione di Church fanno da contraltare le considerazioni di Emil Post: se le perplessità 'a caldo' del primo, si spiegano con il fatto che egli stava tentando di "fare esattamente ciò che Gödel dimostrò essere impossibile", è forse a causa del tentativo a lungo coltivato di anticipare la scoperta gödeliana, che, dal canto suo, il logico polacco realizzò invece immediatamente la portata epocale dei teoremi di incompletezza.

Lo scambio epistolare tra Gödel e Post risale a qualche anno più tardi rispetto a quello con Church, quando a seguito di un incontro tra i due, avvenuto il 29 ottobre 1938 ad una riunione dell'American Mathematical Society a New York, Gödel ricevette una prima lettera dal logico polacco scritta quello stesso giorno.

Con quella nota, Post intendeva scusarsi di essersi preso la libertà di aver avvicinato, al loro primo incontro, "l'uomo principalmente responsabile del dissolvimento di quel sogno", da lui caldeggiato, di "stupire il mondo matematico" con le proprie idee "poco ortodosse" e dimostrare l'esistenza di problemi irresolubili.

Post si riferiva indubbiamente a quei pionieristici lavori, rimasti a lungo inediti, che lo avevano portato agli inizi degli anni '20 del '900 ad accostarsi al problema della decisione, e dei quali doveva aver messo a conoscenza Gödel nel corso del loro incontro. Considerato il precedente di Finsler, Gödel sarebbe potuto rimanere comprensibilmente sorpreso dal tono di rispetto e ammirazione del logico polacco, il quale oltretutto, quasi a voler a

[47]La lettera è riportata per intero in [Sie97], pp. 177-178, ed a questa fonte si riferisce la citazione sottostante.

[48][Sie97], p. 177.

rassicurare Gödel da questo punto di vista, concludeva il breve messaggio escludendo esplicitamente ogni rivendicazione di priorità al proposito, nel riconoscere (non senza lusinghe nei confronti del proprio interlocutore) che "avrei *dimostrato* il Teorema di Gödel nel 1921 - fossi stato Gödel"[49].

I dettagli della vicenda a cui questa prima breve nota allude, sono ricostruiti dallo stesso Post in una lettera più distesa, scritta a Gödel il giorno seguente, ovvero il 30 ottobre 1938.

Alcuni dati tecnici e cronologici per completare il quadro a cui Post allude, possono essere brevemente riassunti come segue: sulla scia del teorema di completezza semantica e decidibilità per il calcolo proposizionale attraverso il metodo delle tavole di verità (risultato pubblicato nel 1921), Post si dedicò al tentativo di estendere i risultati ottenuti.

Come Church, anch'egli pervenne ad una propria formulazione della logica mediante i cosiddetti 'sistemi di produzioni', cioè sistemi di regole che permettono di operare opportuni cambiamenti (spostamenti di simboli) sulle 'parole' (successioni finite di simboli) di un certo 'alfabeto' (un insieme infinito numerabile di simboli)[50].

L'interesse per problemi combinatori di questo genere deriva, da un lato, dal fatto che i sistemi di Post consentono uno svluppo della teoria degli insiemi ricorsivamente enumerabili e risultano essere, quindi, un modo del tutto equivalente alle altre alternative per caratterizzare la nozione di 'funzione ricorsiva generale'; dall'altro, perché certi problemi algebrici possono essere ricondotti a problemi di manipolazione di parole costruite su un alfabeto, nel senso dei sistemi di produzione[51].

Post aveva individuato diversi tipi di sistemi (diversi a seconda delle produzioni, cioè delle regole di manipolazione, ammesse), ed aveva ottenuto un teorema che risulta rilevante proprio per le ricadute matematiche della teoria: ogni sistema basato su una produzione canonica, può essere ridotto ad un sistema di produzioni normale[52].

---

[49] [Göd03b], p. 169.

[50] I sistemi di produzioni, sono introdotti e trattati da Post in [Pos65] e [Pos43] (il primo dei due lavori, inviato dal logico polacco per la pubblicazione, fu respinto e pubblicato solo successivamente in [Dav65]).

[51] Si veda al proposito [Dav58], cap. 6.

[52] Ai fini del presente lavoro, i dettagli su cosa siano le produzioni 'canoniche' e 'normali' è tuttavia ininfluente. Si rimanda perciò ai riferimenti bibliografici indicati in precedenza per i dettagli.

Per mezzo di una congettura, la Tesi di Post, secondo la quale ogni insieme di successioni finite di simboli di un alfabeto generato da una procedura effettiva, e quindi ogni sistema logico, può essere generato da una produzione canonica (e coincide quindi con un sistema canonico), il risultato precedente poteva essere utilizzato a fini metamatematici: a partire da esso, infatti, egli concentrò i propri sforzi sull'indagine circa l'esistenza di problemi irresolubili in modo assoluto (ossia rispetto ad ogni sistema logico).

A tal fine, il ruolo della Tesi da lui avanzata risultava essere assolutamente cruciale, come, nella lettera del 30 ottobre del '38, Post riconosce in modo esplicito:

> Il problema in questione era un [E]ntscheidungsproblem. Presi quindi in considerazione ogni sistema logico nel quale il problema in questione potesse venire formulato. Sulla base delle riduzioni a cui ho fatto riferimento [nella nostra conversazione di ieri] assunsi che sistemi di questo genere fossero riducibili ad una 'forma normale' (equivalente, ne sono sicuro, alla ricorsività generale). L'assunzione che il sistema in questione fosse in grado di assegnare un'unica risposta si o no ad ogni problema costituente l'[E]ntscheidungsproblem, conduceva a contraddizione per mezzo dell'argomento diagonale di Cantor. Da qui la mia conclusione circa l'irresolubilità di quel problema in ogni 'logica simbolica' e quindi la sua indecidibilità assoluta.[53]

L'analogia delle proprie speculazioni con la scoperta di Gödel, il quale nella conversazione avvenuta il giorno precedente, come lo stesso Post ricorda all'inizio della lettera, "ogni volta che veniva fatto un paragone tra [i teoremi di Gödel] e i problemi assolutamente irresolubili" tendeva a sottolineare come "nel primo caso un enunciato in particolare risulta essere indecidibile"[54], è a detta di Post addirittura più profonda:

> Ma esaminando ulteriormente la fonte della contraddizione mi accorsi che essa conduceva ad un problema specifico della classe di problemi che costituiva l'[E]ntscheidungsproblem che non riceveva né un si né un no come risposta - assumendo che la logica fosse consistente, cioè che essa non assegnasse sia si che no come risposta. [...] Da qui la mia conclusione che non esiste una logica simbolica completa - [...] [nel senso che] relativamente ad un preciso e semplice insieme di enunciati, quelli dell'[E]ntscheidungsproblem in questione, [nessuna logica] riesce ad includere una risposta si o no per ognuno di essi.[55]

[53][Göd03b], p. 170.
[54]*Ibid.*, p. 170.
[55]*Ibid.*, p. 170.

Dunque, l'incompletezza sintattica di *ogni* sistema logico, seguiva come conseguenza del risultato di Post relativo all'irresolubilità assoluta della classe di problemi da lui individuata.

Come nel caso della breve nota citata in apertura, Post, al di là dell'espressione di un comprensibile rammarico, non intende così asserire una qualche priorità in materia, considerando che, come egli riconosce, "non sono le idee bensì la loro esecuzione a costituire il segno della grandezza"[56].

O ancora, come Post spiega più diffusamente (e indicativamente) in un altro passaggio:

> Vorrei dire infine che niente di ciò che io ho fatto potrebbe aver rimpiazzato la splendida attualità della Sua dimostrazione. Perché per quanto il Suo teorema sia un corollario dell'esistenza di problemi assolutamente irresolubili, l'irresolubilità assoluta del problema ha il solo fondamento, almeno nel mio lavoro e, ancora ritengo, in ogni altro lavoro, nella natura dell'induzione fisica. Naturalmente, con sufficiente operosità quest'induzione potrebbe essere portata al punto da includere teorematicamente il Suo sistema particolare. Che ciò potesse essere fatto per Principia Mathematica lo vidi già all'epoca.[57]

Ricapitoliamo: Post era dunque in possesso, almeno 'potenzialmente', di una dimostrazione dell'esistenza di una classe di problemi irresolubili mediante ogni sistema normale di produzioni; osservando più da vicino la dimostrazione, egli era giunto ad individuare un particolare tipo di problemi che non potevano essere decisi in alcun sistema normale e quindi in nessun sistema canonico; per mezzo della propria Tesi, egli era così arrivato ad individuare una classe di enunciati indecidibile in modo assoluto, cioè rispetto ad ogni sistema logico consistente.

Considerato che Post sottolinea di essere convinto che i sistemi normali di produzioni non sono altro che una caratterizzazione equivalente della ricorsività generale, la Tesi di Post è una versione della Tesi di Church, formulata esplicitamente da quest'ultimo nel 1936[58]: ogni funzione intuitivamente calcolabile è ricorsiva generale.

Cosa intendeva dunque dire Post, sottolineando come, dal momento che l'intera inferenza relativa all'esistenza di problemi irresolubili in senso assoluto dipendeva dalla propria Tesi, l'intero argomento (e ogni possibile al-

[56] *Ibid.*, p. 172
[57] *Ibid.*, p. 171.
[58] In [Chu36].

ternativa ad esso) aveva il "solo fondamento" in un ragionamento di tipo induttivo?

A tale proposito, Post si era già pronunciato nel 1936 nel dare la sua "formulazione 1" dell'analisi dei processi combinatori finiti, un tentativo al quale egli allude, nella lettera spedita a Gödel, come all'idea di "analizzare 'tutti i processi finiti della mente umana' (quel tipo di cosa che Turing fa nel suo articolo sui numeri computabili[59])"[60].

In quell'occasione, Post aveva sottolineato come, al fine di verificare la correttezza di una simile definizione (di offrire cioè una verifica delle sue ambizioni di "fedeltà psicologica"), "vengono prese in considerazione formulazioni sempre più ampie", rispetto alle quali "il nostro scopo sarà quello di dimostrarne la riducibilità logica alla formulazione 1". Sebbene Post riconoscesse a questa prospettiva allo stato il ruolo di "*ipotesi di lavoro*", egli indicava altresì il fatto che il successo del programma descritto ne avrebbe cambiato radicalmente la natura in una *legge naturale*, e concludeva significativamente:

> Solo in questo modo, sembra all'autore che il teorema di Gödel relativo all'incompletezza delle logiche simboliche di un certo tipo generale e il risultato di Church sull'irresolubilità ricorsiva di certi problemi, possono essere trasformati in conclusioni riguardanti tutte le logiche simboliche e tutti i metodi di irresolubilità"[61]

Dunque, è proprio attraverso una procedura di progressivo convincimento su base induttiva, derivante dal fatto che per quante estensioni della formulazione data si possano prendere in considerazione, è la stessa Tesi, a meno di versioni ad essa equivalenti, ad essere verificata, che si raggiunge la certezza circa l'assoluta generalità di certi teoremi logici.

E che, nel caso del teorema di incompletezza, di questo si tratti, ossia di un risultato di una tale portata generale, Post lo dice chiaramente nella lettera inviata a Gödel nell'ottobre del 1930:

> Naturalmente la Sua proposizione indecidibile in una logica ha un suo interesse intrinseco nella reinterpretazione come un enunciato che esprime la consistenza del sistema formale. Ho paura che proprio quest'interesse specifico ha portato sia ad un'interpretazione sbagliata circa il significato del Vostro teorema rispetto alle possibili dimostrazioni di consistenza, sia a negare ciò che credo sia il punto cruciale del Vostro risultato, l'esistenza di

[59] *NdT*: [Tur36].
[60] *Ibid.*, p. 171.
[61] [Pos36], p. 105.

> un enunciato indecidibile in ogni logica sufficientemente generale e, dunque, in una logica simbolica.[62]

Quindi, per Post, il risultato di Gödel correttamente inteso ha essenzialmente lo stesso significato del fatto da lui intuito anni prima, e quindi possiede quella valenza generale che Zermelo e Finsler, a causa della loro erronea lettura, non erano nella posizione di vedere, e che Church, a causa della (supposta) dipendenza dal formalismo dei *Principia*, era restio a riconoscere.

È pur vero che nel momento in cui Post scrive a Gödel, Church si era ovviamente convinto del fatto che il 'carattere universale' ascritto ai teoremi di incompletezza era perfettamente giustificato, come testimonia il seguente passo tratto dalla lettera datata 23 gennaio 1935 da lui spedita a Paul Bernays[63]:

> Gödel mi ha poi mostrato che il suo argomento può essere modificato in modo tale da rendere irrilevante l'uso di questa proprietà particolare del sistema dei Principia. Egli ha presentato questa forma generalizzata del suo argomento in una serie di conferenze a Princepton la scorsa primavera, ed è stato in grado di indicare un insieme di condizioni tale che il suo teorema si applichi ad ogni sistema logico che le soddisfi.

Almeno per quel che riguarda i dubbi avanzati da Church dunque, già con le conferenze tenute da Gödel presso l'Institute for Advanced Studies di Princeton nel 1934, la questione della generalità dei teoremi di incompletezza, sarebbe stata del tutto risolta.

Come risulterà chiaro al termine del prossimo paragrafo, quel che Gödel disse in quella circostanza può sì essere considerato un notevole chiarimento della questione, ma solo nel senso che quell'intervento mise in luce proprio quanto segue dalla lettera Post: l'esistenza di un passaggio cruciale, che poteva essere superato solo al costo di un impegno nei confronti di una Tesi, di un'assunzione di tipo concettuale analoga a quella adottata dal logico polacco.

[62] [Göd03b], p. 171.

[63] La lettera fa parte di materiale contenuto nel *Nachlass* di Paul Bernays conservato presso la Eidigenössische Technische Hochschule di Zurigo. Il passo riportato è citato in [Sie97], p. 161 e da questa fonte è stato ripreso e tradotto.

1.1.2 La nozione di sistema formale

Cosa aveva dunque detto Gödel durante le conferenze di Princeton per convincere Church dell'indipendenza dei teoremi di incompletezza dalle caratteristiche particolari del formalismo dei *Principia mathematica*?

Il confronto diretto tra il testo che riassume le lezioni di Gödel tenutesi tra il febbraio ed il maggio del 1934 e [Göd31b], fa emergere una differenza sostanziale fin dalle prime battute: anziché contenere una introduzione intuitiva alla dimostrazione del risultato, Gödel scelse presumibilmente in quell'occasione di prendere le mosse dalla definizione della nozione di "sistema matematico formale"[64].

Con questa espressione, sottolinea Gödel, si intende "un sistema di simboli con le regole per impiegarli"[65]. Accanto alla specificazione delle categorie sintattiche di 'termine', 'formula', 'assioma', 'dimostrazione', 'teorema', compare la seguente osservazione:

> Si richiede che le regole di inferenza, e le definizioni di formula ben formata e assioma siano costruttive; cioè, [si assume] che per ogni regola di inferenza esista una procedura finita per determinare se una data formula B sia una conseguenza immediata (rispetto alla regola in questione) di certe formule $A_1, \ldots, A_n$, e che esista una procedura finita per determinare se una formula A data sia una formula ben formata o un assioma.[66]

La nozione così individuata è a tutti gli effetti quel concetto inteso dai 'padri fondatori' della disciplina, da Frege a Peano.

Si tratta comunque, come si sarà notato, di un definizione intuitiva: un sistema formale si caratterizza per la decidibilità, mediante una "procedura finita", dell'insieme delle formule, degli assiomi e dell'insieme di quelle formule che sono conseguenza immediata di questi ultimi secondo le regole di inferenza del sistema.

Poco oltre nel testo, dopo aver esposto i dettagli dell'aritmetizzazione della sintassi utilizzando un sistema di funzionali come esempio per il proce-

[64] Il testo delle conferenze ([Göd65a]), come ricorda Kleene nella sua nota introduttiva ([Göd86a], p. 338), ha origine dagli appunti presi da Kleene stesso e da Rosser. Successivamente questi appunti furono integrati con due pagine di "Note e Correzioni" ad opera delle stesso Gödel che diede quindi la propria approvazione al testo. Al fine della pubblicazione in [Dav65], le due pagine di Gödel furono incorporate nel testo insieme a correzioni e integrazini ulteriori ad opera di quest'ultimo.

[65] [Göd65a], p. 346.

[66] *Ibid.*, p. 346.

dimento[67], si legge di un paragrafo dedicato alle "Condizioni che un sistema formale deve soddisfare" affinché ad esso si applichi l'argomento col quale si dimostrano i teoremi di incompletezza, e che sembra il candidato ideale per rappresentare il riferimento sottinteso da Church nel passo della lettera a Bernays di poc'anzi. Tra i requisiti in questione, Gödel indica in primo luogo il fatto che un sistema per il quale sia stata definita un'assegnazione di numeri ai simboli e alle successioni finite di essi, sia tale che l'insieme di assiomi e le relazione di conseguenza immediata (definita rispetto alle regole di inferenza), siano primitivi ricorsivi. E aggiunge:

> Questa è una precisa condizione che è in pratica un sostituto sufficiente per l'indefinito requisito del §1 secondo il quale la classe degli assiomi e la relazione di conseguenza immediata devono essere costruttivi.[68]

È pur vero che né questa né le altre condizioni enumerate da Gödel fanno riferimento alle specificità del formalismo presentato in precedenza, e che quindi, su questa base, Church poteva ben scrivere a Bernays che Gödel gli aveva in effetti più dettagliatamente mostrato la generalità dell'argomentazione.

Tuttavia, il criterio indicato permette di concludere, con Post, l'esistenza di un enunciato indecidibile "in ogni logica simbolica", solo se si hanno anche ragioni che rendono *necessaria* l'identificazione tra i sistemi formali la cui classe degli assiomi e la cui relazione di conseguenza immediata di una regola di inferenza sono costruttive, cioè decidibili mediante una 'procedura finita', e i sistemi in cui le stesse sono primitive ricorsive.

Che questo passaggio fosse improprio se compiuto rispetto alla nozione di ricorsività primitiva, Gödel lo indica peraltro esplicitamente nella celebre nota 3 al passo in cui egli aveva, poco prima rispetto alla citazione precedente, alluso alla fondamentale proprietà delle funzioni primitive ricorsive che consiste nel fatto che il loro valore, per dati argomenti, può essere calcolato mediante una 'procedura finita':

> Il viceversa sembra essere vero se, oltre a recursioni in accordo con lo schema [di recursione primitiva], vengano ammesse recursioni di altra forma (ad esempio rispetto a due variabili simultaneamente). Questo fatto non può essere dimostrato, dal momento che la nozione di computazione finita non è definita, ma può servire come un principio euristico.[69]

[67] *Ibid.*, §3.
[68] *Ibid.*, p. 361.
[69] *Ibid.*, p. 348.

Laddove questo passo indica la necessità di estendere la nozione di funzione ricorsiva primitiva, esso si riconnette alla proposta gödeliana in questa direzione contenuta nel paragrafo conclusivo (§9) delle lezioni di Princeton, in cui si trova la definizione della classe delle funzioni "ricorsive generali", ottenuta da Gödel modificando un suggerimento, secondo quanto egli stesso sostiene, venuto dal matematico francese Jacques Herbrand[70]. La classe in questione viene definita come la collezione di quelle funzioni il cui valore per ogni possibile argomento possa venire calcolato mediante certi sistemi di equazioni.

La definizione adottata da Gödel può essere illustrata come segue.

Si consideri un alfabeto costituito dai simboli: 0 (la costante zero), succ (il funtore unario di successore), $v_0, \ldots, v_n, \ldots$ (variabili individuali), $f_0, \ldots, f_n$, ... (variabili per funzioni n-arie, per ogni $n \in \mathbf{N}$), $=$ (relazione binaria di uguaglianza) e simboli ausiliari. La nozione di *termine* è induttivamente definita a partire dallo 0, per chiusura sotto il successore e l'applicazione delle variabili funtoriali f_i. Tra i termini, i *numerali* sono, come al solito, le espressioni ottenute dalla sola costante 0 per applicazione reiterata dell'operazione succ.

Un'*equazione in* $g_1, \ldots, g_n$, è un'espressione della forma $a = b$ dove a e b sono termini le cui variabili funtoriali sono (al massimo) le $g_1, \ldots, g_n$. Con l'accezione *sistema di equazioni* si intende una successione finita di equazioni.

Si scrive $A \vdash B$ per dire che "l'espressione B è ottenuta dall'espressione A per applicazione delle regole **R1** e **R2**", dove le regole in questione, data un'espressione A, sono le seguenti:

(R1) sostituire le variabili individuali $x_1, \ldots, x_k$ di A con numerali $\overline{n}_1, \ldots, \overline{n}_k$;

(R2) se $f_i(\overline{n}_1, \ldots, \overline{n}_k) = \overline{n}$ (per qualche $i \in \mathbf{N}$), sostituire qualsiasi occorrenza di $f_i(\overline{n}_1, \ldots, \overline{n}_k)$ in A con $\overline{n}$.

L'individuazione di queste regole non era ovviamente casuale: si tratta infatti delle inferenze modellate su quelle sufficienti a calcolare il valore di una funzione numerica F per ogni possibile argomento, nel caso in cui questa sia definita ricorsivamente (cioè definendone il valore su un argomento per

[70] In realtà la vicenda relativa al ruolo di Herbrand per la nascita del concetto di ricorsività alla Gödel-Herbrand è più complessa. Si veda al proposito il carteggio tra i due personaggi riprodotto in [Göd03b], pp. 14-25, la nota introduttiva di Wilfried Sieg (*Ibid.*, pp. 3-13) e la letteratura ivi citata.

mezzo dei valori sull'argomento ad esso precedente) a partire da funzioni note[71].

Stabilito ciò, si dice che una funzione aritmetica F è *computabile* se e solo se esiste un sistema di equazioni $E_1, \ldots, E_k$ in $g_1, \ldots, g_l$ (in cui ogni E_i è della forma $b = g_l(a_1, \ldots, a_m)$ e g_l non occorre in $a_1, \ldots, a_m$), tale che, per ogni $n_1, \ldots, n_m \in \mathbf{N}$ valga:

$$F(n_1, \ldots, n_m) = n \Leftrightarrow E_1, \ldots, E_k \vdash \overline{n} = g_l(\overline{n}_1, \ldots, \overline{n}_m)$$

In una lettera a Kleene del 29 novembre 1935, Church ipotizza che la modifica della definizione di Herbrand, "che non ha alcuna connessione con il concetto di computabilità effettiva", potesse in qualche modo avvicinarsi alla "sola idea" che Gödel considerava all'epoca plausibile a proposito del problema di dare una definizione della nozione di 'funzione computabile', ovvero di 'procedura finita': quella di "indicare un insieme di assiomi che includano le proprietà comunemente accettate di questa nozione, e cercare di fare qualcosa su questa base"[72].

Sappiamo tuttavia che neanche rispetto a questo concetto esteso, Gödel intendeva all'epoca proporre, in accordo con la nota 3 di [Göd65a] citata in precedenza, l'equivalenza con la nozione informale di 'procedura effettiva di computo'. Questo fatto è stato spiegato dallo stesso Gödel in una lettera a Martin Davis del 15 febbraio 1965:

> Non ero affatto convinto, all'epoca delle conferenze [di Princeton], che il mio concetto di ricorsività comprendesse tutte le possibili forme di recursione: e infatti la dimostrazione di equivalenza tra la mia defìnzione e quella di Kleene [...] non è così banale.[73]

Chiaramente la definizione di Gödel possiede un alto grado di intuitività ed aderenza al concetto inteso per il fatto di proporre, come interpretazione della nozione di un procedimento eseguito secondo regole, l'idea di una funzione che sia calcolabile mediante inferenze che rendono ammissibili solo sostituzioni elementari. D'altra parte, se questo è tutto ciò che si può trovare nelle lezioni di Princeton e anche ammettendo che quanto detto da Gödel sia stato sufficiente a far superare a Church i propri dubbi (almeno fin quanto essi erano relativi alla dipendenza dei teoremi di Gödel dal formalismo della

[71] Un'esemplificazione di questo fatto la si può trovare in [Kle52], p. 262.
[72] Il brano è citato in [Dav82], p. 9.
[73] *Ibid.*, p. 8.

teoria dei tipi), un aspetto essenziale della questione rimaneva ancora aperto: come eliminare i dubbi residui relativi all'identificazione della nozione informale di 'procedura finita', che è alla base dell'idea di sistema formale d'assiomi, con l'estensione della ricorsività primitiva proposta da Gödel?

Nel breve articolo del 1936 dal titolo "Sulla lunghezza della dimostrazioni"[74], il cui contributo fondamentale è il riferimento al risultato del cosiddetto *speed up* delle teorie assiomatiche formali (la prova del fatto che, in parole povere, passando da una teoria assiomatica ad un sistema più forte è possibile accorciare considerevolmente la lunghezza delle dimostrazioni del primo sistema), Gödel stesso sembra voler riprendere il filo del discorso interrotto a Princeton, introducendo il concetto di "funzione computabile in un sistema formale".

L'idea appare a tutti gli effetti come un'elaborazione dell'*Entsheidungsdefinitsheit*, la rappresentabilità formale per funzioni e relazioni ricorsive primitive che Gödel aveva individuato nel 1931, e può essere quindi resa nel modo seguente: una funzione $f(x)$ è computabile in un sistema formale S se esiste un termine t_f del linguaggio di S tale che, per ogni $m \in \mathbf{N}$, $f(m) = n \Leftrightarrow \mathsf{S} \vdash t_f(\overline{m}) = \overline{n}$[75]. Il testo prosegue poi, con la definizione di una gerarchia di sistemi formali $\langle \mathsf{S}_i \rangle_{i \in \Omega}$ (dove Ω indica la classe dei numeri ordinali), rispetto alla quale Gödel descrive il risultato dello *speed up*, che è data nel modo seguente: "S_i è il sistema che contiene, insieme agli opportuni assiomi logici, variabili e quantificatori per i numeri naturali, per le classi di numeri naturali, per le classi di classi di numeri naturali, e così via, fino alle classi di i-esimo tipo, ma non contiene variabili per i tipi superiori"[76].

Dunque, in questa gerarchia S_1 è, in particolare, un sistema formale del prim'ordine per l'aritmetica, S_2 il sistema del second'ordine per la teoria degli insiemi (o, in alternativa, per l'analisi) ottenuto dal precedente estendendone il linguaggio nel modo indicato e aggiungendo l'assioma di comprensione, e così via.

In una nota aggiunta al testo nel corso della revisione delle bozze per la

---

[74] [Göd36].

[75] La definizione di Gödel non è così precisa e recita, testualmente, "una funzione $\phi(x)$ verrà detta essere *computabile in S* nel caso in cui per ogni numerale m esista un numerale n tale che $\phi(m) = n$ è dimostrabile in S" ([Göd36], p. 396.). La definizione data nel testo costituisce dunque una interpretazione del passo. Essa tuttavia, vuole avere solo valore illustrativo ed equivale, da questo punto di vista, ad un'eventuale formulazione alternativa per ciò che concerne la problematica concettuale che si intende far emergere.

[76] *Ibid.*, p. 396.

stampa, Gödel osserva a proposito delle definizioni date:

> È possibile inoltre dimostrare che una funzione computabile in uno dei sistemi S_i, o anche in un sistema di ordine transfinito, è già computabile in S_1. Dunque, la nozione 'computabile' è in un certo senso 'assoluta', mentre quasi tutte le nozioni metamatematiche altrimenti note (per esempio, dimostrabile, definibile, e così via) dipendono in modo essenziale dal sistema adottato.[77]

Nessun argomento viene fornito per la dimostrazione dell' 'assolutezza' del concetto definito, cioè una dimostrazione del fatto che una funzione 'computabile' in un qualsiasi sistema S_i della successione definita, sia già 'computabile' in S_1. Nel testo, Gödel si limita a sottolineare, confermando il legame con la rappresentabilità delle funzioni primitive ricorsive su indicata, come "tutte le funzioni definite per recursione, ad esempio, sono già computabili nell'aritmetica classica (cioè nel sistema S_1 della successione definita sotto)"[78].

Si osservi, tuttavia, che anche una prova dell'indipendenza del concetto così definito dalla gerarchia dei formalismi introdotta da Gödel, non sortirebbe effetti decisivi per la questione di cui ci stiamo occupando. Come infatti potremmo giovarci a tal fine della suddetta assolutezza? Potremmo ad esempio, per mezzo del Teorema di Forma Normale di Kleene dimostrare la coestensività della classe di funzioni così definita e le funzioni ricorsive generali alla Gödel-Herbrand. La forza di un simile argomento discenderebbe allora dal riconoscere nella prima nozione (computabilità in un sistema logico arbitrario - in quella gerarchia), una generalizzazione della seconda (computabilità in un sistema di equazioni). Si avrebbe cioè proprio quel tipo di giustificazione induttiva segnalato da Post: passando da una certa formulazione ad una apparentemente più generale, si dimostra che la classe di funzioni definita in quest'ultimo modo è riducibile a quella di partenza.

Con un elemento critico, però: la definizione di Gödel del 1936, infatti, chiama in causa sistemi formali nei quali, in linea con quanto sostenuto da Gödel a Princeton, l'insieme degli assiomi e la relazione di conseguenza immediata (rispetto alle regole di inferenza) sono ricorsive.

Saremmo quindi ancora debitori di un argomento indipendente a sostegno del passaggio dal concetto informale di *sistema matematico formale* alla nozione matematicamente chiara che si definisce in tal modo. Si dovrebbe

[77] *Ibid.*, p. 398.
[78] *Ibid.*, p. 396.

in pratica fornire proprio un argomento indipendente a sostegno dell'identificazione tra le 'procedure finite' e le funzioni ricorsivamente definibili, la stessa tesi che la dimostrazione in questione vorrebbe supportare. Ed è difficile immaginare, d'altra parte, come la dimostrazione di 'assolutezza' annunciata da Gödel possa non fare uso di strumenti dimostrativi che, come il Teorema di Forma Normale, già di per sé non rendano necessario assumere l'effettività delle inferenze dei sistemi d'assiomi che la definizione di 'funzione computabile in un sistema formale' chiama in causa.

I limiti dell' 'assolutezza' paventata da Gödel, risiederebbero quindi nella relatività della nozione rispetto al formalismo.

L'inferenza indicata poc'anzi, basata sull'equiestensionalità delle due classi di funzioni di Gödel, dipenderebbe infatti dalla seguente assunzione:

- I passi inferenziali alla base del processo deduttivo di un sistema formale devono essere ricorsivamente definibili.

ed appare dunque, in relazione al problema in questione, ad alto rischio di circolarità[79].

Sembra così che, dopotutto, il criterio basilare offerto da Gödel nelle conferenze di Princeton non rappresentasse poi una soluzione *definitiva* al problema di poter considerare i teoremi di incompletezza un risultato generale, valido cioè per "ogni logica simbolica", e che i dubbi di Gödel a proposito dell'inevitabile Tesi sottostante fossero giustificati: per quanto una simile congettura potesse apparire implausibile su base 'empirica', si poteva ancora pensare che la sostituzione del requisito di decidibilità mediante una procedura finita dell'insieme degli assiomi e della relazione di conseguenza immediata del sistema formale con la nozione analizzata in termini di ricorsività, fosse in un certo modo arbitraria.

[79] Riferendoci ai soli testi di Gödel, non si è fatto altro che cercare di far emergere quello che Wilfried Sieg, autore di analisi approfondite ed accurate dei vari aspetti del problema concettuale alla base della Tesi di Church (si vedano [Sie94], [SB96], [Sie97], [Sie01]), indica come "lo scoglio" che si pone di fronte al genere di argomento a cui si è fatta allusione nel testo. L'argomento che chiama in causa un concetto del tutto analogo a quello di 'funzione computabile in un sistema formale', è offerto da Church in [Chu36], pp. 100-102 (*NdA*: il riferimento è alla ripubblicazione dell'articolo di Church in [Dav65], come da voce bibliografica), a sostegno della propria versione della Tesi (cioè, l'identificazione delle funzioni computabili con le funzioni ricorsive alla Gödel-Herbrand). La dimostrazione dell'equivalenza tra le funzioni ricorsive generali e le funzioni calcolabili in un sistema formale definite come in [Göd36], viene presentata in [HB39], suppl. 2.

Qualcosa di realmente nuovo doveva accadere, nel panorama logico, per superare l'*empasse*.

1.2 L'analisi di Turing

Il ragionamento che costituisce un sostegno plausibile per l'affermazione gödeliana relativa all' 'assolutezza' del concetto di 'funzione computabile in un sistema formale' sembra essere, come si è già notato in precedenza, davvero in sintonia con l'idea espressa da Post al termine del suo [Pos36]: a prescindere da come si estenda un sistema d'assiomi per mezzo di variabili per ordini superiori e relativi assiomi, la classe di 'funzioni computabili' individuata per mezzo della definizione data rispetto al formalismo di partenza rimane invariata. Quindi la nozione in questione è indipendente dalla gerarchia dei sistemi così costruita, ed è, in tal senso, 'assoluta'[80].

Si può fare ricorso sempre alle parole di Post, quelle, stavolta, della lettera a Gödel del 30 ottobre 1938, per introdurre la 'svolta' rappresentata dal contributo di Alan Turing. Tra le ragioni infatti che Post adduce a parziale scusante per non aver pubblicato le proprie conclusioni relative all'esistenza di proposizioni formalmente indecidibili nel sistema dei *Principia mathematica*, egli inserisce la convinzione che il risultato di maggior rilievo, la dimostrazione dell'esistenza di problemi assolutamente irresolubili, potesse discendere da un'analisi di "tutti i processi finiti della mente umana".

"[*Q*]*uel tipo di cosa*", egli aggiunge parenteticamente, "*che Turing fa nel suo articolo sui numeri computabili*"[81].

Nella frase di Post è racchiuso l'elemento che distingue l'articolo di Turing dal titolo "Sui numeri computabili con un'applicazione all'*Entscheidungsproblem*"[82], dagli altri lavori ad esso affini, e che consiste nel fatto che il suo contributo non solo contiene una caratterizzazione ulteriore della classe delle funzioni ricorsive generali per mezzo di una traduzione in termini matematici dell'idea di un dispositivo meccanico calcolatore, ma fornisce un argomento a favore dell'adeguatezza della nozione matematica rispetto al concetto inteso

[80] Come si è detto, Gödel non offre alcuna dimostrazione del risultato sulle funzioni al quale egli allude. La terminologia utilizzata richiama alla mente un risultato di Georg Kreisel (cfr. [Kre65b]) che permette di caratterizzare da un punto di vista della teoria dei modelli le funzioni ricorsive come quelle *invariabilmente definite* su $\mathbf{N}$.

[81] [Göd03b], p. 171.

[82] [Tur36].

(quello di un uomo che esegue un computo), *mediante l'analisi diretta* di quest'ultimo.

Data la rilevanza del contributo di Turing per una vicenda che, come abbiamo visto, interessa un aspetto concettualmente essenziale della questione della generalità dei teoremi di incompletezza, riteniamo non sia superfluo dedicare questo paragrafo ad un breve esame della proposta di Turing, abbandonando così per il momento gli scritti di Gödel.

Il nucleo concettuale del lavoro di Turing, è quanto contenuto in un paragrafo, il nono, nel quale si offrono tre diversi tipi di argomentazioni volti a mostrare come "i numeri 'computabili' includano tutti i numeri che verrebbero naturalmente considerati come computabili"[83], ovverosia, sostituendo la terminologia dell'articolo con quella più familiare, come le funzioni Turing-computabili includano tutte le funzioni intuitivamente calcolabili.

Il problema dal quale dipende la correttezza di questa Tesi di Turing, è individuato in modo chiaro e diretto: si tratta di rispondere al quesito "Quali sono i possibili processi che vengono messi in atto nel calcolare una funzione?". Dei tre argomenti presentati in risposta, è il primo fra essi, che è indicato come "Un appello diretto all'intuizione" e che si propone di esaminare il più direttamente possibile il procedimento messo in atto da un 'calcolatore umano', sul quale vogliamo concentrare la nostra attenzione.

La disamina di Turing può essere schematicamente suddivisa in quattro parti: (1) l'individuazione delle condizioni che regolano il calcolo, (2) l'analisi delle operazioni coinvolte, (3) la conseguente definizione di una controparte matematica, (4) l'equivalenza tra questa e la nozione di 'macchina di Turing'.

L'oggetto dell'analisi è un uomo, il 'calcolatore', che, munito di carta, penna e gomma, si appresta ad eseguire un computo. Il punto di partenza di Turing è il problema di individuare i passi elementari, cioè *non ulteriormente scomponibili*, che governano il procedimento, e a tal fine egli adotta alcune restrizioni di partenza:

- anziché nelle due dimensioni di un normale foglio di carta, si assume che il calcolo venga eseguito su un supporto unidimensionale (un nastro), suddiviso in celle;

- si assume che solo un numero finito di simboli possa essere scritto nel corso della computazione.

[83] [Tur36], p. 135 (*NdA*: la numerazione delle pagine si riferisce alla ripubblicazione del lavoro di Turing in [Dav65], come indica la voce bibliografica).

Mentre la prima restrizione viene ritenuta inessenziale ai fini dell'argomentazione, la seconda viene ricondotta da Turing direttamente all'intento indicato: se le operazioni individuate devono essere elementari, allora occorre assumere che le configurazioni simboliche vengano riconosciute *in modo immediato*, così ché la fase del riconoscimento possa essere considerata come un'operazione basilare e non debba essere ulteriormente analizzata.

Questa convenzione nasconde in realtà uno degli elementi fondanti dell'intera analisi: il 'calcolatore' che esegue il processo è sensorialmente e fisicamente limitato nelle proprie capacità[84].

Un'idea del tutto simile è alla base di un'ulteriore restrizione, relativa ad un'altra componente essenziale del processo di calcolo:

- il numero degli "stati mentali" di colui che esegue il computo e che devono essere considerati rilevanti ai fini dell'esecuzione del medesimo, è finito.

Rispetto a ciò occorre notare due cose: la prima, è la motivazione addotta per questa restrizione, "[s]e ammettessimo un numero infinito di stati mentali, alcuni di essi sarebbero 'arbitrariamente vicini' e verrebbero confusi"[85], e che viene ricondotta da Turing alla questione delle limitazioni fisico-sensoriali a cui si alludeva in precedenza; la seconda è l'osservazione che Turing aggiunge poco dopo, proponendo lo stesso argomento sotto una nuova veste, la quale dipende essenzialmente dal fatto che "si evita di introdurre lo 'stato mentale' considerandone una controparte più fisica e definita" rappresentata da "una nota di istruzioni [...] che spieghino come continuare il lavoro"[86] nel caso in cui il 'calcolatore' si interrompa per una pausa. Questo, per sottolineare come il ricorso ad una simile nozione sembri più un 'accidente terminologico' dettato dalle circostanze, che non pare sottintendere alcuna velleità di tipo 'psicologista' o 'mentalista' dietro l'intento del matematico britannico.

Turing procede a questo punto, ad isolare le condizioni alle quali avviene

[84] È significativa in questo senso, la motivazione data da Turing per giustificare la seconda restrizione indicata nel testo (*Ibid.*, pp, 135-136): l'incapacità di cogliere configurazioni simboliche complesse in ogni loro aspetto ed in modo immediato, come si può facilmente illustrare considerando due successioni di simboli, ad esempio 99999999999999 e 999999999999999, e riflettendo sulle difficoltà che si hanno nel dirimere la questione se esse indichino o meno la stessa cifra, con un solo colpo d'occhio.

[85] *Ibid.*, p. 136.

[86] *Ibid.*, p. 139.

il processo di computo e dalle quali dipende il fatto che il procedimento risulti essere *deterministico* e *finitario*[87]:

(D) "Il comportamento del calcolatore [umano] è determinato in ogni momento dai simboli che sta osservando, e dal suo "stato mentale" nell'istante."

(F1) "Si può supporre che c'è un limite B al numero di simboli o delle celle che il calcolatore [umano] può osservare in ogni momento."

(F2) "Si supporrà inoltre che il numero degli stati mentali che occorre prendere in considerazione sia finito."[88]

Per quel che riguarda poi le operazioni coinvolte, quelle "che sono così elementari che non sia facile immaginare come possano essere ulteriormente suddivise", esse consistono in "qualche cambiamento nel sistema fisico costituito dal calcolatore [umano] e dal suo nastro"[89].

Si tratta quindi delle seguenti alternative:

- la sostituzione di un simbolo presente su una delle celle osservate
- il cambiamento delle celle osservate con altre celle del nastro
- una variazione dello 'stato mentale' del 'calcolatore'

e di quelle derivanti dalla combinazione delle prime due con la terza.

In accordo con l'idea delle limitate capacità sensoriali del 'calcolatore' di cui si è detto, Turing ritiene di dover assumere che le operazioni indicate debbano rispettare le seguenti condizioni di *località*:

(L1) "Si potrebbe quindi supporre, senza perdita di generalità, che le celle i cui simboli vengono cambiati sono sempre celle osservate."

(L2) "Penso sia ragionevole supporre che [le nuove celle osservate] possano essere solo celle la cui distanza da quelle più vicine alle celle osservate immediatamente prima, non ecceda una certa quantità."[90]

[87]Si sta procedendo, sulla falsariga della chiara e convincente analisi che dell'argomento di Turing ha dato Wilfried Sieg in una serie di lavori (si vedano ad esempio [Sie94], [SB96], [Sie01]).
[88]*Ibid.*, p. 136.
[89]*Ibid.*, p. 136.
[90]*Ibid.*, p. 136.

Il passo successivo di Turing consiste quindi nell'indicare un modello di macchina che sia in grado di svolgere il computo alle condizioni individuate.

La descrizione di quello che dovrebbe costituire la controparte matematicamente trattabile del 'calcolatore'[91], è, sulla base dell'analisi compiuta, estremamente sommaria e procede semplicemente facendo corrispondere agli elementi isolati le parti di un dispositivo che è direttamente ispirato alle 'macchine' definite da Turing nei paragrafi iniziali del lavoro: il nastro diviso in celle corrisponde al supporto unidimensionale sul quale avviene il computo, le operazioni della macchina (con le dovute restrizioni) alle operazioni del 'calcolatore', gli stati del dispositivo, o configurazioni, agli stati mentali.

L'osservazione che "le macchine così descritte non differiscono in modo essenziale dalle macchine calcolatrici definite al §2", cioè dalle 'macchine di Turing' comunemente note, conduce così alla conclusione dell'argomentazione secondo la quale "in corrispondenza di ogni macchina di questo tipo, può essere costruita una [macchina di Turing] che computi la stessa successione, ossia la stessa successione computata dal calcolatore [umano]"[92].

La letteratura critica al proposito[93], indica la differenza tra le 'macchine di Turing', che comunemente vengono definite per operare su un simbolo alla volta (e per tale motivo possono essere dette *letter machines*), e le macchine che operano nel rispetto delle condizioni **(D)**, **(F)** ed **(L)**, con il fatto che queste ultime possono invece operare su successioni finite di simboli (e per questo motivo, vengono chiamate *string machines*).

Quindi, il risultato conclusivo a cui Turing allude, altro non è che la direzione non banale di una dimostrazione di equivalenza tra le 'macchine di Turing' e queste *string machines*.

Una volta che il testo di Turing venga scomposto ed analizzato nel modo proposto, si giunge così ad individuarne la struttura così come visualizzato dalla figura 1.1: il procedimento di computo eseguito da un calcolatore umano viene esaminato direttamente; quest'analisi concettuale conduce ((1) in figura) alla tesi secondo la quale 'ogni computo di questo tipo, soddisfa le condizioni di determinatezza, finitezza e località'; si afferma che un calcolo che soddisfa **(D)**, **(F)** e **(L)** può essere eseguito da una 'macchina di Turing' che operi su configurazioni finite di simboli (passo (2)); si dimostra, o, meglio, si allude alla dimostrazione del fatto che ogni computo eseguito da una

[91]Cfr. *Ibid.*, p. 137.
[92]*Ibid.*, p. 138.
[93]Si veda la nota 87.

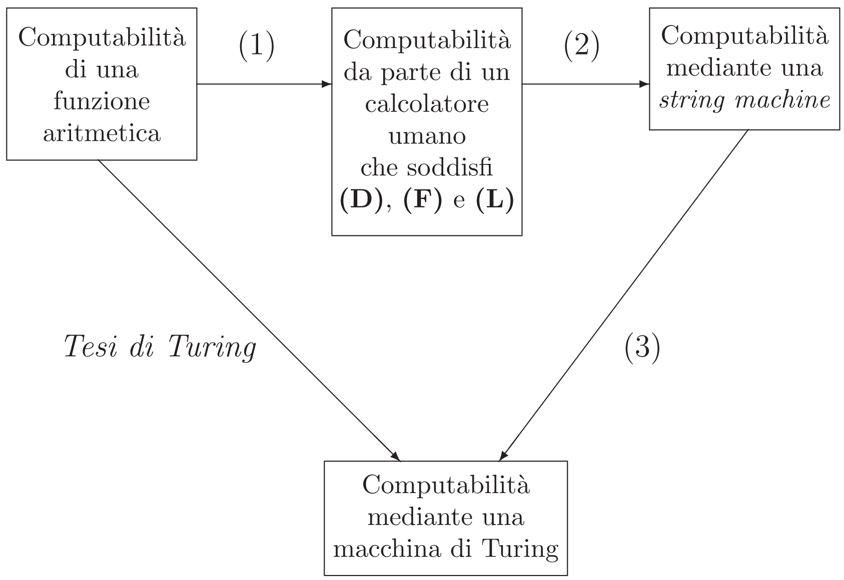

Figura 1.1: La struttura dell'argomento di Turing secondo Sieg.

string machine possa essere eseguito su una normale 'macchina di Turing' ((3)).

Non c'è alcuna traccia, per tornare alla questione sulla quale avevamo chiuso il paragrafo precedente, né del fondamento induttivo di Post, né tantomeno di un sentore di circolarità: sulla base dell'esame della nozione intesa, se ne individua una precisa controparte matematica e si dà un teorema di l'equivalenza tra questa e la definizione proposta inizialmente.

Si sarà notato come l'intera argomentazione dipenda da assunzioni, 'ogni calcolo eseguito da un matematico umano soddisfa le condizioni isolate' e 'ogni calcolo che soddisfa le condizioni isolate può essere eseguito da una *string machine*', che (per quanto indubitabilmente esenti da un vizio di circolarità) denotano il carattere essenzialmente informale dell'inferenza. Un fatto, quello dell'impossibilità di fornire una soluzione *matematicamente* soddisfacente, di cui Turing era ben consapevole, e, come lui, anche tutti gli altri protagonisti di cui si è detto finora. Un fatto che, in ogni caso, nulla toglie al merito del matematico britannico: aver assottigliato in modo decisivo le distanze tra la nozione informale e la sua pretesa controparte matematica, al punto tale da rendere trascurabili quelle rimanenti (a detta della gran parte della comunità logica del suo tempo).

È stato lo stesso Gödel, d'altra parte, ad indicare, con una precisione che si attaglia perfettamente alla struttura individuata, l'aspetto innovativo dell'argomento di Turing che ne spiega il valore 'rivoluzionario'.

Nel poscritto ideato per la pubblicazione del testo delle lezioni di Princeton in [Dav65], egli scrive infatti che:

> Il lavoro di Turing fornisce un'analisi del concetto di 'procedura meccanica' (alias 'algoritmo' o 'procedura di computazione' o 'procedura combinatoria finita'). Si mostra che questo concetto è equivalente a quello di 'macchina di Turing'[94]

È grazie a questo (e all'opera di Kleene) che, "se 'procedura finita' viene ritenuto significare 'procedura meccanica' ", alla questione sollevata nella nota 3 di [Göd65a], se cioè ogni 'procedura finita' sia calcolabile mediante una funzione definita per recursione, "può essere data una risposta affermativa per la ricorsività definita come nel §9 [ricorsività alla Gödel-Herbrand] che equivale alla ricorsività così come viene definita oggi [cioè, alla μ-ricorsività]"[95].

Ed è grazie a ciò, soprattutto, se la questione della formulazione generale dei teoremi di incompletezza può dirsi definitivamente risolta:

> [Poiché] una precisa ed indiscutibilmente adeguata definizione di sistema formale può essere adesso data, è possibile dimostrare rigorosamente per *ogni* sistema formale consistente che contenga una certa quantità di aritmetica finitaria, l'esistenza di proposizioni indecidibili e l'indimostrabilità della [proposizione che esprime la] consistenza di un sistema nel sistema stesso.[96]

[94] [Göd65a], pp. 369-370.
[95] *Ibid.*, p. 370.
[96] *Ibid.*, p. 369.

1.3 Una versione generale dei teoremi di incompletezza (2): i matematici e le macchine

All'alba del 1937[97] dunque, il problema della formulazione dei teoremi di incompletezza in tutta la loro generalità non nascondeva più alcuna questione in sospeso. Anche se il risultato di Gödel aveva già acquisito pienamente la propria fama, un tassello di quella scoperta, rilevante da un punto di vista concettuale, veniva così a trovare la propria definitiva collocazione.

Fino alla pubblicazione di parte del *Nachlass* gödeliano nel terzo volume dei *Collected Works*, poteva certo stupire che, al contrario di quanto era lecito aspettarsi considerato il peso da lui stesso riconosciuto in più occasioni al ruolo essenziale ricoperto dal contributo di Turing, Gödel non avesse colto l'occasione per soffermarsi sul significato della propria scoperta subito dopo quella data.

D'altra parte, dal 1931 l'attenzione del logico austriaco si era rivolta ad altro, sia in campo squisitamente logico (i rapporti tra logica e matematica classica ed intuizionista con la cosiddetta 'traduzione negativa', il problema della decisione per particolari classi di formule) che in quello più matematico-fondazionale (il problema dell'Ipotesi del Continuo di Cantor, l'interpretazione 'Dialectica' e le possibili revisioni del programma hilbertiano). Solo molto più tardi, a partire dagli anni '50, esaurita la 'spinta' delle questioni squisitamente insiemistiche e sulla scia del prevalere dell'interesse per problematiche di carattere filosofico, egli tornerà sulla questione della corretta lettura dei teoremi di incompletezza, vuoi perché direttamente stimolato da quanti lo interpellavano al proposito, vuoi perché preoccupato di aggiornare il quadro delle acquisizioni scientifiche rilevanti, in vista della ripubblicazione di alcuni articoli storici[98].

[97] Nonostante il volume 42 dei *Proceedings of the London Mathematical Society* sia stato pubblicato a cavallo tra il 1936 ed il 1937 (e consistentemente risulti annotato come volume del 1936-37), l'articolo di Turing fu ricevuto nel maggio del 1936, letto nel novembre dello stesso anno e pubblicato in due parti, la prima il 30 novembre, la seconda il 23 dicembre 1936.

[98] Come nel caso di [Göd31b] e [Göd65a], rispettivamente in [vH67] e [Dav65] (il primo dei due in traduzione inglese, il secondo pubblicato per la prima volta). I due volumi dei *Collected Works* dedicati alla corrispondenza, ad esempio, offrono molteplici spunti per apprezzare l'attenzione che Gödel prestava ogni qual volta era costretto a tornare sul significato dei teoremi di incompletezza che, anche a causa della fortuna acquisita,

Tra gli scritti inediti, esiste tuttavia un piccolo lavoro che appartiene presumibilmente al periodo storico del quale ci stiamo occupando e che contiene, invece, proprio alcune preziose indicazioni relativamente all'interpretazione di Gödel di una versione generalizzata del risultato di indecidibilità[99].

Il testo, evidentemente ancora allo stato di bozza seppure per certi aspetti già definito nel suo contenuto, può essere suddiviso in tre parti principali ed una coda. Le tre parti sono: (1) l'esposizione del significato, matematico e fondazionale, dei teoremi di incompletezza[100]; (2) la ricostruzione del percorso che ha portato alla precisa definizione dei concetti di 'procedura meccanica' e 'sistema formale'[101]; (3) la prova di un teorema che permette di stabilire in modo matematicamente rilevante il risultato di indecidibilità[102]. La coda[103], contiene invece un'osservazione estremamente interessante che mette in relazione i teoremi di incompletezza con il lavoro di Gödel nel campo della teoria assiomatica degli insiemi.

È lo stesso Gödel, in apertura, a sottolineare come l'interesse del contributo non stia tanto nella presentazione di un nuovo risultato, quanto nel valore che esso assume nel riassumere quanto emerge da pubblicazioni separate:

> Quasi tutto quello che andrò a dire in questa conferenza è stato dimostrato e pubblicato diversi anni fa, ma le pubblicazioni sono sparse in differenti articoli e non è quindi forse inutile raccogliere tutti questi risultati insieme così da valutare l'intero spettro delle questioni. La sola cosa nuova è una certa semplificazione dell'enunciato indecidibile che di principio non è di grande importanza ma che può essere interessante per i teorici dei numeri.[104]

---

venivano sottoposti talvolta a bizzarre letture (si possono vedere ad esempio i documenti relativi alla corrispondenza con Balas, Brutian, Church (lettera 3), Follett in [Göd03a], Nagel, Rappaport, Plummer in [Göd03b]. Alcune di queste testimonianze, verranno citate ed utilizzate nel seguito).

[99] Il materiale proveniente dal *Nachlass* gödeliano a cui intediamo fare riferimento, è stato pubblicato nel terzo volume dei *Collected Works* ([Göd95], pp. 164-175). Si tratta di poche pagine manoscritte ritrovate senza indicazione di titolo o data. L'impostazione del testo (che contiene anche l'indicazione di materiale da scrivere sulla lavagna), fa pensare che fosse stato preparato per una conferenza. Approssimativamente, esso appartiene alla fine degli anni '30 (si veda la nota introduttiva di Martin Davis - in particolare, p. 163 - per qualche congettura al riguardo).

[100] [Göd38c], pp. 164-166.

[101] *Ibid.*, pp. 166-168.

[102] *Ibid.*, pp. 170-174.

[103] *Ibid.*, p. 175.

[104] *Ibid.*, p. 164.

Gödel passa quindi ad illustrare il risultato, partendo però dalla ricostruzione del quadro concettuale che lo rende significativo.

Lo spunto a cui egli ricorre per introdurre la questione è inconsueto ed intrigante ad un tempo. Si tratta infatti dell'"assioma di risolubilità di ogni problema matematico" che David Hilbert aveva formulato al Congresso Internazionale dei Matematici svoltosi a Parigi nel 1900 come "la convinzione che ogni problema matematico determinato debba essere necessariamente passibile di una rigorosa sistemazione, o riuscendo a dare la risposta alla questione posta oppure mostrando l'impossibilità di una sua soluzione e quindi la necessità dell'insuccesso di ogni tentativo"[105].

L'affermazione di Hilbert, soprattutto se letta in continuità con il tipo di programma fondazionale da lui proposto negli anni successivi, può essere vista come il frutto di un sogno accarezzato lungamente nel corso della storia del pensiero, e il cui precedente più illustre è probabilmente l'idea di Leibniz, esposta per la prima volta nella *Dissertazione sull'arte combinatoria* del 1666, che fosse possibile dare vita ad una 'lingua universale' o 'caratteristica', comune a tutti gli uomini, la quale, privata delle componenti emotive proprie dei linguaggi naturali, avrebbe esaltato la possibilità di comunicazione delle conoscenze razionali fino al punto da garantire la risoluzione di ogni disputa scientifica mediante strumenti puramente deduttivi.

Gödel ricorda anche come Hilbert fosse così convinto della propria posizione da ritenere che di questo fatto si potesse persino giungere ad una dimostrazione matematica (almeno per il caso dell'aritmetica).

Quello che segue è uno stringato sunto del percorso concettuale svoltosi in quelli che sono stati alcuni tra i decenni più fertili in campo logico, e che, grazie allo sviluppo della teoria della dimostrazione di marca hilbertiana, conduce da quella 'primordiale' affermazione alla sua formulazione precisa come *Entscheidungsproblem*, o 'problema della decisione'.

Immaginiamo infatti, dice Gödel, di voler verificare la convinzione di Hilbert circa l'esistenza di una prova matematica per la sua assunzione. Il problema deve essere innanzi tutto formulato in modo tale da renderlo passibile di una trattazione matematica. Occorre allora, in primo luogo, riconoscere come, almeno "[p]er ogni uomo privo di pregiudizi", l'affermazione di Hilbert non abbia altro significato se non che "[d]ata una proposizione matematica

[105] [Hil78], p.153 (*NdA*: il riferimento è alla ripubblicazione in traduzione italiana del testo dell'intervento di Hilbert nella raccolta curata da V. M. Abrusci, come da voce bibliografica).

arbitraria A esiste o una prova di A o una prova di non-A, dove con 'prova' si intende qualcosa che parte da assiomi evidenti e procede per inferenze evidenti"[106].

Posto in questo modo tuttavia, il problema non ha ancora assunto quella forma che possa permettere una soluzione matematica, in quanto in esso ricorre comunque la nozione informale di 'evidenza'.

Il passo successivo, allora, è quello di applicare a quest'ultima il tipo di analisi che proviene dalla logica, così da ottenere la formulazione seguente della tesi (che è lasciata implicita nello scritto gödeliano): esiste un sistema formale S tale che, per ogni enunciato matematico A, o esiste in S una dimostrazione di A, o una dimostrazione di $\neg A$. È in questa versione che la convinzione di Hilbert (anche limitata alla sola teoria dei numeri) può essere confutata sulla base dei teoremi di incompletezza, parafrasando i quali, nessun sistema 'adeguato' per l'aritmetica può essere anche sintatticamente completo, come invece richiesto.

Tuttavia, nota Gödel, una simile risposta negativa all'aspettativa hilbertiana può avere *due* significati diversi. Può cioè voler dire in primo luogo "che il problema nella sua formulazione originaria ha una risposta negativa", oppure "che nel passaggio dall'evidenza al formalismo qualcosa è andato perduto". Ed è facile convincersi, secondo Gödel, che è proprio questo secondo il caso, "dal momento che le questioni aritmetiche che sono indecidibili in un dato formalismo sono sempre decidibili mediante inferenze evidenti che non sono esprimibili in esso"[107].

L'osservazione di Gödel richiama alla mente altre affermazioni analoghe, prima fra tutte l'osservazione già presente alla nota 48a di [Göd31b]:

> Come verrà mostrato nella Parte II di questo articolo, la vera ragione dell'incompletezza inerente a tutti i sistemi formali della matematica è il fatto che la formazione di tipi superiori può essere proseguita nel transfinito [...], mentre in ogni sistema formale è disponibile solo un numero al massimo numerabile di essi. Infatti può essere mostrato che l'enunciato indecidibile qui costruito diviene decidibile ogni qual volta vengano aggiunti tipi superiori appropriati (ad esempio, il tipo ω al sistema P). Una situazione analoga si verifica nel caso degli assiomi della teoria degli insiemi.[108]

È possibile quindi decidere formalmente l'enunciato, indecidibile in un sistema formale di partenza, estendendone il linguaggio mediante un ulteriore

[106] [Göd38c], p. 164.
[107] *Ibid.*, p. 154
[108] [Göd31b], p. 180.

sorta di variabili (per gli oggetti di tipo esattamente superiore a quelli per i quali il sistema possiede già variabili) e opportuni nuovi assiomi.

Ciò che segue quindi dai teoremi di incompletezza è quindi il fatto che non è possibile racchiudere l' 'evidenza' matematica in un unico sistema formale. Equivalentemente, prosegue Gödel nella sua disamina, si può dire che dai teoremi di incompletezza discende il fatto che "non è possibile formalizzare il ragionamento matematico, cioè non sarà mai possibile rimpiazzare il matematico con una macchina, anche confinandosi ai problemi aritmetici"[109] (dove appare essenziale, per riconciliare questa lettura con l'osservazione precedente, che Gödel sottolinei come non sia possibile sostituire il matematico con *una* macchina, ossia un *singolo* formalismo).

Questo significa in particolare che nel campo dell'evidenza, cioè nel campo della nozione di base presa nel suo senso intutivo, l'affermazione di Hilbert rimane ancora del tutto plausibile. Non è quindi quell'ottimismo razionalista che pervade la visione hilbertiana che può essere contestato sulla base dei teoremi di incompletezza, bensì l'idea, che tuttavia Hilbert avrebbe forse sostenuto, che gli strumenti delle logica colgano in modo adeguato quella nozione di 'prova' a cui l'affermazione originaria fa riferimento, e quindi che una riformulazione di quella convinzione come 'problema della decisione', relativo all'esistenza di una cornice formale entro la quale poter decidere ogni enunciato matematico, ne sia una corretta controparte rigorosa.

L'analisi gödeliana di una simile soluzione negativa del 'problema della decisione' di Hilbert, che sembra volto ad evidenziare un carattere peculiare del procedimento dimostrativo in matematica che non può essere rappresentato formalmente, si combina peraltro in modo perfetto con le conclusioni raggiunte al proposito da altre personalità in campo logico.

Può infatti essere messa in relazione, ad esempio, con l'interpretazione data da von Neumann anni prima e che risulta essere speculare a quella di Gödel, considerando essa gli effetti (ipotetici) di una soluzione positiva dell'*Entscheidungsproblem*:

> [L]'indecidibilità [del fatto se una formula sia dimostrabile o meno] è anche la *conditio sine qua non* perché la pratica matematica contemporanea, che usa cioè come accade, metodi di tipo euristico, abbia un qualche senso. Il giorno in cui l'indecidibilità non sarà più valida, la matematica così come noi la intendiamo oggi cesserebbe di esistere; sarebbe rimpiazzata da una prescrizione assolutamente meccanica, per mezzo della quale ognuno

[109][Göd38c], p. 164-165.

> potrebbe decidere circa la dimostrabilità o l'indimostrabilità di una certa proposizione.[110]

La conclusione di Gödel appare molto simile anche a quell'osservazione che Post, nella più volte citata lettera del 30 ottobre del 1938, aveva indicato essere per lui l'implicazione della 'scoperta' dell'esistenza di problemi indecidibili in senso assoluto, ossia rispetto ad ogni sistema logico:

> Conclusi di conseguenza che la dimostrazione matematica è essenzialmente creativa per il fatto che, una volta messo a punto un sistema formale per esprimere il succitato [E]ntscheidungsproblem, si può sempre trascendere il sistema in questione [...] una conclusione raggiunta anche, credo, dal Vostro lavoro.[111]

In più, rispetto a questi esempi, Gödel sottolinea, passando a considerare l'aspetto più tecnico di [Göd38c], come una volta che questo tipo di soluzione negativa è stato raggiunto, "[l]'interesse sta naturalmente nel trovare, per così dire, la più piccola porzione della matematica che non è formalizzabile"[112].

A tal fine, egli isola la classe A dei problemi della forma:

$$\bigwedge x_1, \ldots, x_m \bigvee y_1, \ldots, y_n F(x_1, \ldots, x_m, y_1, \ldots, y_n) = 0$$

dove F è un polinomio a coefficienti interi e $\bigwedge \ldots \bigvee \ldots$ un'abbreviazione per 'per ogni ... esistono ...', rispetto alla quale egli enuncia due teoremi:

1. Non esiste una procedura meccanica di decisione per ogni enunciato della classe A.

2. In ogni sistema formale (corretto) che permette di esprimere tutte le proposizioni della classe A, ne esiste almeno una indecidibile.

Prima di passare alla dimostrazione però, Gödel si sofferma su un aspetto non secondario del risultato così formulato:

> Quando ho pubblicato per la prima volta il mio lavoro sugli enunciati indecidibili, il risultato non poteva essere annunciato con questa generalità dal momento che le nozioni di procedura meccanica e di sistema formale non avevano all'epoca ricevuto una definizione matematicamente soddisfacente. Questa lacuna è stata da allora colmata da Herbrand, Church e Turing.[113]

[110] [vN27], p. 12
[111] [Göd03b], p. 170-171.
[112] [Göd38c], p. 165.
[113] *Ibid.*, p. 166.

Gödel ripercorre quindi puntualmente tutte le tappe di cui ci siamo fino a questo momento occupati: dalla definizione della nozione di 'sistema formale' al rapporto tra questa e quella di 'funzione effettivamente computabile'; dall'individuazione delle funzioni ricorsive generali come soluzioni di opportuni sistemi di equazioni, all'analisi di Turing che ne stabilisce l'adeguatezza con la nozione informale. La conferma dell'assoluta rilevanza del percorso tracciato, traspare dalle parole riservate per quest'ultimo passaggio e per il ruolo ricoperto dal contributo di Turing in particolare:

> Che questa [ricorsività alla Gödel-Herbrand] sia davvero la definizione corretta di computabilità meccanica è stato stabilito oltre ogni dubbio da Turing. [...] Egli ha mostrato come le funzioni calcolabili definite in questo modo sono esattamente quelle per le quali è possibile costruire una macchina con un numero finito di parti che sarà in grado di fare quanto segue. Scritti su un nastro di carta numeri $n_1, \ldots, n_r$ qualsiasi, messo il nastro nella macchina e avviato il meccanismo, dopo un numero finito di passaggi la macchina si fermerà ed il valore della funzione sull'argomento $n_1, \ldots, n_r$ sarà stampato sulla carta.[114]

Segue, quindi, la dimostrazione dei due teoremi indicati.

La prova si compone di due passaggi essenziali: dal Teorema di Enumerazione di Kleene[115], segue che, per ogni insieme ricorsivamente enumerabile D, esiste un $m \in \mathbf{N}$ tale che:

$$x \in D \Leftrightarrow \bigvee yT_1(m, x, y) \tag{1.3}$$

dove T_1 è un predicato ricorsivo primitivo.

È una semplice applicazione di questo risultato, concludere allora che l'insieme $K := \{n \in \mathbf{N} \mid \bigwedge y \neg T_1(n, n, y)\}$ non può essere ricorsivamente enumerabile (perchè se lo fosse allora, per (1.3), per qualche $k \in \mathbf{N}$, si avrebbe $x \in K \Leftrightarrow \bigvee yT_1(k, x, y)$, da cui la contraddizione segue immediatamente per $x = k$). Ciò implica che la classe di problemi $B := \{\bigvee yT_1(n, n, y) \mid n \in \mathbf{N}\}$ non è decidibile, ossia vale per essa il teorema 1 enunciato in precedenza.

Quel risultato (rispetto alla classe A) discende allora come corollario dal seguente lemma fondamentale che Gödel dimostra: per ogni relazione primitiva ricorsiva $S(x_1, \ldots, x_l)$ esiste un polinomio a coefficienti interi F con $m + n + l$ variabili tale che, per ogni $k_1, \ldots, k_l \in \mathbf{N}$:

$$S(k_1, \ldots, k_l) \Leftrightarrow \bigwedge x_1, \ldots, x_m \bigvee y_1, \ldots, y_n F(k_1, \ldots, k_l) = 0 \tag{1.4}$$

[114] *Ibid.*, p. 166.

[115] Gödel si riferisce presumbilmente alla versione del risultato contenuta in [Kle36].

Sostanzialmente, la tecnica di prova è la stessa che aveva dato a Gödel la possibilità di dimostrare risultati simili in precedenza[116], e dei quali, quindi, il lemma (1.4) può essere visto come un raffinamento.

La dimostrazione del lemma viene ricostruita da Gödel come segue.

Data la relazione primitiva ricorsiva $S(x_1, \ldots, x_l)$, consideriamone la funzione caratteristica f_S anch'essa ricorsiva primitiva e, dunque, computabile (nel senso della definizione Gödel-Herbrand). Sia dunque $E_1, \ldots, E_m$ il sistema di equazioni in $\phi_1, \ldots, \phi_k$ tale che, per ogni l-pla $\mathbf{n} \in \mathbf{N}^{l}$ [117]:

$$f_S(\mathbf{n}) = u \Leftrightarrow E_1, \ldots, E_m \vdash \overline{u} = \phi_k(\overline{\mathbf{n}}) \tag{1.5}$$

Dalla definizione segue che, per ogni possibile scelta dell'argomento, solo un numero finito di valori noti di ognuna delle ϕ_j è necessario per calcolare il valore della f_S.

Questa osservazione equivale a considerare le ϕ_j definite solo in un certo dominio (quello nel quale cadono gli argomenti in corrispondenza dei quali le funzioni assumono i valori che occorrono), e ciò fa sì, osserva Gödel, che (1.5) possa essere equivalentemente espressa come:

$$f_S(\mathbf{n}) = u \quad \Leftrightarrow \quad \bigvee \phi_1, \ldots, \phi_k \bigvee x_1, \ldots, x_k \bigwedge \mathbf{y}$$
$$\left.\begin{array}{c} [\mathbf{y} < x_1 \rightarrow E_1 \wedge \\ \mathbf{y} < x_2 \rightarrow E_2 \wedge \\ \wedge \ldots \wedge \\ \overline{\mathbf{n}} < x_k \ \wedge \ \overline{u} = \phi_k(\overline{\mathbf{n}}) \wedge \\ B_1 \wedge \ldots \wedge B_r] \end{array}\right\} W(\phi_j, x_i, \mathbf{y}, \overline{\mathbf{n}})$$

dove le condizioni aggiuntive $B_1, \ldots, B_r$, risultano essere necessarie nel caso in cui, tra le ϕ_j, vi sia qualche funzione, ad esempio ϕ_s, che sia definita in termini di qualche altra, sia essa, per limitarsi al caso più semplice, la sola ϕ_p.

Se così fosse, si avrebbe allora, per qualche $1 \leq q \leq r$:

$$B_q \equiv \mathbf{y} < x_p \rightarrow \phi_s(\mathbf{y}) < a_s$$

[116]Intendiamo riferirci ai risultati citati in [Göd31b] (Teor. VII, pp. 182-184) e [Göd65a] (§8, pp. 363-367) relativi alla reinterpretazione come formule aritmetiche degli enunciati formalmente indecidibili.

[117]Nel seguito si assumerà che le lettere $m, n, \ldots, u, v, \ldots$ varino su elementi di $\mathbf{N}$, e le corrispettive in grassetto $\mathbf{m}, \mathbf{n}, \ldots, \mathbf{u}, \mathbf{v}, \ldots$ su successioni (di lunghezza appropriata) di numeri naturali. Stessa convenzione, vale per le variabili e per i numerali.

che permette di esprimere come la conoscenza di ϕ_p nel suo dominio, sia una condizione sufficiente per calcolare la ϕ_s nel proprio.

Fissato allora un certo argomento $\mathbf{n}$ della f_S, siano indicati con $a_1, \ldots, a_{r_j}$ le successioni dei valori delle ϕ_j in questione (cioè le funzioni occorrenti nel calcolo di $f_S(\mathbf{n})$), per ogni $1 \leq j \leq k$.

Mediante la tecnica già adottata da Gödel nei lavori precedenti, si dimostra, usando il Teorema Cinese del Resto, che è possibile codificare ognuna di queste successioni mediante la funzione $\beta(x, y, z)$ che calcola il valore del resto della divisione di x per $1 + (1 + z)y$[118].

Si dimostra cioè che esistono $c_j, d_j \in \mathbf{N}$ tali che, per ogni $1 \leq i_j \leq r_j$ e $1 \leq j \leq k$:

$$a_{i_j} = \beta(c_j, d_j, i_j)$$

Segue dunque da (1.5) che:

$$f_S(\mathbf{n}) = u \Leftrightarrow \bigvee \mathbf{c}_j, \mathbf{d}_j, \mathbf{x}_j \bigwedge \mathbf{y} W(\overline{\mathbf{n}}, \mathbf{c}_j, \mathbf{d}_j, \mathbf{x}_j, \mathbf{y}) \tag{1.6}$$

dove $W(\overline{\mathbf{n}}, \mathbf{c}_j, \mathbf{d}_j, \mathbf{x}_j, \mathbf{y})$ è una relazione costruita mediante le operazioni booleane a partire da uguaglianze e disuguaglianze contenenti le applicazioni della funzione β (eventualmente 'annidate', nel caso in cui qualcuna delle ϕ_j sia definita in termini di qualche altra funzione di quella stessa successione).

A questo punto si ricava immediatamente che vale:

$$f_S(\mathbf{n}) \neq u \Leftrightarrow \bigwedge \mathbf{c}_j, \mathbf{d}_j, \mathbf{x}_j \bigvee \mathbf{y} \overline{W}(\overline{\mathbf{n}}, \mathbf{c}_j, \mathbf{d}_j, \mathbf{x}_j, \mathbf{y})$$

da cui segue[119] l'equivalenza di $S(\mathbf{n})$ con un'espressione che possiede intanto un prefisso della forma desiderata.

Per ottenere il risultato voluto, basta quindi osservare che la matrice può essere trasformata spostando il complemento lungo W secondo le opportune leggi per le operazioni booleane ed eliminando le conseguenti disuguaglianze, disequazioni e occorrenze 'annidate' della β mediante le seguenti equivalenze:

$$\begin{aligned} x \neq 0 \quad &\Leftrightarrow \quad \bigvee y(x = y + 1) \\ x \leq y \quad &\Leftrightarrow \quad \bigvee z(y = x + z) \\ x = \beta(c, d, \beta(c', d', i)) \quad &\Leftrightarrow \quad \bigvee y(x = \beta(c, d, y) \wedge y = \beta(c', d', i)) \end{aligned}$$

[118]Per una dimostrazione di questo fatto si veda, ad esempio, [Göd65a], p. 365.
[119]Ovviamente per il fatto che $S(\mathbf{n}) \Leftrightarrow f_S(\mathbf{n}) \neq 1$.

L'equivalenza di (1.4) può essere quindi ottenuta dall'espressione risultante, utilizzando le seguenti corrispondenze:

$$\begin{aligned} A = 0 \wedge B = 0 \quad &\Leftrightarrow \quad A \times B = 0 \\ A = 0 \vee B = 0 \quad &\Leftrightarrow \quad A^2 + B^2 = 0 \end{aligned}$$

e riducendo infine alla forma prenessa.

Come si diceva, dato (1.4), il teorema 1 enunciato da Gödel segue come un semplice corollario, dal momento che quel risultato implica l'esistenza di un opportuno polinomio G tale che:

$$\overline{T}_1(x, x, z) \Leftrightarrow \bigwedge x_1, \ldots, x_m \bigvee y_1, \ldots, y_n G(x, z) = 0$$

che permette di stabilire una corrispondenza tra i problemi appartenenti al complemento della classe B su indicata e gli elementi della classe A a cui si riferisce l'enunciato del teorema 1, e di far quindi seguire l'indecidibilità di quest'ultima dalla non ricorsività di B.

Per quel che riguarda poi il teorema 2, esso segue dall'osservazione banale che se esistesse un sistema formale corretto e sintatticamente completo rispetto alla classe A, avremmo anche, contrariamente a quanto appena dimostrato, una procedura meccanica di decisione per essa (ossia la macchina di Turing che enumera l'insieme dei teoremi del sistema in questione).

Quel che risulta essere rilevante rispetto all'argomentazione proposta da Gödel, che come si è fatto notare utilizza metodi di prova e risultati già acquisiti, è piuttosto il fatto seguente: i teoremi di incompletezza seguono come corollario dall'indecidibilità algoritmica della matematica o, più precisamente, della teoria delle equazioni diofantee, un fenomeno al quale Gödel, come abbiamo visto, si riferisce ricorrendo alla suggestiva formula secondo la quale il matematico, in tale ambito, "non può essere sostituito da una macchina". È poi il carattere peculiare degli enunciati formalmente indecidibili, i quali cioè divengono dimostrabili o refutabili nel sistema di tipo immediatamente superiore a quello considerato, a spingere Gödel a ritenere che la convinzione di Hilbert circa la resolubilità di ogni questione matematica non venga intaccata da questa conclusione.

L'osservazione che chiude questo lavoro inedito, torna però sulla questione:

> Non lascerei tuttavia passare sotto silenzio il fatto che esistono apparentemente questioni con una struttura molto simile che sono con molta probabilità indecidibili nel primo dei due sensi che ho esposto precedentemente. La

> differenza nella struttura di questi problemi sta soltanto nel fatto che in questi polinomi compaiono anche variabili per i numeri reali. Questioni connesse con l'ipotesi del continuo di Cantor portano a problemi di questo tipo. Non sono ancora riuscito a dimostrarne l'indecidibilità, ma ci sono considerazioni che rendono altamente plausibile il fatto che lo siano effettivamente.[120]

La nota si chiarisce parzialmente se confrontata con alcuni passi di tenore analogo contenuti nei contributi inediti dello stesso periodo e dedicati alla dimostrazone di consistenza dell'Ipotesi del Continuo e dell'Assioma di Scelta con gli assiomi della teoria degli insiemi ZF.

In una conferenza tenuta a Göttingen nel 1938, ad esempio, Gödel sottolinea come questa dimostrazione di consistenza "non sia affetta" dall'estensione del formalismo rispetto al quale essa viene data, per mezzo di nuovi assiomi evidenti e si può dimostrare, anzi, che essa "continua a valere immutata" anche per il formalismo esteso, "così ché essa vale in senso assoluto, per quanto possa avere senso, oggi, parlare di consistenza assoluta"[121].

Poco dopo, quasi alla fine dell'intervento, si legge ancora:

> La consistenza della proposizione A (che ogni insieme è costruibile) ha anche un interesse intrinseco, specialmente perché è estremamente plausibile che con A si abbia a che fare con una proposizione assolutamente indecidibile, rispetto alla quale la teoria degli insiemi si biforca in due sistemi differenti, simili alla geometria Euclidea e a quella non-Euclidea.[122]

Come nel caso dell'inedito analizzato in precedenza, Gödel conclude sottolineando come egli non sia ancora in possesso della dimostrazione di consistenza anche per la negazione di A (e quindi dell'indipendenza della medesima), ma si dice convinto di questo fatto.

Del tutto simile alle precedenti, è l'osservazione contenuta nella conferenza alla Brown University dell'anno successivo, con la quale Gödel intende evidenziare come la dimostrazione di consistenza dell'assioma di costruibilità e dell'Ipotesi del Continuo sia assoluta, ossia "indipendente dal particolare formalismo scelto per la matematica"[123]: ciò risulta dal fatto che essa "funziona per sistemi di tipo arbitrario", e 'trascende' quindi il "modo più generale" per completare, in relazione ai teoremi di indecidibilità, ogni formalismo.

---

[120] *Ibid.*, p. 175.
[121] [Göd38b], p. 129.
[122] *Ibid.*, p. 155.
[123] G1940a, p. 184.

Egli prosegue successivamente, accennando, come nel passo di [Göd38c], alla struttura di queste proposizioni indecidibili in senso assoluto e riassumendo in una suggestiva disgiunzione la situazione derivante dal non essere ancora riuscito a fornire una prova completa di quest'indecidibilità:

> Si ha poi che A è equivalente a proposizioni della forma seguente: $(P)[F(x_1, \ldots, x_k, n_1, \ldots, n_l) = 0]$ dove F è un polinomio con coefficienti interi dati e due sorta di variabili $x_i, n_{[j]}$ dove le x_i sono variabili per numeri reali e le $n_{[j]}$ variabili per interi, e dove P è un prefisso, ossia una successione di quantificatori composti da queste variabili x_i e $n_{[j]}$. Non sono ancora riuscito a dimostrare che A, e quindi che questa proposizione relativa al polinomio in questione, è effettivamente indecidibile, ma quello che si può dimostrare dato il risultato da me presentato in questa conferenza è naturalmente questo: O questa proposizione è assolutamente indecidibile, oppure [esiste un sistema formale ottenuto da ZF per aggiunta di nuovi, evidenti assiomi nel quale A è dimostrabile e] l'ipotesi del continuo di Cantor è [anch'essa] dimostrabile (dal momento che A implica l'ipotesi del continuo).[124]

Da quanto si dice nei passi qui citati si deve desumere quanto segue: che alla fine degli anni '30 Gödel era convinto dell'indecidibilità dell'assioma di costruibilità e dell'Ipotesi del Continuo; che egli fosse in possesso di una dimostrazione dell'indipendenza quantomeno della prima proposizione rispetto ad un sistema arbitrario nella gerarchia dei formalismi tipati[125]; che egli ritenesse questo fatto un indizio decisivo della natura assoluta dell'indecidibilità della proposizione in questione, in contrasto con quella solo relativa degli enunciati formalmente indecidibili della memoria del 1931, cioè un indizio, vale la pena sottolinearlo, del fatto che né per la proposizione in questione, né per la sua negazione può esistere una prova, *nel senso informale del termine*.

[124] *Ibid.*, p. 185.

[125] Una testimonianza diretta su questo punto in particolare, potrebbe celarsi dietro il brano tratto dalla corrispondenza con Church risalente alla metà degli anni '60 e motivato dalla richiesta di Church, al quale era stato affidato il compito di tenere un breve discorso in occasione della consegna della medaglia Fields a Paul Cohen nell'agosto del 1966, di un chiarimento circa certe dicerie sui risultati non pubblicati da Gödel sull'Ipotesi del Continuo (in special modo delle voci giuntegli circa l'esistenza di una dimostrazione di Gödel di indipendenza, precedente a quella di Cohen). Nell'ultima lettera, datata 29 settembre 1966, Gödel suggerisce a Church ([Göd03a], p. 373) di inserire un passaggio che si riferisce al fatto di come egli fosse in possesso di una dimostrazione della consistenza dell'assioma di costruibilità dalla teoria dei tipi che egli riteneva potesse essere estesa ad una dimostrazione di indipendenza dell'assioma di scelta dallo stesso formalismo. Egli aggiunge anche di aver deciso di non pubblicare niente al riguardo perché, tra l'altro, riteneva il risultato "di scarso interesse filosofico"[126].

È pur vero che alla luce delle affermazioni contenute nei lavori successivi a quelli fin qui considerati e dedicati in particolare all'attività gödeliana nel campo della teoria assiomatica degli insiemi, sembra plausibile ritenere che la posizione di Gödel così articolata non sarà destinata a reggere alla prova del tempo: come si vedrà più avanti[127], è lecito pensare che tale mutamento d'opinione possa essere avvenuto in relazione alla riflessione da parte di Gödel intorno ai cossidetti 'assiomi dell'infinito'.

Tuttavia, è rilevante che il decennio durante il quale l'opera di profonda chiarificazione concettuale intorno alla nozione di 'funzione computabile', permise infine di enunciare i teoremi di incompletezza in tutta la loro generalità (che, a detta di Gödel, si traduce in una limitazione fondamentale delle 'capacità matematiche' di una macchina astratta), si chiuda con una riflessione sull'esistenza di questioni indecidibili in modo assoluto per ogni matematico (umano). Un fatto che è significativo, sia, come si vedrà più approfonditamente nel capitolo seguente, perché ciò costituisce un passaggio importante per lo sviluppo del pensiero di Gödel, sia, più in generale, per il tipo di speculazione a cui i teoremi di incompletezza darranno adito (alcune delle quali saranno al centro della seconda parte del presente lavoro).

[127]Cfr. §2.3.

Capitolo 2

I teoremi di incompletezza e le loro implicazioni

Nel ricostruire il percorso storico e concettuale che ha condotto alla formulazione dei teoremi di incompletezza nella loro massima generalità e ad un'analisi, da parte di Gödel, delle ricadute filosofiche e fondazionali del risultato, si sono incontrate una serie di questioni rilevanti che, a causa del taglio che si è voluto dare al capitolo, non sono state di fatto approfondite.

È quello che ci si propone invece di fare adesso, sempre sulla base degli scritti di Gödel, isolando in modo particolare quelle che permettono di fare luce sulla genesi del risultato e delineino meglio la posizione di Gödel relativamente alla valenza della propria scoperta.

2.1 I principi euristici della scoperta: Gödel e Tarski

Tra i documenti che attestano uno sguardo retrospettivo di Gödel al quadro storico e culturale nel quale egli si trovò ad effettuare le proprie scoperte, certamente le due lettere spedite ad Hao Wang il 7 dicembre del 1967 e il 7 marzo 1968, che sono solo una piccola testimoniananza della stretta relazione che avrebbe legato i due personaggi nell'ultima fase della vita di Gödel e che sono divenute ormai familiari dal momento della loro parziale pubblicazione in [Wan74], meritano un posto d'onore per almeno due motivi.

In primo luogo, perché come nel caso del passo dell'intervento alla conferenza di Königsberg con il quale si è voluto aprire il capitolo precedente, ma persino in modo più significativo, si tratta di documenti nei quali Gödel cerca di fornire una ricostruzione d'insieme, che comprenda cioè alcuni dei risultati più importanti da lui ottenuti (nel caso specifico si tratta, oltre che del teorema di completezza semantica e dei risultati di incompletezza, anche della prova di consistenza dell'Ipotesi del Continuo e dell'assioma di scelta).

In secondo luogo, perché l'operazione viene condotta all'insegna della volontà di stabilire un preciso 'sodalizio euristico' tra convinzioni filosofico-fondazionali e ricerca matematica, un fatto che è probabilmente uno degli elementi alla base dell'interesse che continua a suscitare l'opera gödeliana nel suo complesso.

Nello specifico, le lettere in questione dovevano essere una risposta alla richiesta, da parte di Wang, di rivedere e correggere una prima bozza di introduzione ad una raccolta di scritti del logico norvegese Thoralf Skolem[128]. L'occasione servì a Gödel soprattutto per chiarire ulteriormente[129] la relazione tra la propria dimostrazione del teorema di completezza semantica del calcolo predicativo classico del prim'ordine, e certi lavori di Skolem, nei quali, apparentemente, tutte le nozioni ed i risultati necessari allo scopo, erano già contenuti.

La prima parte del passo seguente, dunque, nella quale Gödel individua ciò che a suo dire costituì il vero motivo alla base della mancata individuazione della soluzione alla questione della completezza semantica nei lavori di Skolem, è dunque riferita in particolare a questo risultato.

Tuttavia, è lo stesso Gödel poco dopo (così come documentato nella citazione, dal prosieguo della lettera), ad individuare nella medesima ragione l'elemento di novità contenuto nella dimostrazione dell'esistenza di proposizioni formalmente indecidibili:

> Questa cecità (o pregiudizio, o comunque lo si voglia chiamare) dei logici [nei confronti delle conseguenze dei lavori di Skolem] è davvero sorprendente. Ma credo che la spiegazione non sia difficile da trovare. Giace nella mancanza diffusa, a quel tempo, della necessaria attitudine epistemologica verso la metamatematica e verso il ragionamento non-finitario.
>
> Il ragionamento non finitario in matematica era infatti generalmente considerato dotato di senso solo fino a quando fosse possibile 'interpretarlo' o 'giustificarlo' in termini di una metamatematica finitaria. Questo punto di vista, conduce quasi inevitabilmente all'esclusione del ragionamento non finitario dalla metamatematica. [...] Inoltre, ammettere elementi transfiniti,

[128] [Sko70].

[129] Allo stesso argomento sono dedicati alcuni documenti che fanno parte della corrispondenza di Gödel con J. van Heijenoort (cfr. [Göd03b], pp. 307-325).

‘privi di senso’ nella metamatematica, risultava incoerente con la stessa idea di scienza prevalente all’epoca. Perché, secondo quest’idea, la metamatematica era *la* porzione significativa della matematica, attraverso la quale i simboli matematici (di per sé privi di senso) acquisiscono un sostituto del significato, ossia certe regole d’uso. [...]

> Vorrei aggiungere che la mia concezione oggettivista della matematica e della metamatematica in generale, e del ragionamento transfinito in particolare, è stata fondamentale anche per gli altri miei lavori in logica.
>
> Come si poteva pensare, altrimenti, di *esprimere* la metamatematica *negli* stessi sistemi matematici, se questi ultimi sono considerati come costituiti di simboli privi di senso che acquisiscono un sostituto di significato solo *attraverso* la metamatematica?[130]

In parte, l’interesse del luogo citato risiede, come già accennato poc’anzi, nel fatto di essere una preziosa fonte di informazioni sull’approccio filosofico di Gödel (e su certi aspetti della psicologia del personaggio, convinto di operare in un ambiente culturalmente ‘ostile’), tanto da far passare in secondo piano ogni valutazione di merito circa la veridicità del quadro delineato, sia per quel che riguarda il contesto culturale, sia per ciò che concerne l’effettiva integrità della propria posizione.

Ma questo non è l’unico elemento significativo che sia riconducibile all’argomento principale del presente lavoro. Qualche informazione in più, infatti, la si ricava da quanto Gödel dice in un passo di poco successivo a quello della citazione precedente, e che recita:

> Si dovrebbe infine notare che il principio euristico per la mia costruzione degli enunciati aritmetici indecidibili nei sistemi formali per la matematica, risiede nel concetto altamente non costruttivo di ‘verità matematica oggettiva’, *in opposizione* a quello di ‘dimostrabilità’, con il quale veniva generalmente confuso prima del mio lavoro e di quello di Tarski.[131]

Sullo stesso punto, Gödel torna anche nella lettera a Wang del 7 marzo 1968, laddove egli, nel ribadire l’elemento alla base dell’ ‘attitudine espistemologica’ che di fatto impediva, a coloro i quali condividessero l’impostazione di fondo della ricerca scientifica prevalente all’epoca, di fare le sue stesse scoperte, sottolinea come “i formalisti consideravano la dimostrabilità formale come un’*analisi* del concetto di verità matematica e quindi non erano naturalmente nella posizione di *distinguere* le due”[132].

[130] [Göd03b], p. 397-398.
[131] *Ibid.*, p. 398.
[132] *Ibid.*, p. 404.

Esistono già nella letteratura alcuni luoghi dai quali si possono trarre informazioni per gettare luce sulle affermazioni di Gödel[133]. Il punto cruciale è racchiuso nella cronaca storica della genesi della scoperta gödeliana secondo la quale, concluso il proprio *Habilitationschrifft*[134], Gödel intendeva in realtà tentare inizialmente, ispirato dallo studio degli scritti del matematico tedesco, di risolvere il problema di Hilbert riguardo all'esistenza di una dimostrazione di consistenza per l'analisi mediante metodi di prova finitari. Egli si trovò così a cercare di dimostrare la consistenza dell'analisi nel sistema per l'aritmetica.

Gödel si accorse però che la tecnica a cui egli aveva deciso di ricorrere, identificare le variabili di ordine superiore (per i numeri reali o, equivalentemente, per gli insiemi) con formule aritmetiche e verificare l'assioma di comprensione mediante il predicato di verità per queste formule, portava, nel formalismo scelto, a contraddizioni del tutto analoghe ai paradossi del linguaggio naturale come quello di Richard o 'del mentitore'[135]. Da qui, egli dovette trarre l'indicazione fondamentale, il "principio euristico", che lo avrebbe condotto poi alla dimostrazione dei teoremi di incompletezza.

Sebbene la storia sia nota, crediamo non sia vano, sia per una questione di merito che, più metodologicamente, di corretta indicazione delle fonti, ricostruirla nello spirito che abbiamo fin qui deciso di adottare, seguendo cioè le tracce ed i riferimenti ad essa che emergono dal materiale riconducibile allo stesso Gödel.

[133] Ad esempio, [Wan81] e [Daw96].

[134] Cioè [Göd29].

[135] Seguendo l'interpretazione di Feferman ([Fef84]), si può spiegare quanto affermato nel testo come segue: si consideri un'enumerazione delle formule aritmetiche contenenti una variabile libera e si assuma che le variabili insiemistiche varino su $\mathbf{N}$. Si ha allora che "$x \in y$" viene aritmeticamente interpretato con "$\phi_x(y)$ è vera". Ciò conduce quindi alla necessità di definire il predicato di verità per gli enunciati aritmetici e alla possibilità di formalizzare i paradossi del linguaggio naturale. Ad un livello storico, il ruolo non secondario che Gödel riserva alla problematica rappresentata dai paradossi, potrebbe far nascere il quesito: come era egli venuto a conoscenza della questione? La risposta più immediata e naturale, sarebbe quella di indicare la fonte primaria negli scritti di Russell che Gödel indubbiamente conosceva. Non è improbabile tuttavia che, data l'importanza via via acquisita dalla disamina delle antinomie, essa fosse divenuta patrimonio comune anche di testi matematici di carattere manualistico.

2.1.1 Il ruolo della verità nei teoremi di incompletezza

A conferma di quanto appena ricordato e a ulteriore chiarimento di quanto Gödel dice nelle due lettere a Wang succitate, è possibile indicare due passi estrapolati dai documenti sempre facenti parte il materiale della corrispondenza gödeliana, e che appartengono agli stessi anni dello scambio epistolare con Wang.

Il primo in ordine di tempo, fa parte della copia ritrovata tra le carte del *Nachlass* di una lettera spedita a Jean van Heijenoort il 22 febbraio 1964 e si inserisce nel quadro dello scambio epistolare tra i due dedicato alla revisione per approvazione della traduzione di [Göd31b], che sarebbe poi stata pubblicata in [vH67]. Il passaggio a cui intendiamo riferirci, è il poscritto al corpo della lettera che recita[136]:

> Forse Lei è stato confuso dal fatto che una volta ho affermato che il tentativo di dare una dimostrazione di consistenza relativa dell'analisi condusse alla dimostrazione dell'esistenza di proposizioni indecidibili, e in un altra occasione [ho detto] che il principio euristico e la prima versione della dimostrazione furono quelli indicati nel par. 7 delle mie conferenze di Princeton del 1934. Tuttavia fu proprio la prova di consistenza relativa che rese necessario formalizzare sia "vero" che "dimostrabile" e condusse forzatamente ad un confronto da questo punto di vista tra le due [nozioni].[137]

Il secondo passo è tratto invece dalla bozza per una lettera di risposta, mai inviata, a Yossef Balas, studente presso la University of Northern Iowa, che era in cerca di informazioni relative alla genesi dei teoremi di incompletezza[138]. Il passaggio scelto, seppur con qualche dovizia in più nei dettagli tecnici, conferma tutte le indicazioni provenienti dalla citazione precedente:

> Ho spiegato il principio euristico per la costruzione di enunciati indecidibili in un sistema formale dato, nelle conferenze che ho tenuto a Princeton nel 1934. [...] L'occasione per confrontare la verità e la dimostrabilità venne dal tentativo di dare una dimostrazione di consistenza model-teoretica e relativa dell'analisi nell'aritmetica. Ciò conduce quasi per necessità ad

[136] Come informa la nota editoriale al passo della citazione seguente ([Göd03b], p. 313), c'è su questo punto una discrepanza tra la copia della lettera rinvenuta tra le carte dell'Institute for Advanced Studies di Princeton, e quella poi effettivamente spedita. È dunque probabile che l'osservazione di Gödel non sia mai arrivata a van Heijenoort.

[137] *Ibid.*, p. 313.

[138] La lettera, ritrovata in evidente forma preparatoria, non è datata. La lettera di Balas, inviata all'American Mathematical Society e da questa girata a Gödel, è del 27 maggio 1970. La risposta di Gödel deve essere stata approntata non molto più tardi.

> un simile confronto. Perché, un modello aritmetico per l'analisi non è altro che una $\in$-relazione aritmetica che soddisfi l'assioma di comprensione: $\exists n(x)[x \in n \equiv \phi(x)]$. Ora, se qui "$\phi(x)$" viene rimpiazzato da "$\phi(x)$ è dimostrabile", una simile $\in$-relazione è facilmente definibile. Quindi, se la verità fosse equivalente alla dimostrabilità, si sarebbe raggiunto lo scopo. Tuttavia (e questo è il punto decisivo) segue dalla corretta soluzione ai paradossi semantici che la "verità" delle proposizioni di un linguaggio *non può essere espressa* nel linguaggio stesso, mentre la dimostrabilità (essendo una relazione aritmetica) *può* [*esserlo*].[139]

Come nella ricostruzione sommaria della vicenda da parte nostra di poco fa, Gödel conferma che fu proprio la metodologia della prova di consistenza da lui adottata, a rendere chiaro il fatto fondamentale: mentre ogni tentativo di esprimere il predicato di verità per le formule di un linguaggio formale è destinato all'insuccesso a causa della contraddizione basilare che, come è noto, dipende dalla tecnica cantoriana della diagonalizzazione, lo stesso non si dà nel caso del predicato di dimostrabilità formale, proprio perché, con questa seconda nozione, si ha a che fare con una relazione aritmetica e quindi aritmeticamente definibile.

Il riferimento a [Göd65a] che accumuna i due passi citati, ci riporta al paragrafo che, nel testo delle conferenze tenute da Gödel a Princeton, ha come titolo "Relazione del precedente argomento con i paradossi". Il paragrafo, contiene infatti una riflessione che connette l'argomento di esistenza di proposizioni formalmente indecidibili al problema delle antinomie, sulla base dello spunto offerto dal fatto che le proposizioni in questione sono in realtà (logicamente equivalenti a) enunciati costruiti a partire dalle proprietà metateoriche del sistema formale e che, dunque, 'dicono qualcosa' sul sistema formale e su esse stesse.

La loro costruzione è lo spunto per una critica della soluzione ai paradossi adottata da Whitehead e Russell (e permette quindi a noi di riconciliare l'affermazione di Gödel nel passo della lettera a Balas, su quella che lui considerava la soluzione "corretta" al problema[140]): il criterio che una proposizione non possa 'affermare alcunché' su se stessa è giudicato da Gödel "troppo drastico". Infatti, un risultato come il Lemma di Diagonalizzazione[141] esclude

[139][Göd03a], pp. 9-10.

[140]Si dovrebbe comunque tenere presente la visione articolata che Gödel aveva del problema delle antinomie (si veda, al proposito, [Wan96], in particolare pp. 269-273).

[141]Il riferimento da parte di Gödel a questo risultato, la nota al passo che attribuisce la paternità del medesimo a Carnap (l'indicazione è [Car34], p. 91), le tracce di una discussione tra Gödel e Carnap al proposito presenti nella corrispondenza tra i due (si

che si possano considerare privi di senso enunciati che, in relazione ad ogni proprietà metateorica del sistema dato ed esprimibile nel suo linguaggio, "dicono di se stessi di possedere quella proprietà". Tuttavia, nota Gödel, "questa costruzione è valida solo se la proprietà [in questione] può essere espressa nel sistema", quindi "la soluzione del paradosso di Epimenide", ad esempio, "sta nel fatto che questo non vale per tutte le proprietà metamatematiche"[142].

Quanto segue, che permette di "stabilire questo fatto in modo più formale", è presentato come "l'argomento euristico dell'esistenza di proposizioni indecidibili", ed altro non è che l'argomento 'non costruttivo' che rappresenta un modo alternativo per guardare alla dimostrazione gödeliana, che abbiamo incontrato nell'esaminare la corrispondenza con Zermelo nel §1.1: considerato che la classe A dei gödeliani delle formule vere di un linguaggio non può essere espressa mediante un predicato appartenente al linguaggio stesso (come dimostra, osserva Gödel, la medesima procedura di diagonalizzazione all'origine dell'enunciato indecidibile applicata però alla formula $\neg T(sost(x, y))$, dove T è l'ipotetico predicato di verità e $sost$, la rappresentazione formale della funzione di sostituzione); considerato che, al contrario, risulta esprimibile formalmente l'insieme B dei codici dei teoremi, si ha che $A \neq B$. Dunque, se si assume la correttezza del sistema, $B \subseteq A$, si ha $B \subset A$: esiste almeno una formula vera α non dimostrabile; $\neg\alpha$ è perciò falsa, e, per l'assunzione di correttezza, anch'essa non dimostrabile; dunque, α è indecidibile.

La ricostruzione di Gödel intende di fatto spingere a vedere i teoremi di incompletezza come un risultato che, al pari dei lavori di Tarski dedicati all'indagine sistematica intorno al problema della verità, ha condotto ad un mutamento significativo nel modo di guardare alla questione, ed in particolare all'abbandono dell'illusione che, con la precisazione del concetto di 'dimostrabilità in un sistema formale', si fosse in possesso di uno strumento adeguato per la sua soluzione.

È noto peraltro come l'osservazione di Gödel sia importante anche per un altro verso.

Tra le affermazioni gödeliane più frequenti relative al risultato di incompletezza, vi è quella che tende a sottolineare come gli enunciati formalmente indecidibili in un dato sistema formale, divengono decidibili formalmente in

veda in particolare, [Göd03a], p. 349), sono di particolare interesse storico tenendo conto del fatto che il lemma fa parte ormai integrante di ogni presentazione standard dei teoremi di incompletezza, ma che invece non compare nella dimostrazione originaria data da Gödel.

[142] [Göd65a], p. 363.

un sistema del tipo immediatamente superiore, anche se, rispetto a quest'ultimo formalismo, emerge la questione dell'indecidibilità di nuove proposizioni, dello stesso 'tipo' delle precedenti. In molti dei lavori e delle testimonianze documentali fin qui presi in considerazione, si trovano riferimenti di questo genere.

È il caso infatti, per citare qualche esempio, dell'accenno contenuto in [Göd38c], da noi già messo in evidenza, circa l'esistenza di inferenze "altrettanto evidenti di quelle date", che permettono di decidere formalmente gli enunciati indecidibili in un formalismo di partenza. Oppure, è un caso ancor più chiaro quello di [Göd36], in cui il risultato principale dell'articolo, la possibilità, in un certo sistema S_i della gerarchia dei sistemi tipati ivi introdotta da Gödel, di abbreviare la lunghezza delle dimostrazioni dei teoremi dei sistemi precedenti, viene presentato come un fatto che completa, in un certo senso, quanto segue dai teoremi di indecidibilità: l'esistenza in S_i di teoremi *nuovi*, rispetto a quelli dei sistemi precedenti nella gerarchia[143].

Il luogo più celebre, anche perché costituisce il primo caso di un'affermazione simile, è però probabilmente la nota 48a di [Göd31b], in cui la "vera ragione dell'incompletezza" è indicata "nel fatto che la formazione di tipi superiori può essere sempre proseguita nel transfinito [...] mentre in ogni sistema formale sono disponibili al massimo un'infinità numerabile di essi", e dove si riconduce ciò alla decidibilità della proposizione formalmente indecidibile "qualora vengano aggiunti i tipi superiori appropriati"[144]. Di questo, dice Gödel in quella circostanza, avrebbe dovuto trattare la parte II di [Göd31b].

Dalla corrispondenza con Carnap, in particolare dal passo della lettera dell'11 settembre 1932 che abbiamo già riportato nel §1.1, risulta che Gödel era intenzionato a dare, in quella seconda parte, una definizione della 'verità' per un linguaggio formale[145].

C'è forse un legame con quanto egli afferma nella nota 48a?

Come si accennava in precedenza, è un fatto ormai noto che sussiste una stretta relazione tra le due affermazioni di Gödel. Questo legame può essere facilmente esemplificato per il sistema formale classico per l'aritmetica PA.

[143] La costruzione di una gerarchia di sistemi formali analoga a quella di [Göd36], ed un'osservazione simile sulla decidibilità delle proposizioni formalmente indecidibili, eccettuato ovviamente il riferimento al teorema sullo *speed up* delle teorie formali, si trova anche in [Göd32].

[144] [Göd31b], p. 181.

[145] [Göd03a], p. 347.

In rapporto a questo sistema, il sistema di tipo superiore che ci permette di verificare l'affermazione della nota 48a, è un'opportuna formulazione del sistema PA_2 per l'aritmetica al second'ordine, ossia della teoria che, dal punto di vista del linguaggio, possiede una nuova sorta di variabili per insiemi e la relazione di appartenenza ($\in$), e che contiene poi, oltre agli assiomi ordinari per lo zero ed il successore (gli stessi di PA) e lo schema di induzione esteso a tutte le formule del linguaggio del secondo ordine, anche una qualche formulazione dello schema di comprensione. Allo scopo di decidere gli enunciati formalmente indecidibili in PA, è infatti sufficiente ammettere una forma ristretta di questo schema, e cioè un principio del tipo:

$$\exists X \forall x [x \in X \Leftrightarrow \alpha(x, \vec{y}, \vec{Y})] \quad (\Phi\text{−CA})$$

dove Φ è una classe di formule del linguaggio di PA_2 e α è una formula di Φ le cui variabili libere sono (al massimo) quelle dell'insieme:

$$\left\{x, \vec{y} := y_1, \ldots, y_n, \vec{Y} := Y_1, \ldots, Y_m\right\}$$

Indicheremo nel seguito, come è uso fare, con (Φ−CA) la versione di PA_2 in cui sia ammessa la restrizione dello schema di comprensione alla classe di formule Φ.

Il modo più naturale di procedere è allora quello di definire in PA_2 il predicato di verità per gli enunciati (formule chiuse) di PA. Ciò può essere fatto introducendo innanzi tutto la formula $T(X, n)$, che esprime il fatto che X è un predicato di verità (l'insieme delle formule vere) per formule la cui complessità logica (il numero di simboli logici) sia minore o uguale a n.

Il passo cruciale, consiste nell'osservazione che è possibile definire in PA_2 ricorsivamente la funzione Int, che associa ad ogni termine chiuso di PA la sua interpretazione in $\mathbf{N}$, mediante le condizioni seguenti:

$$\forall x [Int(\dot{\overline{x}}) = x]$$

$$\forall x_1, \ldots, x_n [Int(f(\dot{\overline{x}}_1, \ldots, \dot{\overline{x}}_n)) = f(Int(\dot{\overline{x}}_1), \ldots, Int(\dot{\overline{x}}_n))]$$

dove $\dot{\overline{x}}$ indica il termine del linguaggio che sta per l'$x+1$-esimo numerale[146].

A questo punto, facendo uso di alcune delle funzioni e dei predicati primitivi ricorsivi derivanti da un'aritmetizzazione standard del linguaggio di PA,

[146] Cioè, $\dot{\overline{x}}$ è il termine ottenuto mediante l'espressione che rappresenta nel linguaggio di PA la funzione aritmetica che, ad ogni $n \in \mathbf{N}$, associa il gödeliano del numerale $\overline{n}$.

in particolare $Cl(a)$ (che associa ad a la sua complessità logica), $En(a)$ (il predicato 'a è un enunciato'), $Tc(a)$ ('a è un termine chiuso'), $sost(a,x,t)$ (il risultato della sostituzione in a di t a x), si ha che la formula $T(X,n)$, assumendo che il linguaggio di PA sia ristretto al frammento $\{\neg,\wedge,\forall\}$ e indicando con $\lceil a \rceil$ il gödeliano in PA_2 dell'espressione a, si definisce come:

$$\begin{aligned}
T(X,n) \quad := \quad & \forall x \forall y [Tc(x) \wedge Tc(y) \rightarrow \\
& (\lceil x = y \rceil \in X \leftrightarrow Int(x) = Int(y))] \wedge \\
& \forall a [En(\lceil \neg a \rceil) \wedge Cl(\lceil \neg a \rceil) \leq n \rightarrow \\
& (\lceil \neg a \rceil \in X \leftrightarrow \neg a \in X)] \wedge \\
& \forall a \forall b [En(\lceil a \wedge b \rceil) \wedge Cl(\lceil a \wedge b \rceil) \leq n \rightarrow \\
& (\lceil a \wedge b \rceil \in X \leftrightarrow a \in X \wedge b \in X)] \wedge \\
& \forall x [En(\lceil \forall v(a) \rceil) \wedge Cl(\lceil \forall v(a) \rceil) \leq n \rightarrow \\
& (\lceil \forall v(a) \rceil \in X \leftrightarrow \forall y (sost(\lceil a \rceil, \lceil v \rceil, \dot{\bar{y}}) \in X))]
\end{aligned}$$

Si verifica quindi che: $(\Delta^1_1\text{–CA}) \vdash \forall x \exists ! X T(X,x)$.

Il predicato di verità $T(x)$ per gli enunciati del linguaggio di PA può essere a questo punto definito nei due modi equivalenti:

$$\begin{aligned}
T(a) \quad &:= \quad En(a) \wedge \exists x \exists X [Cl(a) \leq x \wedge T(X,x) \wedge a \in X] \\
T(a) \quad &:= \quad En(a) \wedge \forall x \forall X [Cl(a) \leq x \wedge T(X,x) \rightarrow a \in X]
\end{aligned}$$

Dallo schema di comprensione, segue quindi che:

$$(\Delta^1_1\text{–CA}) \vdash \exists X \forall x [x \in X \leftrightarrow T(x)]$$

Il passo successivo consiste nel verificare in $(\Delta^1_1\text{–CA})$ la verità di ogni enunciato dimostrabile in PA, ossia che vale:

$$(\Delta^1_1\text{–CA}) \vdash \forall x [En(x) \wedge Teor_{PA}(x) \rightarrow T(x)]$$

da cui, considerando che per le definizioni date si dimostra $\neg T(\lceil 0 = 1 \rceil)$, si ricava:

$$(\Delta^1_1\text{–CA}) \vdash Con_{PA} := \neg Teor_{PA}(\lceil 0 = 1 \rceil)$$

L'argomento qui schizzato è chiaramente generalizzabile ad ogni sistema formale S che verifichi le ipotesi dei teoremi di incompletezza. Si è così ad un tempo mostrato in che senso il passaggio al tipo immediatamente superiore rende decidibili enunciati dimostrabilmente indecidibili nel sistema di partenza, e che quest'osservazione passa in modo naturale dalla definizione di un predicato di verità per le formule chiuse del linguaggio del sistema dato. Si è cioè stabilito quel legame a cui accennavamo tra le affermazioni di Gödel al proposito e dalle quali abbiamo preso le mosse[147].

Che un simile legame sussistesse, lo si poteva certo già inferire dal carteggio con Zermelo nel quale si ritrovano, nella risposta alla prima lettera da parte del matematico tedesco, il riferimento al predicato di verità per le formule di un linguaggio formale (come versione corretta dell'obiezione sollevata da Zermelo), l'argomento euristico, e cioè la dimostrazione non costruttiva dell'incompletezza (come conseguenza del fatto che la definizione della verità non può mai essere data nello stesso formalismo a cui si riferisce), e l'indicazione della decidibilità delle formule indecidibili in sistemi di tipo superiore (come conseguenza, invece, del fatto che la definizione della verità per il formalismo di partenza può essere data in questi sistemi).

Il luogo nel quale, tuttavia, questo legame emerge dagli scritti di Gödel nel modo più chiaro, fa parte dell'ingente epistolario con Paul Bernays.

2.1.2 Sul predicato di verità: la corrispondenza con Bernays

Il materiale a cui intendiamo fare riferimento possiede anche un interesse indipendente dal tipo di problematica che è all'attenzione di questa parte del presente lavoro. Lo scambio di vedute avviene infatti contestualmente alla preparazione per la stampa di [Göd31b]. Come è facile immaginarsi, Bernays era particolarmente interessato alle implicazioni tecniche e, soprattutto, alle conseguenze fondazionali della scoperta di Gödel. Le prime lettere che fanno parte dell'epistolario tra i due, quindi, sono quasi interamente occupate dalle riflessioni suscitate in Bernays dalla progressiva familiarizzazione con la metodologia di prova usata da Gödel. Particolarmente rilevanti, sono quelle osservazioni che intendono misurare gli effetti dei teoremi di incompletezza sul programma fondazionale di Hilbert. In tal senso, questo materiale verrà ripreso e discusso nel §2.2 che è dedicato al problema del finitismo.

[147]Solomon Feferman, che segnala questa lettura nel suo [Fef84], accredita Saul Kripke di avergli fatto osservare la possibilità di interpretare la nota 48a di [Göd31b] così come si è fatto nel testo.

Ciò non toglie che alcuni commenti di Gödel siano importanti per la questione che si sta esaminando.

Di particolare interesse è la lettera del 2 aprile 1931 che risponde ad una lunga missiva da parte di Bernays nella quale quest'ultimo discuteva gli effetti della prova di Gödel, alla luce di due estensioni del formalismo per la teoria dei numeri a cui di recente lui ed Hilbert avevano pensato.

Si tratta di due sistemi d'assiomi $\Im^*$ e $\Im^{**}$ ottenuti da una formalizzazione standard dell'aritmetica $\Im$, per aggiunta, rispettivamente, di una versione ristretta e di una versione generalizzata di ω-regola (una regole con un'infinità numerabile di premesse).

Più precisamente, il sistema $\Im^*$ è ottenuto da $\Im$ per aggiunta della regola:

$$\frac{\begin{array}{c}\vdots\\ A(\overline{n}_1, \ldots, \overline{n}_k)\end{array}}{\forall x_1, \ldots, x_k A(x_1, \ldots, x_k)}$$

dove $A(x_1, \ldots, x_k)$ è una formula primitiva ricorsiva, per ogni istanza della quale ($A(\overline{n}_1, \ldots, \overline{n}_k)$, per ogni $n_1, \ldots, n_k \in \mathbf{N}$) si possiede una *dimostrazione finitaria*[148].

$\Im^{**}$ è invece ottenuto da $\Im$ mediante una regola analoga, ma priva della restrizione della precedente: $A(x_1, \ldots, x_k)$ è cioè una formula arbitraria del linguaggio.

Piuttosto che l'argomento in sé, mediante il quale Gödel mostra a Bernays come nemmeno $\Im^*$ e $\Im^{**}$ siano sintatticamente completi, e che, considerata la data della lettera di Gödel, anticipa di gran lunga un risultato che discende ufficialmente solo dall'estensione dei teoremi di incompletezza dovuta Rosser e pubblicata nel 1937[149], ci interessa per il momento l'affermazione di Gödel il quale, nel notare come il metodo su cui questa prova si basa "è lo stesso che conduce alla decisione delle proposizioni precedentemente indecidibili per mezzo di variabili di tipo superiore", dice:

> [L'argomento] consiste in quanto segue: Nel sistema S [ottenuto da $\Im$ per l'aggiunta di variabili insiemistiche e assiomi appropriati] si può definire una

[148] Il sistema formale $\Im^*$ è introdotto e discusso in [Hil31b], [Hil31a]. Il primo dei dua articoli fu recensito proprio da Gödel (la recensione è pubblicata come [Göd86b] in [Göd86a], pp. 212-215) il quale tuttavia omise, in quell'occasione, gran parte delle riserve critiche delle quali mette invece al corrente Bernays nello scambio epistolare in esame (la questione, come si è detto, verrà ripresa nel §2.2).

[149] [Ros36].

classe di numeri naturali che (in breve) contiene tutti i numeri assegnati alle formule corrette di [$\Im$] (in modo analogo [a quanto avviene] per due sistemi qualsiasi il secondo dei quali superi il primo nella formazione dei tipi).[150]

Sul predicato $W(x)$ che Gödel introduce al proposito, egli dice qualcosa subito dopo aver schizzato l'argomento, che segue la citazione precedente, mediante il quale si può mostrare per diagonalizzazione l'incompletezza sintattica di $\Im^*$ e $\Im^{**}$.

Più precisamente, egli nota:

> [I]l principio in accordo al quale la classe $W(x)$ viene definita è ricorsivo; definisco in primo luogo cosa W significhi in rapporto alle proprosizioni più semplici (equazioni numeriche, ecc.) e poi procedo con le proposizioni più complesse, cioè in accordo agli schemi seguenti:
>
> $$W[\mathrm{Neg}(a)] =_{\mathrm{Df}} \overline{W}(a)$$
> $$W(a \,\mathrm{Dis}\, b) =_{\mathrm{Df}} W(a) \vee W(b)$$
> $$W(v \,\mathrm{Gen}\, a) =_{\mathrm{Df}} (x) W Sb(a^{v}_{Z(x)})$$
>
> (dove, nell'ultima formula, a è un segno di classe con la variabile libera v).[151]

Notazione a parte (è chiaro che $\mathrm{Neg}(x)$, $x \,\mathrm{Dis}\, y$ e $x \,\mathrm{Gen}\, y$ sono le funzioni che, rispettivamente, associano al gödeliano di una formula quello della sua negazione, ai gödeliani di A e B quello di $A \vee B$, al gödeliano della variabile v e della formula $A(v)$ il gödeliano di $\forall v A(v)$, e $Sb(x^{y}_{z})$ e $Z(x)$ sono i corrispettivi delle funzioni da noi indicate con $sost(x, y, z)$ e $\dot{\overline{x}}$), si sarà notata l'analogia con le definizioni date in precedenza.

Gödel torna quindi, alla luce di ciò, sulla questione della proposizioni formalmente indecidibili:

> Rispetto alla decidibilità nei sistemi superiori delle proposizioni indecidibili, ciò segue immediatamente dalle proprietà del concetto $W(x)$ [(]che può essere sempre definito nel sistema superiore in relazione a quello più debole[)]. Infatti, [da quanto detto in precedenza] segue che la consistenza del sistema inferiore (in questo caso [$\Im^*$]) è formalmente dimostrabile in quello più esteso (cioè [$\Im^{**}$]).[152]

Non manca neanche, infine, il riferimento al lavoro di Tarski:

[150] [Göd03a], p. 94.
[151] *Ibid.*, p. 96.
[152] *Ibid.*, p. 96.

> Simultaneamente ed indipendentemente da me (come ho saputo da una conversazione), il Sig. Tarski ha sviluppato l'idea di definire il concetto di "proposizione vera" in questo modo (sicuramente per altri scopi).[153]

Il sostegno documentale, che deriva soprattutto dalla corrispondenza gödeliana, all'interpretazione comunemente offerta per spiegare certi luoghi degli scritti editi di Gödel in cui compaiono brevi riferimenti allusivi alla problematica della verità, è dunque completo. Da notare ad esempio che, come si è già accennato commentando la corrispondenza con Zermelo, sulla base dell'evidente consapevolezza di Gödel rispetto a questa questione, il celebre passo di [Göd31b] nel quale Gödel sottolinea come la proposizione indecidibile, che 'dicendo di se stessa di non essere dimostrabile' e risultando tale in quanto indecidibile, diviene decidibile, essendo vera, "mediante considerazioni metamatematiche"[154], passo che pare essere all'origine di molti fraintendimenti, acquisisce un senso preciso.

A questo punto, non resta che tirare le fila della disamina.

2.1.3 Una considerazione conclusiva (parziale)

Dunque, Gödel era in possesso del *know how* per la dimostrazione di quello che viene ricordato come il Teorema di Tarski: il predicato di verità per la teoria dei numeri non è aritmeticamente definibile, ma lo diviene, mediante la formalizzazione delle clausole della sua definizione induttiva, nel sistema di tipo immediatamente superiore (e questo vale, più in generale, per tutti i livelli della gerarchia dei tipi). Fatto ancora più rilevante, egli riteneva che ciò indicasse la "vera ragione dell'incompletezza".

Dal quadro che emerge dalle lettere a Wang citate all'inizio di questa parte, la possibilità di definire il predicato di verità per un linguaggio formale, che si riconnette in modo decisivo all'estensione di un formalismo mediante nozioni di ordine superiore, costituisce dunque una parte essenziale di quell' "attitudine epistemologica necessaria verso la metamatematica ed il ragionamento non-finitario", alla quale Gödel ascrive le ragioni del carattere profondamente innovativo di molte delle proprie scoperte.

Di tutto questo, l' "opposizione" tra verità e dimostrabilità formale non è altro che un esempio, il quale dischiude però un aspetto essenziale della concezione gödeliana della matematica.

[153] *Ibid.*, p. 96.
[154] [Göd31b], p. 150.

Essa rappresenta infatti solo uno dei modi attraverso il quale si svela un carattere costitutivo della disciplina: come il contributo di certi concetti astratti (nell'esempio qui considerato, la nozione di insieme, in ultima analisi) risulti necessario per la dimostrazione dei teoremi relativi ad aree della matematica (nel caso, l'aritmetica) per le quali esiste, in relazione alle nozioni in gioco, un'evidenza molto vicina a quella alla base del rapporto percettivo.

Altrove[155], questa è stata chiamata la "dottrina di Gödel".

Come vedremo meglio nei §§2.3 e 2.4 in modo particolare, questa convinzione, che in relazione al tema della verità si delinea soltanto sullo sfondo, emerge con maggiore chiarezza dagli scritti più tardi dedicati (solo apparentemente, dunque) ad altro genere di problematiche.

2.2 Riflessioni sul programma di Hilbert e il finitismo

Tra le conseguenze che vengono comunemente riconosciute ai teoremi di incompletezza, vi sono gli effetti, considerati devastanti, sulle aspettative riposte nel programma fondazionale di David Hilbert. In particolare, si ritiene che il risultato di Gödel, laddove esso indica l'indimostrabilità della proposizione che esprime la consistenza di un sistema d'assiomi nel sistema stesso, abbia inferto un colpo mortale all'idea secondo la quale proprio una dimostrazione della consistenza dei principi della matematica mediante *metodi finitari di prova* (ottenuta cioè con il solo concorso delle 'proposizioni reali', immediatamente evidenti), possa essere il criterio sufficiente a garantire la correttezza soprattutto di quei teoremi per ottenere i quali si sia fatto uso di 'proposizioni ideali', che si riferiscono cioè a concetti astratti (e che, per tale motivo, non sono proposizioni dotate di senso come quelle 'reali').

Questo punto di vista largamente condiviso, si sposa con il tono con il quale Constance Reid, autrice di alcune note biografie di personalità importanti del mondo matematico, scrisse a Gödel il primo settembre 1965, al fine di ottenere informazioni che potessero esserle utili per la biografia di Hilbert sulla quale stava all'epoca lavorando[156]:

> La conclusione drammatica della carriera di Hilbert è ovviamente la pubblicazione del Suo articolo del 1931 e gli effetti di questo sul suo programma

[155] Cfr. [Fef87].

[156] Il volume in questione, pubblicato successivamente al carteggio con Gödel, è [Rei70].

> Formalista. Nello scrivere dell'argomento, mi piacerebbe dire qualcosa anche sulle circostanze personali dell'evento.
>
> Mi chiedo se Lei abbia mai incontrato Hilbert personalmente o se abbia avuto una qualche discussione con lui sul Suo lavoro. Se ciò si fosse verificato, gradirei che mi inviasse una descrizione dell'incontro e delle Sue impressioni relative allo stesso Hilbert.[157]

La risposta di Gödel (la data è quella del 22 marzo 1966), oltre a confermare il fatto già noto che non ci fu mai alcun incontro né scambio epistolare con Hilbert, contiene alcuni interessanti commenti relativi alle conseguenze dei teoremi di incompletezza per il programma hilbertiano.

Il bilancio di Gödel finisce per essere equilibrato, considerando da un lato gli effetti ineludibili delle proprie scoperte, dall'altro gli aspetti concettuali ugualmente attuali e non privi di interesse del punto di vista 'finitista', e gli importanti sviluppi matematici seguiti alla revisione dell'impostazione originaria.

Il tenore prudente di Gödel, traspare già dalla formula a cui egli ricorre per descrivere le conseguenze negative dei teoremi di incompletezza:

> Ciò che è stato dimostrato è solo che non si può ottenere lo *specifico* obiettivo *epistemologico* che Hilbert aveva in mente. L'obiettivo in questione era dimostrare la consistenza degli assiomi della matematica classica sulla base della medesima evidenza concreta ed immediatamente convincente dell'aritmetica elementare.[158]

Questa formula, infatti, fa intuire l'esistenza di qualcosa d'altro, nella posizione del matematico tedesco, che mantiene, a dispetto di tale conclusione, una sua ragion d'essere.

Ci sono in particolare due aspetti che Gödel intende rimarcare. Il primo è legato a certi risultati che la ricerca ispirata alle considerazioni di Hilbert ha comunque permesso di acquisire:

> Guardando tuttavia alla situazione da un punto di vista puramente *matematico*, le dimostrazioni di consistenza sulla base di assunti metamatematici opportunamente scelti (così come sono state fornite da Gentzen e altri), sono altrettanto interessanti, e conducono ad intuizioni importanti riguardo alla struttura proof-teoretica della matematica.[159]

[157] [Göd03b], p. 186.
[158] *Ibid.*, p. 187.
[159] *Ibid.*, p. 187.

Oltre a ciò, sottolinea Gödel, vi sono questioni concettuali legate all'impostazione hilbertiana che, opportunamente riformulate, appaiono ancora meritevoli di un'indagine attenta:

> Rimane inoltre aperta la questione se, o fin dove, sia possibile, sulla base dell'approcio formalista, dimostrare la consistenza della matematica classica "costruttivamente", ossia rimpiazzare i suoi assiomi relativi ad entità astratte appartenenti ad un mondo Platonico oggettivo, mediante intuizioni relative ad operazioni date della nostra mente.[160]

Si sente, dietro le parole di Gödel, l'eco dei propri sforzi in questa direzione[161].

Il suggestivo quadro che questi commenti prefigurano, offre molti spunti che richiamano alcuni degli snodi fondamentali tramite i quali la ricerca di Gödel si è intrecciata alle problematiche derivanti dal programma di Hilbert, e che vanno oltre, sotto molti aspetti, una banalizzante contrapposizione in relazione alla scoperta dei teoremi di incompletezza.

Data l'ormai notevole quantità di materiale al proposito, non sarà vano dedicarci alla ricostruzione e all'analisi di quello che, come si diceva in apertura di paragrafo, rappresenta uno degli aspetti più comunemente associati ai teoremi di Gödel.

2.2.1 La prima fase: 1931-1933

Quanto si è appena detto, non intende affatto sminuire una problematica, quella del rapporto tra i teoremi di incompletezza ed il programma di Hilbert, che è certamente all'origine del grande clamore in campo fondazionale suscitato dalla scoperta gödeliana, e pertanto meritevole di un certo spazio.

Anzi, a ben vedere, proprio il tentativo di ricostruire il punto di vista di Gödel al proposito, presenta non pochi motivi di interesse. Sorprenderà taluni, ad esempio, la constatazione del fatto che, a tal fine, occorre distinguere una prima fase nella quale Gödel, pur criticando il valore attribuito alle dimostrazioni di consistenza secondo i dettami del programma finitista, era restio a considerare l'indecidibilità della consistenza come un fatto decisivo per una soluzione negativa delle aspettative hilbertiane, a dispetto di quanti, anche tra coloro che erano vicini alle idee di Hilbert, ritenevano il contrario.

---

[160] *Ibid.*, p. 187.

[161] Intendiamo riferirci ai due lavori dedicati all' 'interpretazione Dialectica', [Göd58] e [Göd90b].

Cronologicamente parlando, questa fase è legata alle osservazioni offerte immediatamente a seguito della scoperta del risultato e si protrae approssimativamente fino al 1933. Considerando infatti anche i documenti inediti, ci sono ragioni per ritenere, come vedremo, che intorno a quell'anno Gödel si fosse convinto più risolutamente degli effetti negativi del secondo teorema di incompletezza.

L'espressione più compiuta dei dubbi di Gödel al proposito, è probabilmente il seguente passo di [Göd31b] a commento del Teorema XI della memoria, ossia, appunto, del risultato relativo all'indecidibilità della consistenza in un sistema dato:

> L'intera dimostrazione del Teorema XI si trasferisce parola per parola al sistema di assiomi della teoria degli insiemi, M, e a quello della matematica classica, A, e anche in questi casi dà luogo al risultato: Non esiste una dimostrazione di consistenza per M o per A che possa essere formalizzata in M o A, rispettivamente, supposto che M, o A, sia consistente. Vorrei notare espressamente che il Teorema XI (e il risultato corrispondente per M e A) non contraddicono il punto di vista formalista di Hilbert. Perché questo punto di vista presuppone solo l'esistenza di una dimostrazione di consistenza nella quale siano usati nient'altro che metodi di prova finitari, e rimane concepibile che esistano dimostrazioni finitarie che non *possano* essere espresse nel formalismo di P (o M o A).[162]

Il punto di vista di Gödel è espresso in modo ancor più perentorio nel poscritto richiesto dagli editori del numero di *Erkenntnis* e contenente il resoconoto della tavola rotonda sui fondamenti della matematica svoltasi al termine del congresso di Königsberg del 5-7 settembre 1930, il quale era stato esplicitamente pensato per dare la possibilità a Gödel di esporre con maggior dovizia di dettagli il risultato di incompletezza frettolosamente annunciato, come si è in parte ricordato, durante il dibattito:

> [In conseguenza dell'indimostrabilità della consistenza] sarebbe dunque impossibile una dimostrazione di consistenza finitaria per un sistema nel quale siano formalizzati tutti i metodi di prova finitari (cioè intuizionisticamente ineccepibili). Tuttavia, sembra discutibile che qualcuno dei sistemi fin qui introdotti, ad esempio *Principia mathematica*, sia così omnicomprensivo (o addiruttura che un sistema siffatto esista).[163]

In questo caso il tono del passo sembra studiato per smorzare il tenore più apertamente 'anti-hilbertiano' della critica che Gödel aveva sollevato nel corso della discussione di Königsberg.

[162] [Göd31b], p. 195.
[163] [Göd31a], p. 205

Con il suo intervento, infatti, Gödel aveva voluto fare presente le proprie perplessità relativamente all'idea di guardare alla consistenza di un sistema d'assiomi, come ad una proprietà in grado di garantire che l'aggiunta di proposizioni "transfinite" ('ideali', secondo la terminologia hilbertiana) ad una teoria assiomatica accettata, non porti alla dimostrazione di proposizioni "dotate di senso", cioè immediatamente intutive (o 'reali', nell'accezione di Hilbert), che risultino essere false:

> Vorrei adesso puntualizzare che non è possibile ritenere questi due requisiti come equivalenti senza alcun indugio. Infatti, se in un sistema formale consistente A (ad esempio quello della matematica classica) si dimostra una certa proposizione p dotata di senso con l'aiuto di assiomi transfiniti, dalla consistenza di A segue soltanto che non-p non è formalmente dimostrabile *entro A*. Nonostante ciò, rimane possibile che si possa accertare non-p mediante considerazioni di tipo contenutistico (intuizionistico) che *non* siano formalizzabili in A.[164]

L'osservazione di Gödel poggia dunque sullo stesso fondamento della precedente: dal momento che non si ha alcuna certezza sul fatto che esista un sistema formale in grado di 'contenere' tutti i metodi di prova contenutisticamente, ovvero intuizionisticamente ammissibili, in presenza di una dimostrazione di una proposizione, ad esempio, della forma $\exists x F(x)$, in cui F esprime una proprietà combinatoria dei numeri naturali, non si potrebbe di fatto concludere dalla semplice consistenza, la sua verità. Essa trova un sostegno ulteriore nella distinzione, all'epoca inevitabilmente inedita, tra la 'consistenza' di una teoria assiomatica e l' 'ω-consistenza' della medesima:

> (Assumendo la consistenza della matematica classica) si può persino fornire esempi di proposizioni [della forma $\forall x F(x)$, dove F è una formula aritmetica] che, sebbene contenutisticamente vere, sono indimostrabili nel sistema formale della matematica classica. Aggiungendo allora la negazione di tali proposizioni agli assiomi della matematica classica, si ottiene un sistema consistente in cui è dimostrabile un enunciato contenutisticamente falso.[165]

Nonostante i dubbi su una possibile applicazione *diretta* dei teoremi di incompletezza alla prospettiva hilbertiana, Gödel era dunque certamente critico nei confronti della posizione di Hilbert relativa alle dimostrazioni finitarie di consistenza.

[164] *Ibid.*, p. 200.
[165] *Ibid.*, p. 202.

Già nel paragrafo introduttivo alla dimostrazione del teorema di completezza semantica[166], egli aveva affrontato la questione prendendo in esame la posizione, non difficilmente riconducibile ad Hilbert, sebbene, nel testo, Gödel non lo faccia in modo esplicito, che guarda alla consistenza degli assiomi come ad un requisito sufficiente per l'esistenza di un modello, ed aveva già prefigurato le difficoltà per essa derivanti da un'ipotetica dimostrazione di incompletezza sintattica (e della conseguente esistenza di due modelli non isomorfi)[167].

Tuttavia, Gödel era anche genuinamente convinto delle proprie perplessità relative all'assunzione che tutti i metodi di prova finitari fossero formalizzabili in un *unico* sistema formale. Non sembra plausibile ritenere cioè che vi sia la sua proverbiale cautela, la volontà di mantenersi il più possibile estraneo a qualsiasi diatriba intellettuale, dietro quelle affermazioni che sembrano offrire una 'via d'uscita' al programma finitista dalle conseguenze dei teoremi di incompletezza.

D'altra parte di riferimenti alla questione ve ne sono anche nell'epistolario di Gödel relativo a quegli anni, segno di una convinzione persistente. Il fatto curioso casomai, consiste nell'osservare come, nonostante i suoi interlocutori rispondano al nome di Herbrand, Bernays, von Neumann, e rappresentino quindi quanto di più vicino ad Hilbert si poteva immaginare nel panorama della ricerca logico-matematica del tempo, saranno proprio questi ultimi a rilevare, a fronte dello scetticismo di Gödel, certe implicazioni della scoperta gödeliana.

Tra i primi a contattare Gödel dopo l'annuncio del teorema ci fu von Neumann, come era peraltro prevedibile essendo egli stato tra i partecipanti alla succitata tavola rotonda di Königsberg.

Già con la lettera del 29 novembre 1930 di quest'ultimo a Gödel, le avvisaglie del dissenso tra i due, sia riguardo alla possibilità di formalizzare i metodi finitari di prova ("Ritengo che ogni considerazione intuizionista possa essere copiata formalmente"[168]), sia, di conseguenza, sul significato dei teoremi di incompletezza ("Perciò penso che il Suo risultato abbia risolto

[166] Cfr. [Göd29], pp 60-65.

[167] Occorre notare tuttavia, che le considerazioni fondazionali contenute in quel paragrafo (Gödel contrapponeva al punto di vista ricordato nel testo, la critica brouweriana delle dimostrazioni di consistenza), furono escluse dall'articolo nel quale il risultato di completezza semantica per il calcolo dei predicati classico venne ufficialmente presentato (cioè [Göd30a]).

[168] [Göd03b], p. 338.

negativamente la problematica fondazionale: non esiste una giustificazione rigorosa per la matematica classica"[169]), appaiono evidenti[170].

È però nella lettera successiva (datata 10,12 gennaio 1931) che la diversità di vedute tra i due emerge con maggiore chiarezza dalle parole di von Neumann:

> Mi trovo in totale disaccordo con il Suo punto di vista riguardo la formalizzabilità dell'intuizionismo. Certamente, per ogni sistema formale ce ne è un'altro, come Lei ha dimostrato, che è [...] più forte. Ma l'intuizionismo non è assolutamente affetto da ciò. [...]
>
> Non posso certo dimostrare che ogni *costruzione dell'aritmetica* intuizionisticamente corretta sia formalizzabile in A [il sistema per l'aritmetica classica], M [il sistema dell'analisi] o in Z [la teoria degli insiemi]- perché l'intuizionismo è indefinito e indefinibile. Ma non è forse un fatto che non sia nota neanche una sola costruzione del tipo menzionato che non possa essere formalizzata in A, e che nessun logico vivente sia nella posizione di indicarne una? Oppure sbaglio, e Lei conosce una costruzione aritmetica effettiva di tipo intuizionista la cui formalizzazione in A crea difficoltà? Se questo fosse il caso per la mia più grande sorpresa, allora la formalizzazione suddetta dovrebbe certamente funzionare rispetto a M o Z![171]

Simili considerazioni, il pieno convincimento della formalizzabilità dei 'metodi finitari di prova' da un lato, rafforzato dall'assenza e dall'implausibilità di un controesempio, ma anche la consapevolezza delle difficoltà derivanti dalla mancanza di una definizione matematicamente soddisfacente della nozione chiave, sono alla base della posizione di Herbrand con il quale Gödel si trovò ad intrattenere un breve ma significativo carteggio di poco più tardo.

Il seguente brano, tratto da una lettera del matematico francese datata 7 aprile 1931, ricalca quasi parola per parola quanto sostenuto da von Neumann nel passo precedente:

> Non capisco affatto come sia possibile che esistano dimostrazioni intuizioniste[172] che non sono formalizzabili nel sistema di Russell [...]. Solo un esempio potrebbe convincermi di ciò [...].
>
> [Il fatto che non sia possibile definire esattamente in cosa consistono i metodi intuizionisti di prova] [r]afforza la mia convinzione che sia impossibile

[169] *Ibid.*, p. 338.

[170] Il contesto sembra rendere inequivocabile che il termine 'intuizionismo' venga considerato da von Neumann equivalente a 'finitismo', come si è implicitamente assunto nel testo. Simili considerazioni valgono d'altra parte anche per lo stesso Gödel (nel caso di citazioni precedenti) e anche per il passo di Herbrand al quale ci si riferirà in seguito.

[171] *Ibid.*, pp. 340-342.

[172] *NdT*: Si veda quanto detto alla nota 170.

> dimostrare che ogni prova intuizionista è formalizzabile nel sistema di Russell, ma che un controesempio a ciò non verrà mai trovato. Saremo forse obbligati ad adottare una forma di postulato logico.[173]

Evidentemente, però, l'assenza di un'analisi soddisfacente del concetto di 'dimostrazione finitaria' era per Gödel una fonte di perplessità maggiore, di quanto l'assenza e l'implausibilità di un controesempio non fosse un motivo di convincimento, per i suoi interlocutori, della ragionevolezza del 'postulato logico' proposto da Herbrand.

Anche il carteggio con Bernays non si discosta in modo decisivo da questo genere di considerazioni.

Tuttavia, occorre notare che la posizione del braccio destro di Hilbert differisce da quella di von Neumann ed Herbrand, dimostrando egli una maggiore condivisione dei dubbi di Gödel[174], per quanto fosse apparentemente propenso a formulare in modo comunque più netto di quanto non facesse Gödel le conseguenze dei teoremi di incompletezza[175].

Piuttosto, due osservazioni di Gödel relativamente a certe considerazioni di Bernays, ma solo marginalmente connesse con il significato dei teoremi di incompletezza per il programma di Hilbert, meritano menzione.

La prima è stimolata dall'introduzione da parte di Bernays delle estensioni $\mathfrak{I}^*$ e $\mathfrak{I}^{**}$ del formalismo standard per l'aritmetica, a cui si è fatto riferimento nel paragrafo sulle relazioni con il lavoro di Tarski[176]. A tale proposito Gödel sottolinea, nella risposta del 2 aprile 1931, come il fatto che essi facciano riferimento essenziale, nella loro stessa definizione (e in particolare, in relazione alla nuova regola di inferenza), al concetto di 'prova finitaria', "senza che questa sia stata resa in modo matematicamente preciso"[177], sia

[173] *Ibid.*, pp. 18-20.

[174] Discutendo il fatto ovvio che la posizione di von Neumann riguardo alla formalizzabilità di ogni dimostrazione di tipo finitario nel sistema P di [Göd31b], implica necessariamente la conclusione che non può esistere una dimostrazione di consistenza finitaria per P, Bernays commenta: "come Voi ritengo che ciò non sia affatto risolto" ([Göd03a], p. 86).

[175] Il paragrafo immediatamente successivo a quello a cui appartiene la citazione della nota precedente recita: "Anche senza l'aggiunta della congettura di von Neumann, per ogni sistema formale che soddisfi la Vostra assunzione 1. [che sia, cioè, ricorsivamente assiomatizzabile], vale l'alternativa: O il sistema è inadeguato per quanto riguarda la sua forza dimostrativa, o la sua consistenza non è dimostrabile in modo finitario" (*Ibid.*, p. 86).

[176] Cfr. §2.1.2.

[177] *Ibid.*, p. 96.

all'origine dell'insoddisfazione che, a detta di Gödel, questi sistemi suscitano come candidati per una fondazione ragionevole della teoria dei numeri. Il fatto acquisisce la propria rilevanza tenendo presente che il lavoro di Hilbert nel quale il sistema $\Im^*$ viene introdotto fu recensito proprio da Gödel[178], il quale mancò tuttavia, in quell'occasione, di fare un simile rilievo critico alla proposta hilbertiana.

L'altra nota, riguarda sempre la discussione dei sistemi $\Im^*$ e $\Im^{**}$, ma è relativa alla descrizione da parte di Gödel dell'argomento mediante il quale è possibile dimostrare l'incompletezza sintattica anche di queste estensioni dell'aritmetica classica. Infatti, proprio il fatto che la regola che li caratterizza chiama in causa nella premessa proposizioni le cui istanze numeriche siano state dimostrate con metodi finitari, costringe Gödel, al fine di rendere i sistemi d'assiomi in questione passibili di un'indagine metamatematica, ad offrire una propria, "altamente plausibile" definizione di cosa significhi per una formula essere 'finitariamente dimostrabile'. È interessante notare come la proposta gödeliana ricorra a quella che indicheremmo oggi come la risoluzione skolemiana di una formula in forma prenessa, per dare conto dell'interpretazione costruttiva della quantificazione esistenziale: una formula del tipo $(Q)A(x_1, \ldots, x_n)$, dove $A(x_1, \ldots, x_n)$ è una matrice primitiva ricorsiva e (Q) un prefisso di quantificatori è 'finitariamente dimostrabile' se, per ogni variabile x_i quantificata esistenzialmente in (Q) esiste una "funzione aritmetica finitariamente costruita" f_i, che abbia per argomenti le variabili quantificate universalmente in (Q) che precedono x_i, e tale che per la formula ottenuta sostituendo f_i a x_i, esista una dimostrazione finitaria[179].

Tornando all'argomento principale del paragrafo, valga, ai fini di un resoconto complessivo della posizione gödeliana in merito alla valenza antifinitista dei teoremi di incompletezza così come essa emerge dalla corrispondenza a cui si è fatto riferimento, il seguente passo, tratto dalla risposta del 25 luglio del 1931 alla succitata lettera di Herbrand del 7 aprile dello stesso anno:

> Ovviamente, non intendo affermare neanche che vi siano con certezza delle dimostrazioni finitarie non formalizzabili nei *Principia mathematica*,

[178] La recensione compare in [Göd86a] (pp. 212-215) come [Göd86b].

[179] Cfr. *Ibid.*, p. 93. L'interesse dell'osservazione di Gödel, consiste anche nel fatto che essa utilizza la procedura di risoluzione skolemiana di una formula in forma prenessa in relazione ad una questione, la caratterizzazione della concezione costruttiva del quantificatore esistenziale, che non ha niente a che fare con i propositi originari alla base dei risultati di Skolem.

> anche se intuitivamente tendo a questa assunzione. In ogni caso, una dimostrazione finitaria non formalizzabile nei *Principia mathematica* dovrebbe essere straordinariamente complicata, e su questo piano puramente pratico ci sono ben poche prospettive di scoprirla; ma è mia opinione che ciò non cambi nulla riguardo alla possibilità di principio.[180]

Si diceva comunque, che le perplessità di Gödel riguardo alla possibilità di formalizzare i 'metodi finitari di prova' entro un'unica cornice formale, rappresentano solo una prima fase della posizione di Gödel al proposito e che queste possono essere considerate apparentemente superate a partire dal 1933.

Alla fine di quell'anno, infatti, Gödel tenne un intervento dal titolo "L'attuale situazione dei fondamenti della matematica", all'incontro congiunto della Mathematical Association of America e dell'American Mathematical Society svoltosi dal 29 al 30 dicembre a Cambridge (Massachusetts). In quell'occasione Gödel, intendendo fornire come da titolo dell'intervento un quadro riassuntivo della ricerca fondazionale, individuò due questioni nelle quali poteva essere scisso il problema di giustificare la "totalità dei metodi di prova usati correntemente dai matematici". Si trattava cioè, (i) "di ridurre al minimo numero di assiomi e regole di inferenza primitive" i metodi di prova in questione, e, (ii), "trovare una giustificazione in un senso o in un altro per questi assiomi"[181].

Dal testo emerge come, se da un lato Gödel giudicava il primo problema essere stato risolto "in modo perfettamente soddisfacente" mediante la formalizzazione della matematica secondo l'indagine iniziata da Frege e Peano, egli volesse contemporaneamente sottolineare come, relativamente alla seconda questione, la situazione risultasse invece "estremamente insoddisfacente".

Dopo aver opportunamente motivato la propria conclusione in merito[182], Gödel introduce il problema della dimostrazione di consistenza di un sistema

[180] [Göd03b], p. 23.

[181] [Göd33], p. 45.

[182] Gödel individua in particolare tre questioni irrisolte, la nozione non costruttiva di esistenza, le definizioni impredicative e l'assioma di scelta, come conseguenza delle quali "i nostri assiomi, se interpretati come proposizioni dotate di senso, presuppongono necessariamente qualche forma di Platonismo, che non può soddisfare alcuna mente critica e che non produce neanche la convinzione che essi siano consistenti." (*Ibid.*, p. 50). Il passo in questione, divenuto ben presto un luogo celebre nella letteratura critica, sembra indicare, per quanto sia isolato in questo senso, un approccio problematico al realismo verso gli enti matematici (si veda al proposito anche la nota introduttiva a [Göd33] di Solomon Feferman).

d'assiomi, come una sorta di obiettivo secondario che discende dal fallimento conseguito nell'ottenere quello primario (trovare una giustificazione degli assiomi per la matematica):

> Sorge allora la congettura che, sebbene non si sia stati in grado di fornire un significato ineccepibile ai simboli dei nostri sistemi formali, si possa quantomeno essere in grado di dimostrare la loro non contraddittorietà mediante metodi ineccepibili di prova.[183]

A tal fine, Gödel introduce un sistema formale A il quale deve rispettare le due restrizioni di carattere generale seguenti:

1. è possibile quantificare universalmente o esistenzialmente solo su collezioni per le quali si dispone di un metodo effettivo di generazione degli elementi (sul modello dell'insieme dei numeri naturali)[184];

2. sono ammesse solo quelle proposizioni esistenziali per le quali si è in grado di produrre concretamente un esempio rispetto al quale esse siano soddisfatte (quindi il quantificatore esistenziale sussiste solo come abbreviazione).

Sulla base della prima restrizione, il metodo per dimostrare proposizioni universali è dunque il metodo induttivo, applicato alla procedura di generazione degli oggetti trattati. Si richiede inoltre che possano essere introdotte "solo quelle nozioni decidibili per ogni elemento" e "solo quelle funzioni calcolabili" per ogni possibile argomento[185].

Se il sistema A rappresenta quella cornice formale entro la quale offrire quella dimostrazione 'ineccepibile' di consistenza per gli assiomi della matematica (classica), questo stesso obbiettivo, sottolinea Gödel, non può essere raggiunto in conseguenza dei teoremi di incompletezza *e* della seguente, cruciale osservazione:

> Tutte le dimostrazioni intuizioniste ottemperanti i requisiti del sistema A che sono state finora costruite possono essere facilmente espresse nel sistema dell'analisi classica e anche in quello dell'aritmetica classica, *e ci sono ragioni*

[183] *Ibid.*, p. 50.

[184] Tale formulazione può apparire ambigua: non è chiaro se si intenda limitare la quantificazione ad oggetti appartenenti a collezioni decidibili o ricorsivamente enumerabili. Presumibilmente, si tratta del primo caso.

[185] Per quanto la definizione di Gödel sia data solo in modo informale, quanto detto richiama alla mente il sistema PRA, per l'aritmetica primitiva ricorsiva.

per ritenere che questo continuerà a valere per ogni prova che si sarà mai in grado di costruire.[186]

Sulla base di quest'affermazione, che incarna quel mutamento d'opinione di Gödel a cui si accennava in precedenza, si può leggere la frase di chiusura dell'intervento alla conferenza del 1933, secondo la quale "rimane la speranza che in futuro si possano scoprire altri e più soddisfacenti metodi di costruzione, che oltrepassino i limiti del sistema A, tali da poter giustificare su di essi l'aritmetica e l'analisi classiche"[187], tanto come un riferimento ad un superamento o revisione dell'originaria impostazione hilbertiana del problema della consistenza, quanto come il riconoscimento dell'interesse di una simile direzione d'indagine.

2.2.2 Il superamento: estendere il programma di Hilbert

Se è vero che almeno fin dal 1933 Gödel aveva superato le proprie iniziali perplessità, passando a considerare i teoremi di incompletezza quell'ostacolo alla realizzazione del programma di Hilbert che ad altri era parso immediatamente, e tenendo presente l'importanza che verrà attribuita all'evento, si potrebbe persino giungere a chiedersi il perché dell'assenza di una memoria più specificamente dedicata all'argomento.

Lo stesso Gödel d'altra parte si era dimostrato consapevole della plausibilità di una simile aspettativa, sottolinenando proprio nella lettera a Constance Reid la propria volontà, "in un'occasione opportuna", di pubblicare un qualche resonconto "sulla relazione tra il mio lavoro ed il programma di Hilbert"[188].

È difficile tuttavia scorgere qualcosa del genere nella prima delle tre "note sui risultati di indecidibilità", che costituiscono l'ultima pubblicazione di Gödel relativa ai suoi celebri teoremi[189]. Si tratta di poche righe, precedute dal titolo, significativo, "La migliore e la più generale versione dell'indimostrabilità della consistenza nel sistema stesso", le quali, con la data del 18 maggio 1966, erano apparse come nota aggiuntiva al testo, in occasione della pubblicazione in traduzione inglese di [Göd32] nella raccolta [vH67].

[186] *Ibid.*, p. 52 (l'enfasi è nostra).
[187] *Ibid.*, p. 53.
[188] [Göd03b], p. 187.
[189] [Göd90c].

Quelle righe, contengono una duplice osservazione sul secondo teorema di incompletezza. Da un lato, Gödel sottolinea come il requisito per un sistema formale S di 'contenere' il sistema Z per l'aritmetica affinché il risultato in questione si applichi ad esso, necessiti una formulazione precisa: sotto l'ipotesi infatti "della traducibilità effettiva di Z in S, in modo che la dimostrabilità sia preservata in questa direzione" (ossia dell'esistenza di una funzione ricorsiva che associ ad ogni formula α di Z una α^* in S in modo tale che $\mathsf{Z} \vdash \alpha \Rightarrow \mathsf{S} \vdash \alpha^*$), "la consistenza (nel senso di non-dimostrabilità di un enunciato e della sua negazione), anche di un sistema S molto forte, *potrebbe* essere dimostrabile in S e persino nel sistema per l'aritmetica ricorsiva primitiva"[190]. Il riferimento obbligato, è a quei lavori[191] che a partire dalla metà degli anni '50, misero in evidenza come fosse possibile di una data teoria formale, fornire una presentazione, cioè un'opportuna 'scelta' degli assiomi e delle regole di inferenza, che a parità di capacità deduttiva, permettesse la derivazione nel sistema di un'enunciato esprimente la non contraddittorietà degli assiomi.

È la seconda parte della nota, tuttavia, a giustificare il titolo scelto da Gödel per essa:

> Ciò che è possibile mostrare essere indimostrabile in S è il fatto che le regole del calcolo equazionale applicate ad equazioni dimostrabili in S tra termini primitivi ricorsivi, determinano solo equazioni numeriche corrette (assumendo che S possieda la proprietà che si asserisce essere indimostrabile).[192]

Nella versione "migliore e più generale", dunque, il secondo teorema di Gödel dice che: assumendo che S sia un sistema formale, il quale possieda la proprietà secondo la quale dati t_f, s_g termini che rappresentano in S due funzioni primitive ricorsive (diciamo, f e g), ogni enunciato della forma $t_f(\overline{n}) = s_g(\overline{m})$, ottenuto unicamente per applicazione delle regole (i) di sostituzione di un termine primitivo ricorsivo ad una variabile e (ii) di un termine con un altro dimostrabilmente uguale ad esso, è vero (cioè vale $f(n) = g(m)$), allora la proposizione che esprime questa stessa proprietà è indecidibile.

Formalmente parlando, ciò equivale a dire che esiste una formula α della forma $\forall x(t(x) = 0)$ (con t che rappresenta in S una funzione primitiva

[190] [Göd90c], p. 305.
[191] In particolare i già citati [Tak55], [Fef60], [Kre65a].
[192] *Ibid.*, p. 305.

ricorsiva), tale che l'enunciato:

$$Teor^*_S(\ulcorner\alpha\urcorner) \to \alpha$$

è indecidibile in S (dove $Teor^*_S(x)$ è il predicato ricorsivamente enumerabile 'x è un teorema equazionale di S', ossia è ottenuto in S per applicazione delle sole regole (i) e (ii) indicate in precedenza)[193].

"Si noti", dice Gödel, "che è necessario dimostrare questa consistenza 'esteriore' per 'giustificare' gli assiomi transfiniti di S nel senso del programma di Hilbert"[194]. La valenza anti-hilbertiana dell'indecidibilità di questa forma di consistenza viene dunque esplicitamente dichiarata[195].

Non pare essere su un piano squisitamente tecnico, comunque, che si ritrovano le ragioni profonde della distanza tra la visione di Gödel e l'approccio hilbertiano (d'altra parte occorre tenere presente che l'osservazione di Gödel nasce come nota al testo, e dunque non si può certo attribuirgli troppe velleità). E comunque non certo mediante un risultato tecnico del tipo indicato da Gödel, come risulta dalla presa d'atto delle linee di ricerca negli anni a cui appartiene la nota gödeliana: i lavori di Takeuti, Schütte, Tait da un lato e l'opera di Kreisel dall'altro, avevano già spostato l'accento sull'estensione al secondo ordine delle indagini di Gentzen e, dunque, avevano considerevolmente alzato il livello del dibattito sul finitismo e le sue possibili estensioni[196].

Gödel era peraltro ben al corrente dell'avanzamento della ricerca in questa direzione, come si evince anche da alcuni accenni presenti soprattutto nella corrispondenza con Bernays come il seguente, tratto da una lettera dell'11 agosto 1960:

[193]La prova di questa versione del secondo teorema di incompletezza, ricalca, nella strategia dimostrativa, la dimostrazione usuale del secondo teorema di Gödel (a parte per una differenza essenziale nella costruzione dell'enunciato indecidibile). Si rimanda alla nota introduttiva a [Göd90c] (in particolare, [Göd90a], pp. 284-284) per i dettagli al proposito.

[194][Göd90c], p. 305.

[195]Il fondamento dell'osservazione di Gödel, consiste nel fatto che Hilbert identificava le proposizioni 'reali' con proposizioni numeriche generali le cui istanze fossero verificabili in modo effettivo ("formule a cui corrispondono informazioni contenutistiche di proposizioni finitarie [...] equazioni e disequazioni numeriche, oppure informazioni più complesse composte da queste"- [Hil28], p. 470), della derivazione delle quali egli intendeva garantire l'ammissibilità mediante la consistenza "esteriore", che, come Hilbert sapeva (cfr. *Ibid.*, p. 474), equivale classicamente alla 'normale' consistenza di un sistema formale.

[196]Tra i riferimenti principali per la discussione del finitismo dei primi decenni del secondo dopoguerra, oltre a [Tak55] e al testo di Scütte [Sch60], possiamo citare anche [Kre65a], in particolare p. 151 e segg.

> Fino a questo momento, apparentemente Tait ha dimostrato la consistenza dell'analisi solo mediante metodi non finitari. Ho avuto interessanti discussioni con Kreisel. Egli sembra aver adesso dimostrato davvero in modo matematicamente soddisfacente che il primo ϵ-numero è un limite preciso di ciò che è finitario. Trovo che il risultato sia molto bello, anche se dovrà forse richiedere una sottostruttura di tipo fenomenologico perché possa essere completamente soddisfacente.[197]

Sfortunatamente nelle lettere successive l'argomento non trova uno spazio adeguato, probabilmente perché, ipotizza Fefereman nella nota introduttiva, la questione veniva dibattuta più esaustivamente nelle conversazioni tra i due.

Questi riferimenti sono comunque sufficienti a far emergere il punto cruciale: piuttosto che vertere su una riformulazione dei teoremi di incompletezza, il 'resoconto' dedicato alle relazioni tra il programma di Hilbert e l'opera di Gödel, avrebbe suscitato maggior interesse se incentrato sull'inadeguatezza "epistemologica" di quella posizione, della quale Gödel riferisce alla Reid, o anche sull'analisi "di tipo fenomenologico" che risulta necessaria per chiarire gli aspetti metamatematici della questione, di cui Gödel parla a Bernays nella citazione precedente.

Sono proprio i testi nei quali è l'aspetto concettuale ad essere privilegiato insomma, che risultano essere maggiormente indicativi. Ed è il materiale inedito, ancora una volta, a fornire gli spunti più interessanti.

Ciò vale, innanzi tutto, per uno dei testi più rappresentativi in questo senso, che riproduce l'intervento di Gödel del 29 gennaio 1938, al seminario organizzato dallo storico della scienza Edgar Zilsel.

L'argomento trattato, si riallaccia perfettamente al bilancio conclusivo sull'esperienza hilbertiana tracciato da Gödel nella lettera alla Reid, dal momento che, come appare evidente fin dalle prime battute, il materiale è imperniato intorno all'indagine sulla praticabilità matematica e sul valore epistemico di una ricerca fondazionale basata sulle prove di consistenza *relativa*, come riformulazione dell'impianto del programma di Hilbert basato originariamente sulla ricerca di dimostrazioni *assolute* di non contraddittorietà.

Il peso dei teoremi di incompletezza è ormai un dato acquisito e lo si intravede dietro le osservazioni di Gödel relative alle dimostrazioni di consistenza: la nota di apertura è dedicata a come simili prove mantengano un significato in quanto *riduzioni* di una teoria ad una sottoteoria, cioè di un

[197] [Göd03a], p. 192.

sistema d'assiomi ad un sistema che abbia (in quanto più debole) maggiore plausibilità.

Il nucleo centrale dell'intervento può essere visto come il tentativo di dare un seguito all'affermazione con la quale Gödel aveva chiuso la conferenza del 1933 sulle prospettive fondazionali del 'dopo Hilbert': definire una cornice formale che possa essere vista come una estensione plausibile dell'aritmetica finitaria, entro la quale si sia in grado di offrire una giustificazione, per mezzo di una dimostrazione di non contraddittorietà, degli assiomi della matematica.

A tal fine, la disamina di Gödel passa attraverso l'indicazioni di tre vie principali: l'introduzione di funzionali di tipo superiore, l'ammissione di principi di induzione fino ad opportuni ordinali transfiniti, l'interpretazione mediante operatori modali.

Due osservazioni sorgono spontanee: la prima consiste nel fatto che Gödel avesse già individuato all'epoca il principio alla base di quella che diverrà nota come l''Intepretazione Dialectica', ossia la traduzione dell'aritmetica di Peano in un sistema di funzionali di tipo finito basato su una logica intuizionista, che consente una dimostrazione 'combinatoria' della consistenza della teoria classica. È una conferma, questa, di quanto emerge più chiaramente dalla conferenza presso l'università di Yale del 1941[198]: l'ideazione di quello che sulla base degli scritti pubblicati rappresenta l'ultimo contributo tecnico di Gödel[199], appartiene in realtà alla fine degli anni '30.

In secondo luogo, si può apprezzare come Gödel avesse assimilato perfettamente la metodologia di prova usata da Gentzen per dimostrare la consistenza dell'aritmetica classica, al punto da alludere ad una possibile estensione a sistemi di ordine superiore (anticipando così quella questione che verrà sistematicamente indagata solo a partire dagli anni '50 e alla quale è dedicato il frammento precedente della lettera a Bernays)[200].

---

[198] Il testo dell'intervento è stato pubblicato nel terzo volume dei *Collected Works* come [Göd41].

[199] Il lavoro che ufficialmente contiene l'esposizione del risultato è del 1958.

[200] Per quel che riguarda la terza via, che risulta anche la meno chiara dal materiale relativo all'esposizione di Gödel da Zilsel, ci limitiamo ad indicare il breve accenno contenuto nella lettera a von Neumann datata 14 marzo 1933, che ha forse qualche attinenza con quanto viene detto in [Göd38d] sulla via 'logico-modale' (e che tratta apparentemente dell'interpretazione di teorie classiche mediante operatori di dimostrabilità). Dice Gödel: "Ho trovato un'interpretazione del calcolo di Heyting per la logica proposizionale per mezzo del concetto 'dimostrabile' ma fino ad ora non sono stato in grado di stabilire l'equivalenza piena con il sistema di Heyting" ([Göd03b], p. 346). Il problema indicato da Gödel è stato

È con il paragrafo conclusivo, nel quale le tre proposte vengono giudicate come possibili riformulazioni del programma finitista, in relazione alla rispettiva valenza fondazionale (nel senso di provvedere ad una giustificazione di certe parti della matematica) e all'interesse epistemologico (come riduzione concettuale delle nozioni astratte ad altre dotate di maggiore evidenza intrinseca), che il punto di vista di Gödel sull'intera 'esperienza' legata alla figura di Hilbert, cioè l'opinione riassunta nella lettera alla Reid, emerge in tutta la sua articolazione.

Punto primo: il punto di vista di hilbertiano ha un elevato significato filosofico e fondazionale:

> Se fosse stato possibile portare a termine il programma di Hilbert originario, esso sarebbe stato di enorme valore epistemologico [...]: (A) la matematica sarebbe stata ridotta ad una piccola porzione di se stessa (e quindi un gran numero di assunzioni indipendenti sarabbero divenute superflue). (B) Tutto sarebbe stato ricondotto a una base concreta, in grado di suscitare l'accordo di tutti.[201]

Punto secondo: un'estensione dell'approccio hilbertiano deve forzatamente rinunciare a simili aspirazioni:

> Relativamente alle dimostrazioni mediante il finitismo esteso, [l'obbiettivo (A)] non può essere più raggiunto, dal momento che per dimostrare una data teoria consistente, si ha sempre bisogno di aggiungere assunzioni al posto di quelle da dimostrare non contraddittorie, così che non si ha una riduzione [nel senso suddetto], ma piuttosto un rimpiazzamento o uno slittamento. Il secondo [scopo] (riduzione ad una base concreta, che si traduce in un accrescimento di evidenza) lo si ottiene in misura diversa per i diversi sistemi [considerati nell'intervento].[202]

Rimane il fatto, terzo, che "il significato matematico dell'indagine sia totalmente intatto", che Gödel consideri questo significato come "straordinariamente rilevante", e che egli ritenga che "i metodi applicati in questa direzione porteranno a risultati molto interessanti sia nel campo della ricerca fondazionale che al di fuori di esso"[203].

---

ripreso e sviluppato da un punto di vista tecnico da Sergei Artëmov con la cosiddetta 'Logica delle Dimostrazioni' (cfr. [Art94] e [Art01] in particolare), che si è rivelata come uno stimolante completamento dell'analisi modale offerta in termini della più nota 'Logica della Dimostrabilità'.

[201] [Göd38d], p. 112.

[202] *Ibid.*, p. 112

[203] *Ibid.*, p. 112.

Aleggiano sullo sfondo di questa disamina alcune questioni che diverranno patrimonio condiviso nel campo della metamatematica solo qualche decennio più tardi: la necessità di distinuere tra vari livelli di costruttivismo, il problema di determinare i limiti precisi (da un punto di vista della teoria della dimostrazione) della matematica finitaria, la valutazione, alla luce dei due punti precedenti, del contenuto costruttivo di quelle inferenze che garantiscono la non contraddittorietà di certi sistemi d'assiomi (cioè, in ultima analisi, la valenza epistemica di certe riduzioni proof-teoretiche). Non sarebbe difficile costruire, su questo, un'ulteriore chiave di lettura dell'opera gödeliana chiamando in causa materiale diverso da quello che si è deciso qui di analizzare (o leggendo il materiale che si è preso in considerazione nel presente lavoro sotto una luce diversa). Tuttavia, ciò ci condurrebbe lontano dallo scopo più modesto che ci si è prefissi: individuare alcuni aspetti cruciali del pensiero di Gödel partendo dall'analisi delle implicazioni dei teoremi di incompletezza.

Da questo punto di vista, in relazione alla tematica del significato dei teoremi di Gödel per il programma di Hilbert, ci si potrebbe ritenere soddisfatti dall'aver trovato con [Göd38d] un bilancio chiaro della ricerca intorno al programma sulle dimostrazioni di consistenza, che contiene, per di più, anche la valutazione di possibili sviluppi dell'indagine metamatematica.

Tuttavia, un elemento di interesse ulteriore lo si può ricavare seguendo 'il filo' dell'elaborazione da parte di Gödel del problema delle prove di consistenza, nella sua naturale prosecuzione rappresentata dai lavori dedicati all' 'Interpretazione Dialectica'[204]. Dall'esame delle pagine iniziali di quei contributi, scaturisce infatti la possibilità di inserire le riflessioni di Gödel sul finitismo nel quadro più ampio della visione gödeliana della matematica.

2.2.3 La matematica finitaria ed i concetti astratti

Quanto ci pare opportuno rilevare in questo contesto, è contenuto già nei primi paragrafi di [Göd58]. Seppur mascherato dall'attribuzione dell'osservazione cruciale a Bernays, la frase di apertura del lavoro ci riporta seduta stante al problema delle implicazioni dei teoremi di incompletezza per il programma di Hilbert:

[204] [Göd58] e [Göd90b]. Il fatto che si consideri questi lavori come la prosecuzione "naturale" del materiale fin qui analizzato, è ovviamente riferito alla presenza in [Göd38d], come precedentemente indicato, di un'anticipazione dei principi alla base del sistema dei funzionali di tipo finito.

> P. Bernays ha sottolineato in diverse occasioni che, dal momento che la consistenza di un sistema non può essere dimostrata mediante metodi di prova più deboli di quelli del sistema stesso, è necessario andare oltre il quadro di ciò che è, secondo Hilbert, la matematica finitaria se si vuole dimostrare la consistenza della matematica classica, o anche quella dell'aritmetica classica. Di conseguenza, dal momento che la matematica finitaria è definita come la matematica in cui l'evidenza poggia su ciò che è *intuitivo*, certe nozioni *astratte* sono necessarie per la consistenza dell'aritmetica (così come è stato esplicitamente formulato da Bernays nel suo [[Ber35]], pagine 62 e 69).[205]

Piuttosto che la paternità dell'affermazione, ci interessa qui lo sviluppo dell'analisi, che è tutto gödeliano.

Vale la pena di riportare per intero il passo rilevante, che recita:

> Qui, per nozioni astratte (o non intuitive) intendiamo quelle che sono essenzialmente del secondo ordine o di ordine superiore, cioè nozioni che non coinvolgono proprietà o relazioni di *oggetti concreti* (per esempio combinazioni di segni), ma che sono legate a *costrutti mentali* (ad esempio, prove, enunciati dotati di senso, e così via); e nelle dimostrazioni si fa uso di intuizioni, relative a questi costrutti mentali, che emergono non dalle proprietà combinatorie (spazio-temporali) delle combinazioni di simboli che rappresentano le dimostrazioni, ma solo dal loro *significato*.[206]

Il supermento della posizione hilbertiana avviene dunque su un duplice piano: per il fatto che le dimostrazioni di consistenza devono necessariamente chiamare in causa concetti non riconducibili all'evidenza immediata, in primo luogo; perché le proprietà che le dimostrazioni di consistenza richiedono, poi, derivano dall'analisi del significato di questi concetti astratti.

L'esempio al quale Gödel ricorre per illustrare quanto detto, è la dimostrazione di consistenza per l'aritmetica classica di Gentzen. In quel caso, il limite del ricorso all'intuizione relativa alle mere proprietà spazio-temporali delle dimostrazioni, intese come combinazioni di simboli di un alfabeto finitario, è rappresentato ovviamente dalla validità del principio di induzione fino all'ordinale ϵ_0: è in relazione a questo principio che "non si è più in grado di passare in rassegna le varie possibilità strutturali che si hanno per le successioni discendenti, e quindi non si è più in grado di riconoscere per mezzo dell'intuizione che ogni successione del genere termini necessariamente"[207].

[205] [Göd58], p. 240.
[206] *Ibid.*, p. 240.
[207] *Ibid.*, p. 242.

È quindi lo stesso aspetto della prova di Gentzen che permette di mettere in luce chiaramente dove il passaggio all'intuizione relativa ai "costrutti mentali", diviene inevitabile:

> In particolare, non si può acquisire una simile conoscenza [la fondatezza dell'ordinamento di tipo ϵ_0] *in modo intuitivo* passando un passo dopo l'altro da ordinali più piccoli a quelli più grandi; si può solo ottenere una conoscenza per via astratta per mezzo di nozioni di tipo superiore. Questo è possibile per mezzo della nozione astratta di 'accessibilità', che si definisce per il nostro essere in grado di dare una dimostrazione compresa informalmente del fatto che una certa inferenza è valida.[208]

Riguardo all' "accessibilità", alla capacità cioè di convincersi della correttezza di una certa inferenza per via informale, ossia facendo ricorso a quelle intuizioni relative ai "costrutti mentali" che derivano dal loro "significato" e che perciò non possono essere ridotte alle proprietà combinatorie di successioni di simboli, Gödel dice qualcosa in più nella nota c alla traduzione inglese del testo ([Göd90b]): egli sottolinea infatti come il concetto di "accessibilità" e le nozioni ad esso affini, quelle a cui "si è fatto ricorso in quasi tutti i casi" di prove di non contraddittorietà[209], "creano l'ingannevole impressione di essere basati su un'*intuizione concreta* di certe *procedure infinitarie*, come il 'contare oltre ω' o 'passare in rassegna' gli ordinali più piccoli di un certo ordinale α"[210].

Per quanto potrebbe già essere sufficientemente chiaro il fatto che ci si trovi di fronte ad osservazioni di tipo essenzialmente nuovo sul superamento del programma di Hilbert rispetto alle precedenti, ci pare opportuno ricostruirne il quadro in modo più suggestivo, superando cioè una certa 'circospezione' della formulazione gödeliana.

Gödel infatti sembra suggerire un'analisi della matematica in una duplice direzione: da un lato alludendo alla suddivisione tra l'indagine sulle relazioni e le proprietà di *oggetti concreti* (matematica finitaria) e l'indagine che coinvolge *concetti astratti* o *strutture di pensiero*; dall'altro, distinguendo corrispondentemente l'*evidenza immediata* e la capacità di acquisire conoscenza mediante le nozioni astratte, che è garantita dall'*accessibilità*.

[208] *Ibid.*, p. 242.

[209] A tale proposito, Gödel aggiunge qui una serie di riferimenti a lavori che contengono dimostrazioni di consistenza e riferimenti riconducibili alla questione dell' "accessibilità" (tra essi, oltre al lavoro di Gentzen, articoli di Lorenzen, Schütte, Kreisel, Takeuti).

[210] [Göd90b], p. 272.

Siamo così condotti in modo naturale, nel riassumere le suggestioni provenienti dagli scritti di Gödel legati alla problematica del finitismo e del rapporto con i teoremi di incompletezza, verso tre ordini di riflessioni.

La prima è una considerazione di tipo squisitamente metamatematico, e dunque più estrinseca rispetto alla problematica concettuale al centro della disamina, e riguarda la possibilità di definire quale fosse, per Gödel, una caratterizzazione precisa, in termini di teoria della dimostrazione, dei confini della matematica finitaria.

Sulla base della definizione data in [Göd33] del sistema A e delle similutidini che possono essere riscontrate con l'aritmetica ricorsiva primitiva, parrebbe logico ritenere che quest'ultima costituisse per Gödel una cornice formale sufficiente per contenere la matematica finitaria[211]. Stabilito quindi che PRA verosimilmente è il limite inferiore per essa, rimane quindi il problema di individuare il limite superiore. Apparentemente, Gödel aveva sviluppato una posizione ben precisa al riguardo, che è già implicitamente contenuta nei passi succitati, come ebbe modo di esprimere in un lettera a Bernays del 24 gennaio 1967, nella quale egli si dice "convinto" del fatto che ϵ_0 rappresentasse un limite, "non meramente in pratica, ma di principio" per il finitismo di Hilbert e che fosse persino possibile "dimostrare ciò in modo convincente"[212].

Tornando al tema principale, la seconda osservazione scaturisce dalle suddivisioni menzionate in precedenza. Ci pare infatti che dietro quelle indicate da Gödel si possa intravedere il porsi di una questione ad un livello che potremmo chiamare *mentale* (o *epistemico*): perché Gödel non sembra limitarsi ad un'analisi della matematica di tipo teoretico, ma chiama direttamente in causa, con l' "accessibilità" in primo luogo, due modalità diverse mediante le quali si stabilisce una relazione, di tipo conoscitivo, con due diverse 'famiglie' di enti matematici (quelli alla base della matematica finitaria da un lato, appunto, e i concetti astratti, o 'costrutti mentali' dall'altro). Si pone quindi

[211] Vale la pena ricordare che il problema di caratterizzare in modo preciso, da un punto di vista della teoria della dimostrazione, la matematica finitaria ha rappresentato una questione sulla quale si dibatte da tempo. Il rifgerimento classico al quale si riconduce usualmente la controversa congettura che tale cornice formale sia da individuare nel sistema per l'aritmetica primitiva ricorsiva è [Tai81].

[212] [Göd03a], p. 254. Questo fatto viene sostanzialmente confermato dalle considerazioni presenti in [Göd90b] (cfr. in particolare, p. 274 nota f). La rilevanza di questa posizione di Gödel nel quadro dell'indagine metamatematica, è discussa da Feferman nella nota introduttiva alla corrispondenza tra Gödel e Bernays ([Göd03a], pp. 41-79).

il quesito circa la natura della facoltà conoscitiva che sovrintende l'intuizione delle proprietà combinatorie da un lato, e la comprensione dei "costrutti mentali" dall'altro che ne rende 'accessibili', ossia giustificabili in termini informali, quegli aspetti essenziali per le dimostrazioni di consistenza.

In terzo luogo, ci pare opportuno sottolineare un richiamo ad una questione, quella dell'intervento necessario di concetti astratti, cioè di ordine superiore, che era già emersa dall'analisi del §2.1. Il fatto che si giunga da due strade diverse (per quanto accumunate dal ruolo ricoperto in entrambe dai teoremi di incompletezza) a considerazioni analoghe, legittima il sospetto di trovarci di fronte ad un aspetto non secondario, ma costitutivo della concezione gödeliana.

Per poter dire qualcosa di più al proposito, occorre tuttavia fare un passo indietro (cronologicamente parlando) rispetto ai contributi sulla 'Dialectica', e considerare i testi più legati agli sviluppi dell'indagine sull'Ipotesi del Continuo di Cantor, che, come è noto, si situano in una fase cruciale per l'elaborazione del pensiero di Gödel.

2.3 L'incompletezza e la prassi della matematica: riflessioni sull'indagine intorno all'Ipotesi del Continuo di Cantor

Come si è visto nel §1.3, alla fine degli anni '30, sull'onda del risultato relativo alla consistenza dell'Ipotesi del Continuo di Cantor e dell'Assioma di Scelta rispetto alla teoria degli insiemi, e sulla base del proprio personale convincimento circa l'indecidibilità delle proposizioni in questione in ZF, Gödel fece alcune osservazioni che ne mettevano direttamente in relazione il caso con quello degli enunciati formalmente indecidibili scoperti in precedenza.

Secondo quanto risulta dalle considerazioni di [Göd38c] e da quelle contenute nelle conferenze di Göttingen e presso la Brown University[213], Gödel riteneva che il discrimine tra i due tipi di indecidibilità consistesse principalmente nel fatto che, laddove gli enunciati formalmente indecidibili di [Göd31b] divengono decidibili in sistemi che intrattengono con il formalismo di partenza una relazione precisa, nel caso delle proposizioni legate all'Ipotesi del Continuo, siamo di fronte ad una indecidibilità che possiede le caratteristiche dell'assolutezza.

[213] [Göd38b], [Göd40].

Ciò significa, si badi bene, che per Gödel all'epoca esistevano casi di proposizioni matematiche rispetto alle quali non solo non si era contingentemente in grado di offrire alcuna dimostrazione o refutazione, ma che non si sarebbe potuto in nessun caso presentare alcuna forma di inferenza che potesse venire riconosciuta come una prova matematica di tale proposizione o della sua negazione.

Alla luce di quanto si è visto più in dettaglio nei §§2.1, 2.2, c'è un'altro aspetto degno di interesse dell'indagine gödeliana di quel periodo, e che emerge dalla caratterizzazione matematica delle proposizioni indecidibili come problemi di tipo diofanteo. Quel risultato che, come si è ricordato, pareva a Gödel ricoprire un ruolo importante soprattutto per i risvolti puramente matematici dell'incompletezza, e quindi per colmare la distanza, diremmo noi, tra la metamatematica, alla quale appartiene originariamente la scoperta gödeliana, e la prassi della matematica, rispetto alla quale poteva non apparire immediatamente chiara la rilevanza dell'indagine sui sistemi formali, acquisisce anche un significato diverso. Nel dimostrare l'esistenza di problemi numerici che risultano essere irresolubili se non attraverso il concorso di assunzioni di ordine superiore (ad esempio particolari istanze dell'assioma di comprensione), si forniva anche un esempio 'tangibile', e, soprattutto, matematicamente pregnante, della necessità del ricorso alle nozioni astratte ai fini del ragionamento matematico, a cui Gödel è giunto attraverso le altre direzioni che si sono fin qui analizzate.

Se si vuole quindi, queste considerazioni rappresentano un salto di qualità decisivo per la riflessione gödeliana sull'incompletezza e per lo sviluppo della concezione filosofica di Gödel: il momento in cui, l'emergere dell'elemento più dichiaratamente matematico dei propri risultati sull'indecidibilità, costituisce l'occasione per il precisarsi di una visione complessiva della natura della disciplina.

Come si ricordava al termine della parte precedente, un ruolo essenziale in quest'ultima direzione è ricoperto dai contributi più legati alle ricerche in campo insiemistico.

2.3.1 Nuovi assiomi per la matematica

Consideratane la forte connessione alla teoria dei tipi russelliana, è curioso che proprio in un lavoro dedicato alla logica matematica di Russell si profili quel superamento della posizione relativa all'esistenza di problemi matematici assolutamente irresolubili.

Si è visto infatti come la convinzione di Gödel relativa al carattere assoluto dell'indecidibilità di proposizioni legate all'Ipotesi cantoriana del Continuo, derivasse essenzialmente dal notare come la dimostrazione della loro consistenza rispetto a ZF trascendesse, ossia continuasse a valere inalterata lungo la gerarchia dei formalismi costruita secondo l'idea della stratificazione in tipi: gli assiomi per i numeri naturali in basso, quelli per gli insiemi di numeri naturali a seguire, gli insiemi di insiemi poi, e così via.

Tuttavia, è proprio l'articolo scritto da Gödel per il volume dedicato alla filosofia di Russell e facente parte della "Library of Living Philosophers" curata da Paul Arthur Schilpp, che contiene il primo riferimento a dei nuovi principi sulla base dei quali risolvere questioni matematiche indecise:

> Appare inoltre plausibile ritenere che per decidere certi problemi della teoria astratta degli insiemi e anche certe questioni della teoria dei numeri reali, saranno necessari nuovi assiomi basati su qualche idea finora sconosciuta. Forse le stesse difficoltà apparentemente insormontabili che altri problemi matematici hanno presentato per molti anni sono dovute al fatto che gli assiomi necessari non sono stati ancora trovati.[214]

Sebbene non venga indicato in modo esplicito, il tono dell'osservazione di Gödel sembra compatibile con quel mutamento d'opinione rispetto al contesto nel quale egli aveva accennato all'esistenza di proposizioni assolutamente indecidibili, pur essendo l'ideazione e la stesura di [Göd44] immediatamente susseguente al materiale sulla teoria degli insiemi di riferimento[215].

Occorre ricordare che il convincimento relativo al carattere assoluto dell'indipendenza di certe proposizioni rispetto agli assiomi per la matematica, aveva presumibilmente un fondamento oggettivo: l'esistenza di una dimostrazione di indipendenza dell'assioma di costruibilità rispetto alla teoria dei tipi[216]. L'allusione a possibili principi, basati su intuizioni "ancora sconosciute" e ad assiomi "non ancora scoperti", sembra indicare dunque che Gödel avesse acquisito, per qualche motivo, una maggiore prudenza nel leggere quel risultato, non vedendovi più, in particolare, una riprova del fatto (come l'avverbio 'assolutamente' lascia pensare) che nessuna prova corretta potesse

[214] [Göd44], p. 121.

[215] Delle due conferenze inedite nelle quali compare l'allusione alle proposizioni assolutamente indecidibili, la più tarda è la conferenza alla Brown University del 1940. L'invito a Gödel da parte di Schilpp a contribuire al volume su Russell è del 18 novembre 1942. La prima bozza del manoscritto di [Göd44], fu spedita invece da Gödel il 17 maggio 1943 e la versione rivista e corretta, il 28 settembre dello stesso anno.

[216] Si veda quanto detto alla nota 125.

mai essere offerta in futuro per risolvere l'indecidibilità di certe questioni matematiche.

Si deve tenere presente che il venir meno del fondamento cruciale della riflessione gödeliana sulla natura della supposta indipendenza di certe proposizioni insiemistiche offerta fino ad allora, si accompagna, nell'articolo sulla logica matematica di Russell, alla sottolineatura di quell'elemento di interesse della versione 'matematica' dell'indecidibilità da noi indicato poc'anzi.

L'occasione è data dall'analogia russelliana tra la matematica e le scienze naturali e tra l'evidenza logica e la percezione sensoriale. Non è difficile scorgere nel commento da parte di Gödel, che scaturisce dall'osservazione di come il confronto proposto da Russell "sia stato ampiamente giustificato dagli sviluppi successivi e c'è da aspettarsi che lo sarà in misura anche maggiore in futuro", il riferimento al risultato inedito di [Göd38c] da un lato, ed un'affinità, anche di linguaggio, con la distinzione tra la matematica 'del concreto' e quella 'dell'astratto' emersa in chiusura del §2.2:

> È risultato che (sotto l'assunzione che la matematica moderna sia consistente) la soluzione di certi problemi aritmetici richieda l'uso di ipotesi che trascendono in modo essenziale la teoria dei numeri, cioè l'ambito di quel tipo di evidenza elementare e indiscutibile che può essere più propriamente avvicinata alla percezione sensoriale.[217]

Dunque, il confronto tra i testi inediti e gli scritti pubblicati offre indicazioni divergenti: da un lato, un'apparente discrasia per ciò che concerne l'allusione a proposizioni assolutamente indecidibili negli uni e l'accenno a nuovi assiomi per la matematica negli altri, dai quali si fa dipendere la possibile soluzione di quesiti irrisolti anche nel campo della teoria dei numeri reali. Sotto un altro punto di vista, quello dell'esistenza di problemi numerici algoritmicamente indecidibili, la situazione è diversa: il risultato inedito viene cioè organicamente inserito, per così dire, nel quadro della riflessione gödeliana sull'incompletezza.

Occorre tuttavia completare il quadro, se si vuole che l'analisi possa essere condotta in modo corretto. E per farlo, bisogna affrontare in modo più deciso un aspetto essenziale che è finora rimasto soltanto sullo sfondo: la concezione platonista di Gödel della matematica.

Prima di farlo, esaminando i contributi di Gödel dedicati all'Ipotesi del Continuo e materiale ad essi legato, vogliamo, al fine di conferirgli maggiore

[217] *Ibid.*, p. 121

sostanza, soffermarci ancora sulle questioni emerse in [Göd44] alla luce di fonti ulteriori.

2.3.2 Gli interventi di Tarski e Gödel alla conferenza di Princeton del 1946

Tra il 17 ed il 19 dicembre 1946 fu tenuta a Princeton una conferenza sui problemi della matematica, nell'ambito delle molteplici iniziative messe a punto per la celebrazione del bicentenario della Princeton University. Furono invitati e parteciparono ai lavori alcuni tra i più eminenti studiosi di varie aree della matematica, i cui interventi furono organizzati in sezioni gestite da *discussion leaders*. Nel caso della sessione dei lavori dedicati alla ricerca nel campo della logica matematica, alla quale Gödel partecipò con un proprio intervento, Alfred Tarski fu scelto come moderatore, ed in tale veste egli tenne il 17 dicembre una sorta di discorso di apertura[218].

Dal materiale relativo alla corrispondenza tra Gödel e Tarski immediatamente precedente la conferenza, risulta chiaro che Tarski inviò a Gödel quantomeno un sunto del proprio intervento. Questo spiega le forti analogie che emergono dal confronto tra il testo del discorso di Tarski e la versione riveduta e corretta dell'intervento di Gödel, che sono tali da far pensare che quest'ultimo sia stato concepito come un commento al primo[219].

D'altra parte, i temi che Tarski intendeva trattare e che si riferivano direttamente alle ricerche di Gödel, non erano affatto estranei al tipo di riflessione a cui lo stesso Gödel era giunto, ed anzi richiamavano da vicino alcune delle osservazioni che abbiamo incontrato finora.

Questa identità di vedute emerge chiaramente dalla lettera del 10 dicembre 1946 di Tarski:

> Riguardo alle questioni alle quali sei interessato[220], non credo che riuscirò a fare niente altro che enfatizzare la differenza fondamentale tra la totalità

[218]Molte notizie relative sul contesto nel quale si inserisce l'organizzazione dell'evento, oltre alla riproduzione integrale dell'intervento di Tarski ed un commento esaustivo ad esso, possono essere trovate, insieme ad altro materiale relativo alla conferenza, in [Sin00] (vi si ritrovano inoltre il programma complessivo della conferenza e ad un resoconto della tavola rotonda conclusiva della sessione di logica, pensata evidentemente in funzione della pubblicazione degli atti, prevista in un primo tempo e poi abbandonata).

[219]L'intervento di Gödel è riprodotto in [Göd65b] che, come recita la voce bibliografica, è apparso per la prima volta in [Dav65].

[220]*NdT*: La scelta, in questo caso, di rendere la lettera con la seconda persona singolare, è dettata dall'indubbia familiarità tra i due (come dimostra il "Caro Kurt" con il quale si

> degli enunciati indecidibili fin qui noti nell'aritmetica elementare da un lato, e alcuni enunciati indecidibili in analisi e nella teoria degli insiemi (come l'ipotesi del continuo); in quanto gli enunciati indecidibili del primo tipo sono chiaramente indecidibili in un senso relativo, mentre quelli del secondo sembrano essere indecidibili in qualche senso assoluto. E in relazione a ciò solleverò i problemi (1) se (e come) la nozione di indecidibilità relativa ed assoluta possano essere precisate e (2) se, sulla base di una definizione adeguata di queste nozioni, sarà possibile dimostrare che un problema di tipo aritmetico può essere indecidibile solo in senso relativo.[221]

Non essendo stata ritrovata tra le carte del *Nachlass* la lettera di Gödel alla quale risponde questa missiva di Tarski, non sappiamo quanto i commenti del logico polacco del passo precedente, che possiedono una somiglianza sorprendente con le osservazioni di Gödel analizzate in precedenza, possano essere state ispirate dal suo contenuto. Salta ovviamente agli occhi in questo senso, la distinzione tra la diversa natura dell'indecidibilità in campo aritmetico e nel campo dell'analisi e della teoria degli insiemi.

In linea con quanto preannunciato a Gödel, l'intervento di Tarski ha nei riferimenti al problema della decisione alcuni passaggi cruciali. A parte un passo che riprende quasi alla lettera quello della citazione precedente[222], spicca un'osservazione ulteriore, dalla quale dipende, a detta del logico polacco, la possibilità di "aumentare il prestigio dei logici agli occhi dei matematici", e che riguarda la natura degli enunciati formalmente indecidibili:

> Nessuno di questi problemi è esattamente soddisfacente per il matematico: sarebbe un grande passo nella direzione di placare i matematici se fosse possibile costruire una proposizione relativa ai numeri interi similmente indecidibile. Dopo tutto, il matematico ritiene che questi enunciati indecidibili siano stati costruiti ai fini della loro stessa indecidibilità, ed ha grandi difficoltà a convincersi del fatto che egli avrà modo di imbattersi in enunciati simili nel corso delle proprie ricerche di ordinaria matematica.[223]

Avendo già sottolineato come la dimostrazione da parte di Gödel dell'esistenza di una classe di problemi diofantei per la quale non esiste una procedura ricorsiva generale di decisione rivesta una certa importanza proprio nell'avvicinare un risultato originariamente confinato in ambito metamatematico alla pratica della disciplina, non dovrebbe sorprendere più di tanto

---

apre la lettera), frutto di una frequentazione decennale (si veda a tale proposito anche la nota introduttiva alla corrispondenza tra Gödel e Tarski).

[221] [Göd03b], p. 272.

[222] Cfr. [Sin00], p. 26, punto 17.

[223] [Sin00], p. 25.

il fatto che, nel seguito, Tarski, avendo ricevuto da Gödel "il permesso di fare riferimento ai suoi lavori non pubblicati" (dei quali Tarski doveva essere stato informato presumibilmente da Gödel stesso), sottolinei come "si sia estremamente vicini al culmine di una simile impresa"[224].

È chiaro comunque che di ben altra suggestione appare essere il richiamo da parte di Tarski alla necessità di precisare la nozione di 'indecidibilità assoluta', al fine di poter evidenziare in modo matematicamente irreprensibile la differenza tra le proposizioni indecidibili nei sistemi adeguati per la teoria dei numeri, e gli enunciati specificamente relativi all'analisi e alla teoria degli insiemi. Dagli accenni in cui traspare la convinzione di Gödel dalla quale abbiamo preso le mosse, pare emergere l'idea, come si è già avuto modo di sottolineare, che la natura assoluta delle proposizioni indipendenti dalla teoria degli insiemi, si legasse al fatto che estendere un formalismo costruendo a partire da esso una gerarchia di sistemi tipati, costituisse per Gödel il modo più generale per ottenere sistemi formali via via più potenti. Essendo Gödel una personalità tutt'altro che avventata, parrebbe logico desumere che egli fosse convinto in prima istanza, di poter dare una definizione sufficientemente precisa di 'indecidibilità assoluta' su questa base, in modo da poter giustificare le proprie affermazioni al proposito.

Alcune indicazioni al riguardo, così come altre relative a quello che abbiamo creduto di dover interpretare come il mutamento di indirizzo successivo, si possono ritrovare nel testo dell'intervento di Gödel a Princeton. Soprattutto perché è proprio la questione dell''assolutezza' che permea tutto il discorso gödeliano: prendendo spunto dalla sottolineatura da parte di Tarski dell'importanza ricoperta dalla definizione del concetto di funzione ricorsiva per gli sviluppi della logica, Gödel sottolinea come, a suo dire, il merito principale di tale definizione consista nel fatto che "con questo concetto si è per la prima volta riusciti a dare di una nozione epistemologicamente interessante una definizione assoluta, cioè indipendente dal formalismo scelto"[225].

Si noti intanto come, intesa come risposta alla richiesta di Tarski, l'affermazione di Gödel, che appare del tutto identica a quella relativa alla nozione di 'funzione computabile in un sistema formale' di [Göd36] che abbiamo incontrato nel primo capitolo, identifica l' 'assolutezza' di un concetto definito formalmente con l'indipendenza di tale definizione dal formalismo scelto

[224] *Ibid.*, p. 26. Siamo d'accordo quindi con Sinaceur (cfr. *ibid.*, p. 31, nota 30) nel ritenere che Tarski si riferisca qui al risultato di [Göd38c].

[225] [Göd65b], p. 150.

(dove 'indipendente' significa, almeno nel caso del concetto di 'funzione computabile', che la classe di oggetti individuata è estensionalmente equivalente nel caso di ogni possibile scelta del formalismo di riferimento).

Da qui, Gödel prende le mosse per sottolineare, da un lato, come un simile successo appaia del tutto isolato nel panorama delle discipline matematiche, e dall'altro, per indicare i propri suggerimenti per la trattazione di due casi significativi, quelli dei concetti di 'dimostrabilità' e di 'definibilità'. Ovviamente, sono le osservazioni relative alla prima delle due nozioni, che offrono qualche spunto di riflessione per l'indagine intrapresa nel presente paragrafo.

In primo luogo: in che senso il concetto di 'dimostrabilità' in matematica non può essere definito in modo indipendente dal formalismo scelto? La risposta dipende dai teoremi di incompletezza: per quanto possa essere definito in modo impeccabile, un qualsiasi sistema formale (che sia consistente e, in un senso preciso del termine, adeguato per la matematica) risulta essenzialmente incompleto, nel senso che "la contemplazione del formalismo stesso dà origine a nuovi assiomi che sono altrettanto evidenti e giustificati di quelli da cui si è partiti", e "questo processo di estensione può essere iterato nel transfinito"[226].

Se si è 'scelto' un sistema formale dandone gli assiomi, ai quali guardiamo quindi come a proposizioni pienamente dotate di senso e corrette, è naturale, ad esempio, guardare alla proposizione che esprime la consistenza del sistema così definito, come ad un enunciato altrettanto dotato di senso e corretto; quindi, essendo esso indipendente dagli assiomi indicati, come ad un nuovo principio che estende i precedenti in modo essenziale. Poiché la stessa operazione può essere ripetuta per il sistema esteso e così via, per tutte le estensioni individuate in questo modo, segue che "non ci può essere nessun formalismo che comprenda tutti questi passaggi", ossia, equivalentemente, non esiste una procedura ricorsiva che possa 'comprendere' (enumerare) tutte le proposizioni dimostrabili per mezzo di quelle estensioni.

Tuttavia, dice Gödel "ciò non esclude che tutti questi passaggi (o almeno quanti, tra questi, forniscono qualcosa di nuovo per il dominio di proposizioni a cui si è interessati) possano essere descritti e raccolti insieme in qualche modo non-costruttivo"[227]. L'attenzione di Gödel si concentra sulla teoria degli insiemi, dove egli nota come il tipo di estensioni in questione possono essere "più convenientemente" rappresentate mediante l'aggiunta ad un sistema

[226] *Ibid.*, p. 151.
[227] *Ibid.*, p. 151.

formale di partenza di 'assiomi dell'infinito' sempre più forti (cioè proposizioni che sanciscono l'esistenza di collezioni di cardinalità sempre maggiore). Il risultato, cioè, non cambia: ogni nuovo teorema che è possibile ottenere estendendo ZF mediante il 'passaggio' al tipo superiore, può essere ricavato aggiungendo alla teoria degli insiemi un opportuno assioma dell'infinito[228].

Ora, se per questo motivo non è possibile caratterizzare in modo costruttivo la totalità degli assiomi dell'infinito, "potrebbe esistere, ad esempio, una caratterizzazione [non costruttiva] del tipo seguente: Un assioma dell'infinito è una proposizione che possiede una certa struttura formale (decidibile) e che, in aggiunta, è vero"[229].

Per mezzo di una simile caratterizzazione, si potrebbe così offrire di conseguenza una definizione del concetto di 'dimostrabilità insiemistica', e rispetto ad esso Gödel osserva quanto segue:

> Non è impossibile che per un tale concetto di dimostrabilità valga qualche teorema di completezza che asserisca come ogni proposizione esprimibile nella teoria degli insiemi risulti decidibile dagli assiomi attuali più qualche asserzione vera sull'estensione dell'universo insiemistico.[230]

Ciò significa quindi in particolare, che la proposizione che esprime l'Ipotesi del Continuo di Cantor, o quelle ad essa legate, che risultano indipendenti da ZF, potrebbero divenire decidibili per questa via.

L'interesse per il luogo analizzato non è dissimile a quello per i passi di [Göd44] citati in precedenza: così come questi ultimi rappresentano il primo momento in cui interviene il riferimento alla necessità di nuovi assiomi per 'completare' la matematica, il periodo succitato di [Göd65b] si caratterizza per il fatto che tra i principi che garantiscono una tale estensione, vengono menzionati gli assiomi dell'infinito intorno ai quali, come è noto, si incentra tutta la riflessione gödeliana sull'indagine relativa all'Ipotesi del Continuo.

Rispetto al problema della natura assoluta dell'indecidibilità di certe proposizioni insiemistiche, il passo analizzato fornisce qualche indicazione, sep-

[228]Come fa notare Feferman nel suo [Fef87], ogni sistema 'ragionevole' S di teoria degli insiemi possiede un modello a qualche livello V_α (per un certo numero cardinale α) della gerarchia cumulativa. Quindi il concetto di verità per S è definibile in un sistema ottenuto da esso mediante un assioma che asserisca l'esistenza di un α che soddisfa qualche opportuna proprietà. A titolo di esempio, è possibile definire il predicato di verità per ZFC in un sistema S:= ZFC + $\exists x Inacc(x)$, dove $Inacc(\alpha)$, esprime il fatto che α è un cardinale inaccessibile.

[229]*Ibid.*, p. 151.

[230]*Ibid.*, p. 151.

pure con una valenza di chiarificazione solo parziale: come ci si poteva aspettare, si ritrova un riferimento esplicito al problema di fornire una qualche descrizione generale del concetto di dimostrabilità che, in linea con la sollecitazione dell'intervento di Tarski, dia la possibilità di parlare in relazione ad esso di 'assolutezza' e 'relatività' con cognizione di causa e precisione (allo stesso modo in cui avevamo letto le affermazioni di Gödel sullo *status* di certe proposizioni insiemistiche indipendenti da ZF, alla luce di una concezione della teoria dei tipi come una cornice entro la quale poter dare una simile descrizione). L'esistenza di una simile velleità dietro l'intervento gödeliano, ci pare di riscontrarla nel resoconto della discussione avvenuta alla conferenza di Princeton al termine dei lavori della sessione di logica[231], laddove si dà testimonianza di un intervento di Gödel il quale volle far notare come la propria conferenza non era stata concepita per definire *un* sistema formale, bensì per trattare "il problema di definire una teoria dell'intera serie transfinita di sistemi formali sempre più potenti per la matematica"[232].

Da questo punto di vista, il riferimento agli assiomi dell'infinito non appare risolutivo: perché se è vero che è in relazione al concetto di dimostrabilità ottenuto mediante l'estensione della teoria degli insiemi con questo tipo di assiomi che Gödel allude alla possibilità di dimostrare un "teorema di completezza", è altrettanto vero che il riferimento ad un'estensione per questa via del formalismo di partenza è presentato semplicemente come "più conveniente" rispetto all'usuale estensione per aggiunta degli assiomi per i tipi superiori.

Sembra una conferma di ciò, il fatto che il riferimento all'estensione della teoria degli insiemi mediante "assiomi forti dell'infinito", sia già presente nella prima presentazione del risultato di consistenza ([Göd38a]), dove tuttavia all'osservazione che "la prova di consistenza per [l'assioma di costruibilità] non viene meno" nel formalismo così esteso, fa seguito l'ormai ben noto commento che, per questo, la dimostrazione "sembra essere in qualche senso assoluta, anche se allo stato attuale non è possibile dare un significato preciso a quest'affermazione"[233].

Se questo è vero, ed il riferimento agli assiomi dell'infinito va preso solo a titolo di esempio di una convinzione più profonda, occorre concludere che i passi fin qui considerati non contengano sufficienti informazioni per poter

[231] Si tratta del resoconto preparato da J. C. C. McKinsey e rivisto da J. W. Tuckey, al fine della pubblicazione surricordata degli atti della conferenza.
[232] [Sin00], p. 34.
[233] [Göd38a], p. 27.

tirare le somme su questa fase della riflessione gödeliana, in cui le considerazioni legate ai teoremi di incompletezza si combinano e si completano mediante quelle derivanti dalle ricerche in campo insiemistico.

Sembra dunque naturale, rivolgere lo sguardo a quei lavori che si propongono in modo più diretto l'analisi dei vari aspetti della congettura cantoriana.

2.3.3 Quali sono le implicazioni del problema del continuo di Cantor?

Nel 1945, e dunque proprio nel periodo al quale appartengono anche i lavori di cui ci siamo occupati finora, a Gödel giunse la richiesta da parte del curatore dell' "American Mathematical Monthly" di un articolo espositivo dedicato all'Ipotesi del Continuo di Cantor. Il risultato, il lavoro del 1947 intitolato "Cos'è il problema del continuo di Cantor?", deve certamente parte della sua fama, oltre che alla chiarezza dell'esposizione, alla presenza di numerose osservazioni da parte di Gödel che risentono dell'interesse per gli aspetti filosofici del problema che egli andava via via maturando[234].

Cercando di isolare in modo decisamente pragmatico, solo quei passi che possano essere rilevanti per la presente discussione (considerando anche che si è fatto riferimento alle ricerche insiemistiche di Gödel in questa monografia, solo in quanto i suoi commenti al riguardo si intrecciano con quelli più direttamente relativi ai teoremi di incompletezza e alle loro implicazioni), ci si accorge facilmente di come molte delle allusioni che abbiamo riscontrato fin qui riemergano, corredate, talvolta, di qualche osservazione significativa ulteriore.

Vi si ritrova menzione innanzi tutto di quella che, a fronte di qualche discrepanza nel pensiero di Gödel che si è voluto mettere in evidenza, rappresenta una vera e propria costante della posizione gödeliana a proposito

[234]Ci sono molte testimonianze, anche dirette, relative ad una vera e propria "svolta filosofica" negli interessi di Gödel, collocata all'inizio degli anni '40 in concomitanza con un attento studio dell'opera di Leibniz. A poco a poco, l'interesse per problematiche di natura squisitamente filosofica cominciò a prendere l'avvento sulle questioni tecniche, al punto da indurre Gödel ad abbandonare i tentativi di una dimostrazione di indipendenza dell'Ipotesi del Continuo di Cantor dagli assiomi della teoria degli insiemi (quest'ultimo fatto è testimoniato direttamente dallo stesso Gödel nel carteggio con Church - cfr., in particolare, [Göd03a], pp. 369 e segg. - fatto del quale si è già avuto modo di parlare). Questo spiega il taglio decisamente filosofico di quasi tutto il materiale gödeliano (soprattutto inedito) dagli anni '40 in poi.

dello statuto dell'Ipotesi del Continuo di Cantor, ossia la convinzione circa la sua indipendenza dagli assiomi della teoria degli insiemi:

> Ovviamente, [...] ci sono (assumendo la consistenza degli assiomi) tre possibilità a priori rispetto alla congettura di Cantor: Essa potrebbe essere dimostrabile, indimostrabile o indecidibile. La terza alternativa (che è solo una formulazione precisa della congettura avanzata in precedenza in riferimento al fatto che le difficoltà del problema non sono forse di natura esclusivamente matematica) appare la più probabile, e cercare una dimostrazione di questo fatto sembra attualmente uno dei modi più promettenti per attaccare il problema.[235]

A questa osservazione, Gödel aggiunge il proprio convincimento che la congettura cantoriana sia *falsa*, sulla base delle "numerose proposizioni plausibili" che ne implicano la negazione a fronte del fatto che "neanche una proposizione plausibile" implica l'Ipotesi del Continuo, rovesciando così quello che sembra essere stato il suo primo orientamento al riguardo. È bene notare che, al contrario di quanto accade con il convincimento sull'indipendenza dell'ipotesi di Cantor dagli assiomi della teoria degli insiemi, dietro *questa* posizione sullo statuto aletico della congettura cantoriana, si celerebbe un altro mutamento d'opinione di Gödel: nella prima presentazione del risultato di consistenza[236], egli parla infatti dell'assioma di costruibilità, dal quale segue logicamente la proposizione che esprime la congettura di Cantor, come di un "completamento naturale" degli assiomi della teoria degli insiemi (un'idea incompatibile, per quanto detto, con la falsità dell'Ipotesi del Continuo)[237].

Accanto a, ed in connessione con, il riferimento alla plausibilità dell'indipendenza dell'Ipotesi del Continuo rispetto agli assiomi della teoria degli

[235][Göd47], p. 181.

[236][Göd38a], p. 27.

[237]Come si spiega un simile mutamento d'opinione? Considerato che l'assioma di costruibilità può essere visto come una 'limitazione' dell'universo insiemistico, ciò vuol forse dire che Gödel era più costruttivista alla fine degli anni '30 di quanto non lo fosse alla metà degli anni '40 del '900? A parte simili disquisizioni, occorre notare, per dovere di completezza, che la 'storia' dell'attitudine di Gödel nei confronti dell'Ipotesi del Continuo ha un appendice in due contributi che risalgono al 1970 circa ([Göd70b] e [Göd70a]), con i quali Gödel ritiene, rispettivamente, di poter fornire una confutazione della congettura cantoriana (ovvero una dimostrazione che la cardinalità del continuo è $\aleph_2$), e una sua dimostrazione, per mezzo di nuovi assiomi insiemistici. Trattandosi di materiale che ha fortemente risentito del precario stato di salute del Nostro (e quindi, per tale motivo, relativamente attendibile, per quanto di indubbio interesse), si è preferito non citarlo, in questa sede, e si rimanda alla nota introduttiva ad esso per ogni commento al proposito.

insiemi, la convinzione che un tale esito della ricerca non elimini affatto l'interesse nel proseguire l'indagine, viene connesso da parte di Gödel (in [Göd47]), da un lato, ad un approccio di tipo platonista, dall'altro all'idea di possibili nuovi assiomi che potrebbero permettere di determinare la correttezza o l'erroneità della congettura cantoriana:

> Si deve notare, tuttavia, che anche se si riuscisse a dimostrarne l'indimostrabilità [da ZFC], ciò [...] non sistemerebbe affatto la questione in modo definitivo. Solo chi (come l'intuizionista) negasse che i concetti e gli assiomi della teoria degli insiemi classica non hanno significato (o non hanno un significato ben preciso) potrebbe ritenersi soddisfatto da una simile soluzione, non chi credesse che essi descrivono una qualche ben determinata realtà. Perché in questa realtà la congettura di Cantor deve essere vera o falsa e la sua indecidibilità dagli assiomi conosciuti oggi può significare solamente che questi assiomi non contengono una descrizione completa di questa realtà; e tale convinzione non è affatto illusoria, dal momento che è possibile indicare dei modi mediante i quali potrebbe essere raggiunta una decisione sulla questione, anche se questa fosse indecidibile rispetto agli assiomi nella loro forma presente.[238]

Relativamente ai principi in questione ed ai possibili modi per poter raggiungere una decisione riguardo alla congettura cantoriana, Gödel sottolinea come la natura di questi principi debba essere, a suo modo di vedere, determinata da un'approfondimento dell'analisi, e di conseguenza della comprensione, delle nozioni chiave di questa branca della matematica:

> Perché in primo luogo la collezione degli assiomi della teoria degli insiemi non formano affatto un sistema chiuso in se stesso, ma, al contrario, lo stesso concetto di insieme sul quale essi si fondano, suggerisce un'estensione mediante nuovi assiomi che asseriscano l'esistenza di iterazioni ancora ulteriori dell'operazione "insieme di".[239]

Se è pur vero che il passo si riferisce ovviamente ai già citati 'assiomi dell'infinito' (come è confermato d'altra parte da ciò che segue nel testo), è anche vero che in relazione al problema specifico, decidere la congettura di Cantor, Gödel aggiunge:

> Riguardo al problema del continuo, ci sono poche speranze di poterlo risolvere sulla base di principi oggi noti (la succitata dimostrazione dell'irrefutabilità dell'ipotesi del continuo, ad esempio, vale per tutti questi senza alcun cambiamento). Ma probabilmente ne esistono degli altri, basati su

[238] [Göd47], p. 181.
[239] *Ibid.*, p. 181.

> qualche principio finora sconosciuto; è possibile anche che, oltre agli assiomi ordinari, agli assiomi dell'infinito e [ad assiomi relativi al concetto di "proprietà di un insieme" - il secondo concetto primitivo della teoria[240]], esistano degli altri assiomi (fin qui sconosciuti) della teoria degli insiemi che una comprensione più profonda dei concetti sottostanti la logica e la matematica ci permetta di riconoscere essere implicati da questi concetti.[241]

Si sarà notato come il riferimento al fatto che la dimostrazione di consistenza dell'Ipotesi del Continuo 'trascende' tutte le possibili estensioni della teoria degli insiemi mediante principi noti, fa del passo in questione (non seguendo dalla prima osservazione la conclusione sulla natura assoluta della prova), un riferimento per quel mutamento d'opinione che si è voluto utilizzare come chiave di lettura lungo questa parte del lavoro.

In relazione ai principi che estendono gli assiomi usuali della teoria degli insiemi, Gödel non manca di far notare poi, passando direttamente attraverso la versione generale dei teoremi di incompletezza, le implicazioni di questi assiomi in campo aritmetico:

> Che questi assiomi hanno conseguenze anche al di fuori del dominio dei numeri transfiniti molto grandi, che costituiscono il loro oggetto [di riferimento] immediato, può essere provato; si può dimostrare, sotto l'assunzione di consistenza, che ognuno di essi (fin quando si tratta di principi noti) aumenta il numero di proposizioni decidibili anche nel campo delle equazioni Diofantee.[242]

Qualche indicazione ulteriore, soprattutto per ciò che concerne alcuni aspetti più filosofici della disamina, proviene dal supplemento alla seconda edizione di [Göd47], scritto in funzione della ripubblicazione del testo in [BP64], e dalle aggiunte a seguito della revisione successiva dell'intero lavoro (intorno al 1966) in previsione di una terza edizione (mai avvenuta).

Discutendo di una differenza essenziale tra la geometria e la teoria degli insiemi, Gödel introduce ad esempio il seguente commento che tradisce una certa visione circa lo statuto ontologico degli enti della teoria degli insiemi, e, più in generale, della matematica[243]:

---

[240] *NdT*: quest'aggiunta è tratta quasi letteralmente dalla nota 17 di [Göd47], alla quale, in questo punto, il passo originario di Gödel rimanda.

[241] *Ibid.*, p. 182.

[242] *Ibid.*, p. 182.

[243] Lo spunto per un confronto tra le due aree del sapere matematico, era la critica dell'accostamento tra la (ipotetica) indecidibilità dell'Ipotesi del Continuo dalla teoria degli insiemi e l'indipendenza del V postulato di Euclide dagli assiomi della geometria

[G]li oggetti della teoria degli insiemi [...] chiaramente non appartengono al mondo fisico, e persino la loro connessione indiretta con l'esperienza sensibile è molto blanda (a causa principalmente del fatto che i concetti insiemistici giocano solo un ruolo minore nelle teorie fisiche di oggi).

Ma a dispetto della loro lontananza dall'esperienza sensoriale, possediamo qualcosa di simile ad una percezione anche degli oggetti della teoria degli insiemi, come si vede dal fatto che gli assiomi ci appaiono come principi veri. Non vedo ragioni per le quali dovremmo avere meno fiducia in questo tipo di percezione, cioè l'intuizione matematica, rispetto alla percezione sensoriale, che ci porta a costruire teorie fisiche e ad aspettarci che in futuro la percezione dei sensi sarà in accordo con esse, e a credere inoltre che questioni non decidibili al momento possiedano un significato e possano essere decise in futuro.[244]

Si sarà notato come nell'alludere, attraverso il paragone con gli oggetti dell'esperienza sensibile, all'oggettività della realtà alla quale appartengono gli enti matematici (in particolare, quelli della teoria degli insiemi), Gödel inserisca anche il riferimento alla facoltà conoscitiva che garantisce, proseguendo il parallelo, la percezione di questa seconda realtà così come i sensi sono i garanti della relazione con la prima.

Sull'intuizione matematica, anche in relazione alla specifica indipendenza ontologica degli enti matematici, Gödel dice anche qualcos'altro:

Si dovrebbe notare che l'intuizione matematica non deve essere necessariamente concepita come una facoltà che fornisce una conoscenza *immediata* dell'oggetto a cui si riferisce. Piuttosto, sembra che, come nel caso dell'esperienza fisica, noi *formiamo* le nostre idee anche di quegli oggetti [della matematica] sulla base di qualcos'altro che *è* dato in modo immediato. Soltanto che questo qualcosa d'altro *non* è, o almeno non in modo primario, la sensazione. [...] Evidentemente ciò che è "dato" al di sotto della matematica è strettamente connesso con gli elementi astratti contenuti nelle nostre idee empiriche. Ciò non significa, tuttavia, che i dati di questo secondo tipo, dal momento che non possono essere associati con l'azione di certe cose sui nostri organi di senso, siano qualcosa di puramente soggettivo, come Kant ha asserito. Piuttosto, anch'essi possono rappresentare un aspetto della realtà oggettiva, ma, in opposizione alla sensazione, la loro presenza in noi può essere dovuta ad un altro tipo di relazione tra noi e la realtà.[245]

(euclidea), che era stato avanzato in [Err52] al fine di argomentare a favore di una perdita di significato dell'indagine intorno alla congettura cantoriana in caso di una sua provata indipendenza. Ci si ricorderà forse, ma siamo ormai avvezzi ai cambi di direzione di Gödel, che proprio Gödel aveva azzardato un simile paragone nella conferenza di Göttingen del 1938.

[244] [Göd64], pp. 267-268.

[245] *Ibid.*, p. 268.

Si era motivato l'analisi dei testi gödeliani più direttamente connessi all'Ipotesi del Continuo di Cantor, sull'onda di quelle osservazioni di Gödel che tendevano a mettere in relazione la scoperta di enunciati formalmente indecidibili, con quel caso 'concreto', ossia più vicino alla pratica matematica, di una proposizione indipendente da un sistema d'assiomi, avendovi ravvisato alcuni utili legami con il quadro più generale della filosofia di Gödel.

Dall'esame, sono emersi passi che sembrano confermare quell'impressione. È giunto il momento allora, di cercare di dire qualcosa di più conclusivo al proposito.

2.3.4 L'inesauribilità della matematica

Il richiamo diretto ai teoremi di incompletezza all'interno di quei testi di Gödel più legati alle problematiche della teorie degli insiemi, è duplice: da un lato si è visto come il riferimento all'indecidibilità, ed in particolare all'indecidibilità essenziale dei sistemi formali minimamente adeguati per la matematica, viene chiamato in causa (in [Göd65b], ad esempio) al fine di sottolineare come non sia possibile 'confinarsi' ad un unico sistema d'assiomi e che sia anzi necessario ammetterne comunque un'estensione che, da un punto di vista della legittimità, è altrettanto giustificata del sistema di partenza; in un secondo senso, Gödel si riferisce ai teoremi di incompletezza attraverso quella formulazione del risultato in termini di esistenza di una certa classe di problemi diofantei algoritmicamente indecidibile, ma i cui elementi divengono decidibili una volta che vengano ammesse nozioni di ordine superiore, per sottolineare il contributo degli 'assiomi dell'infinito' all'aritmetica.

Questo secondo tipo di osservazione dovrebbe sì essere avvicinato alle considerazioni di Gödel che abbiamo incontrato al termine della sezione dedicata al programma di Hilbert ed al finitismo, circa la caratterizzazione della matematica finitaria come 'matematica dell'evidenza concreta', a fronte di una 'matematica dell'astratto' che chiama in causa nozioni di natura diversa, ma si dovrebbe in una certa misura anche contrapporle a quelle stesse riflessioni: perché laddove queste ultime vengano lette esclusivamente solo come allusione ad una suddivisione della matematica, quelle da noi incontrate nel paragrafo precedente, sembrano più propriamente suggerire la volontà di costituire un'unità teoretica tra aree costitutivamente diverse del sapere matematico; indicano quindi la convinzione di Gödel che 'tutto si tiene insieme' nella matematica, che la conoscenza degli enti astratti, appunto, ha un riflesso diretto sulle nostre conoscenze o, meglio, le nostre percezioni relative a quelle nozioni per le quali sussiste un rapporto diretto ed immediatamente intuitivo.

Da questo punto di vista, è chiaramente indicativo il parallelo con le scienze naturali, e la fisica in particolare. Anche perché è attraverso questo paragone che si fa strada, in maniera sempre più decisa, il tema del 'platonismo' di Gödel. Per eliminare qualsiasi tentazione di una ricostruzione parodistica o frettolose attribuzioni di ingenuità filosofica (rischi che permangono, considerata soprattutto l'assenza, almeno nei testi fin qui presi in considerazione, di una trattazione sistematica della problematica da parte di Gödel), si dovrebbe prendere in seria considerazione quanto Gödel dice in [Göd64], dove il paragone tra matematica (teoria degli insiemi) e fisica è usato per rimarcare differenze e vicinanze: una differenza, essenziale, di tipo *ontologico*, derivante dal fatto che gli enti matematici non hanno una connessione diretta con i nostri organi di senso, per quanto la loro realtà sia altrettanto legittima di quella degli oggetti del mondo fisico; una vicinanza più netta, altrettanto rilevante, che emerge su un piano *epistemico*[246], relativamente cioè al modo con cui la conoscenza si dà nei due casi, attraverso un atto di costruzione dell'oggetto (che nel caso dell'oggetto fisico è ottenuto combinando il dato sensibile con la componenete astratta, ad esempio l'idea stessa di oggetto che è essenzialmente diversa dalle sensazioni); una costruzione che, tanto per la fisica quanto per la matematica, nulla toglie, tuttavia, alla completa indipendenza dal soggetto della realtà all'origine dei due tipi di atti conoscitivi.

Il modo in cui la questione del realismo matematico di Gödel si salda naturalmente con la questione di come si realizzi l' 'accesso' alla mondo oggettivo degli enti matematici, e quindi di come, attraverso cioè quale facoltà e quali caratteristiche della facoltà in questione, si verifichi la conoscenza matematica, ci dà conferma della rilevanza del problema di quel livello *mentale* dell'analisi gödeliana che, nella sezione sul finitismo, ci era parso emergere in modo ovvio, seppure rispetto ad uno schema che risulta troppo debitore della riflessione sul programma di Hilbert per essere generalizzato *tout court*. Tuttavia, occorre notare che il riferimento all'intuizione matematica nelle pagine del supplemento alla seconda edizione di [Göd47], pur fornendo quelle suggestioni di cui si è appena detto, non sembra ancora risolutivo.

A tale proposito ci preme proporre una riflessione ulteriore. Ci pare op-

[246] Nel contesto del presente lavoro, con quest'accezione si intende indicare ciò che riguarda la teoria della conoscenza nel campo delle scienze deduttive.

portuno notare cioè, come nell'economia della concezione gödeliana che si va delineando, il riferimento all'indecidibilità di una classe di proposizioni diofantee come segnale del fatto che la matematica non è completabile mediante un unico sistema d'assiomi, appare più marginale. Soprattutto se si tiene conto dell'affermazione di Gödel[247] che il ricorso a nuovi assiomi relativi ad iterazioni ulteriori dell'operazione di "insieme di", viene suggerito dallo stesso concetto di insieme. Un'osservazione, questa, che già di per sé è un sostegno al 'programma' gödeliano per la ricerca dei nuovi assiomi per la teoria degli insiemi. Ed è chiarendo i termini di questa marginalità che si può dire qualcosa d'altro sulla filosofia della matematica di Gödel.

Se in [Göd47] non c'è una vera e propria spiegazione di come il concetto di insieme suggerisca il ricorso a nuovi assiomi, quello che appare a tutti gli effetti come un argomento che illustra l'osservazione gödeliana, fa parte invece del testo della venticinquesima conferenza intitolata a Josiah Willard Gibbs, che Gödel tenne ad un incontro dell'American Mathematical Society svoltosi alla Brown University il 26 dicembre 1951, e della quale avremo modo di riparlare anche più avanti.

Ai fini dell'analisi di questo paragrafo, è sufficiente dire che l'intervento di Gödel è dedicato all'esame delle implicazioni filosofiche di certi teoremi metamatematici ed è totalmente incentrato sul fenomeno dell''inesauribilità della matematica', il quale si manifesta nel tentativo di dare degli assiomi dai quali discenda l'intera disciplina[248].

Il primo dei due argomenti che egli offre a sostegno di tale punto di vista[249] è caratterizzato per il fatto che si riferisce alla matematica "così come è, senza che venga decurtata in ragione di qualche concezione [fondazionale] critica", e si basa, per tale motivo, sull'assunzione che la matematica sia interamente riducibile alla teoria degli insiemi. L'elemento cruciale dell'argomento, che conduce all'individuazione di una successione inesauribile di assiomi (dell'infinito) per la teoria degli insiemi (ovvero, per la matematica), consiste nell'osservazione di Gödel secondo la quale, "se si vogliono evitare

[247] [Göd47], p. 181.

[248] Gödel si riferisce a questa operazione come al tentativo di assiomatizzare la "matematica propriamente detta", ossia isolare le proposizioni matematiche che "valgono in senso assoluto". Un'operazione tuttavia che, come lo stesso Gödel sottolinea, è stata storicamente condotta in modi molto diversi a causa dei "punti di vista ampiamente divergenti", all'origine delle dispute fondazionali, su quali fossero i confini di un simile nucleo di conoscenze.

[249] Cfr. [Göd51], pp. 305-308.

i paradossi della teoria degli insiemi senza introdurre qualcosa di completamente estraneo alla prassi matematica attuale", occorre procedere ad assiomatizzare il concetto di insieme "per passi successivi", ossia per iterazione dell'operazione 'insieme di'.

Come da ciò scaturisca il fatto che nessun sistema d'assiomi possa contenere tutta la matematica è presto spiegato: se si deve procedere per passi successivi, il punto di partenza naturale è il caso più semplice, ossia l'applicazione dell'operazione 'insieme di' ad una collezione di oggetti primitivi, ad esempio i numeri naturali; si costituisce così una nuova collezione, quella degli insiemi di numeri naturali, alla quale applicare nuovamente l'operazione. E così si può procedere per tutte le iterazioni finite della potenza insiemistica.

Ma a questo punto si può considerare la totalità degli enti ottenuti da una iterazione finita qualsivoglia dell'operazione 'insieme di', come nuova collezione alla quale applicare la stessa procedura. E per quanto si ammetta anche un'iterazione transfinita dell'operazione, si può sempre considerare la classe che è chiusa sotto di essa, generando una nuova collezione di partenza.

Ciò significa che: comunque ci si accontenti di un'iterazione dell'operazione di potenza insiemistica, sia essa finita o espletata lungo un qualsiasi ordinale transfinito, una nuova iterazione, e dunque un nuovo assioma viene suggerito in modo naturale, ossia il principio che sancisce l'esistenza della collezione che contiene tutti gli elementi generati fino a quel punto (e che costituisce una nuova base d'avvio per l'applicazione dell'operazione)[250].

L'argomento fornito da Gödel, permette di dare una spiegazione all'insistenza da parte sua sugli assiomi dell'infinito, che, fatto del quale (come si è già notato) lo stesso Gödel era consapevole già scrivendo [Göd47], non ha nulla a che fare, contrariamente a quanto erroneamente si sente dire talvolta, con la possibile soluzione del problema del continuo se non per il fatto che, come si è visto, egli era profondamente convinto che tale soluzione potesse giungere solo da un'analisi più approfondita della nozioni alla base della teoria degli insiemi. Ed è in questo senso che gli assiomi dell'infinito, per l'argomento appena visto, rivestono un ruolo paradigmatico: perché essi sono naturalmente connessi con il significato stesso del concetto 'iterativo' di insieme, che egli vedeva come il fondamento di una teoria matematica corretta (non antinomica) fondata sulla nozione informale di 'aggregato' o 'collezio-

[250] In realtà, l'argomento di Gödel è più generale di quanto non si sia qui ricostruito. Esso, infatti, applica la stessa conclusione alla procedura ottenuta iterando lungo un qualsiasi ordinale (transfinito) una *qualsiasi* operazioni insiemistica. Ai fini dell'esposizione il caso della sola operazione 'insieme di' è tuttavia sufficiente.

ne'. Gli assiomi che asseriscono l'esistenza dei grandi cardinali, nell'economia della concezione gödeliana, svolgono semplicemente il ruolo di esempio (un esempio che Gödel riteneva sicuramente calzante) che è indice di qualcosa di ben più rilevante e costitutivamente rappresentativo: il convincimento relativo al procedere della nostra conoscenza del mondo della matematica, per mezzo di principi che sono espressione della progressiva familiarizzazione con, e comprensione del, significato dei concetti che ne fanno parte; dove con il termine 'significato', Gödel intendeva in realtà la loro *essenza*.

Ma torneremo su questo punto nel prossimo paragrafo, cercando di argomentare a favore di quest'affermazione, ancora una volta, per mezzo degli scritti di Gödel.

Vogliamo invece chiudere il paragrafo presente, con un'ultima osservazione sull'indicazione da parte di Gödel dell'esistenza di possibili principi ancora sconosciuti sulla base dei quali decidere questioni matematiche irrisolte, e che sono all'origine dell'elemento forse maggiormente distintivo della posizione di Gödel in campo insiemistico (cioè dell'idea che la teoria degli insiemi non sia un "sistema chiuso" ma rappresenti un campo inesauribile della matematica). Desideriamo cioè chiudere, richiamandoci alla questione dalla quale siamo partiti: il supposto mutamento di opinione relativamente alla natura dell'indipendenza di certe proposizioni insiemistiche.

Se la caratteristica principale per la quale il materiale analizzato in questa sezione del lavoro viene comunemente ricordato è l'esplicito riferimento da parte di Gödel alla propria concezione platonista, ci pare che lo spunto dal quale si sono prese le mosse, che di per sé potrebbe non avere un grande significato considerato anche il 'complicato rapporto' intrattenuto dal logico austriaco con la congettura cantoriana, permetta di evidenziare un altro aspetto, altrettanto caratteristico, della posizione gödeliana. Non è difficile realizzare che una supposta "svolta platonista", che peraltro, come è noto, mal si concilia con le ricostruzioni retrospettive nelle quali Gödel sosterrà di aver aderito precocemente al realismo in matematica[251], non necessariamente

[251]Sono da considerarsi testimonianze in questo senso già le due lettere a Wang citate nel §2.1. Un riferimento comunemente ricordato al proposito è il cosiddetto 'Questionario Grandjean', ossia una serie di domande scritte volte a chiarire il quadro storico e culturale nel quale Gödel si trovò a fare le proprie scoperte, e rivolte al logico austriaco (per lettera, nel 1974) dall'allora studente in sociologia presso l'università del Texas Burke D. Grandjean. Rispondendo ad uno dei quesiti, Gödel indicò di aver aderito ad una concezione platonista fin dal 1925 (la lettera di Grandjean a Gödel, il questionario e le relative risposte di Gödel, sono adesso riprodotte nei volumi dei *Collected Works* dedicati alla

spiegherebbe perché Gödel abbia apparentemente abbandonato i riferimenti all'esistenza di proposizioni assolutamente indecidibili. Il problema in questione, se di problema si tratta, riguarda infatti specificamente la teoria della conoscenza: Gödel passa ad ammettere in sostanza, attraverso il riferimento a principi sconosciuti allo stato attuale delle conoscenze, l'idea che la conoscenza matematica possa essere sempre estesa fino a realizzare, magari, proprio il convincimento di Hilbert sull' 'onniscienza' in campo matematico che in [Göd38c] veniva messo in pericolo dalla specifica forma di irresolubilità di certe proposizioni legate all'Ipotesi del Continuo.

L'idea che Gödel abbia con maggior decisione aderito alla posizione secondo la quale non esistono problemi matematici di principio irresolubili, rende ancora più rilevante, ci pare, il *vulnus* creato dalla mancanza di una trattazione sistematica dell'intuizione matematica nei testi di Gödel fin qui considerati.

E, da un altro punto di vista, rappresenta un'ottima introduzione per la prossima sezione, nella quale ci proponiamo di analizzare quei luoghi dell'opera gödeliana in cui l'esistenza di proposizioni indecidibili viene misurata in relazione alla concezione di Gödel della mente.

2.4 La mente, le macchine e la natura della matematica

L'argomento 'insiemistico' al quale Gödel ricorse nella cosiddetta '*Gibbs Lecture*', l'intervento alla Brown University del dicembre 1951 di cui si è detto nel paragrafo precedente, era stato concepito per rendere evidente il fenomeno dell' 'inesauribilità della matematica'. Gödel ben sapeva che, presupponendo esso la riducibilità dell'intera matematica alla teoria degli insiemi, come egli notò in quell'occasione, non si trattava di un'osservazione neutrale da un punto di vista fondazionale, ma piuttosto di un'argomentazione compromessa con un punto di vista, una sorta di realismo di tipo pragmatico, che guarda alla matematica "senza che essa venga decurtata in ragione di qualche rilievo critico".

---

corrispondenza - cfr. [Göd03a], pp. 442-450). È pur vero che l'attendibilità di questa fonte può essere contestata: perché si tratta di un documento estremamente frammentario e appartenente ad un periodo delicato per ciò che concerne la salute di Gödel. Diverso il discorso per le lettere a Wang di cui sopra (che però non contengono un'indicazione cronologica altrettanto precisa per l'adesione al platonismo).

Durante lo stesso intervento, egli volle far notare tuttavia come, riferendosi più direttamente ai teoremi di incompletezza, divenisse possibile giungere alla stessa conclusione, e persino a qualcosa di più, in modo del tutto indipendente "dal punto di vista scelto" in materia dei fondamenti della matematica.

In questa seconda versione, l'argomento di Gödel in favore dell'inesauribilità della matematica, prende in considerazione le implicazioni che a tale proposito hanno *entrambi* i teoremi da lui scoperti.

Da un lato, infatti, riformulando appropriatamente il primo dei due teoremi di incompletezza in un modo che ci è ormai divenuto familiare, si ha che "*comunque si scelga un sistema d'assiomi e regole di inferenza ben definito, esistono sempre problemi diofantei* [della forma nota] *che sono indecidibili mediante gli assiomi e le regole in questione, supposto che nessuna proposizione falsa di quel tipo sia derivabile.*"[252]. Da ciò discende, esattamente come nel caso dell'argomento basato sull'analisi del concetto iterativo di insieme, che la matematica è incompletabile nel senso che nessun sistema formale particolare è in grado di contenerla tutta (neanche confinandosi al dominio delle proposizioni elementari sui numeri).

Se ad un'analisi simile si sottopone il secondo teorema di incompletezza, dal quale discende che la proposizione che esprime la consistenza di un sistema d'assiomi non è dimostrabile nel sistema in questione, si ricava, secondo Gödel una conclusione persino più chiara:

> È *questo* teorema che rende l'incompletabilità della matematica particolarmente evidente. Perché, *rende impossibile che chiunque mettendo insieme un sistema d'assiomi e regole ben definito, faccia la seguente affermazione su esso in modo consistente: Percepisco* (*con certezza matematica*) *gli assiomi e le regole in questione come corretti e credo inoltre che essi contengano l'intera matematica.*[253]

L'osservazione di Gödel viene sostenuta attraverso una disamina particolarmente attenta, che merita di essere ripercorsa passo passo.

In primo luogo, la conclusione segue dal fatto che "chiunque sottoscrivesse una simile affermazione, contraddirebbe se stesso", come risulta dal seguente argomento: se fosse vero che egli 'percepisce con certezza matematica' il sistema d'assiomi in questione come una collezione di proposizioni matematiche vere e di regole di inferenza corrette (che trasmettono, cioè la

[252][Göd51], p. 308.
[253]*Ibid.*, p. 309.

verità dalle premesse alla conclusione), allora egli "percepirebbe (con lo stesso grado di certezza) anche che [il sistema è] consistente", cioè che nessuna proposizione falsa può essere formalmente derivata a partire da quegli assiomi; tuttavia, poiché il secondo teorema di incompletezza afferma che la (proposizione che esprime la) consistenza non è un teorema del sistema in questione, "egli possiede un'intuizione matematica che non è dimostrabile dai suoi assiomi"[254].

In altre parole: c'è una proposizione vera, un proposizione cioè che è 'percepita con certezza matematica', che si situa al di là di qualsiasi sistema formale che si suppone contenere tutta la matematica.

È lo stesso Gödel, tuttavia, a sottolineare che nel trarre le conclusioni del precedente argomento, occorre procedere con estrema cautela:

> Ciò significa quindi che nessun ben definito sistema di assiomi corretti può contenere tutta la matematica propriamente detta? È così, se per matematica propriamente detta si intende il sistema delle proposizioni vere; non è così, tuttavia, se per matematica propriamente detta si intende il sistema di tutte le proposizioni dimostrabili.[255]

Come si era anticipato alla nota 248, la 'matematica propriamente detta', così come viene definita da Gödel all'inizio dell'intervento, è la collezione delle proposizioni matematiche che "valgono in senso assoluto", ossia, per usare la terminologia di questa parte della conferenza, che "vengono percepite con certezza matematica".

L'osservazione precedente lascia intendere come il significato dell'argomento dipenda in realtà dall'interpretazione della nozione chiave: perché se, come fa Gödel, distinguiamo una 'matematica in senso oggettivo' (la collezione delle proposizioni matematiche vere) da una 'matematica in senso soggettivo' (l'insieme delle proposizioni matematiche dimostrabili), si raggiungono conclusioni opposte:

> Evidentemente nessun sistema ben definito di assiomi corretti può contenere tutta la matematica in senso oggettivo, dal momento che la proposizione che ne esprime la consistenza è vera, ma non dimostrabile nel sistema. Tuttavia, rispetto alla matematica in senso soggettivo, non è precluso che esista una procedura finita dalla quale discendano tutti gli assiomi evidenti per essa.[256]

[254] *Ibid.*, p. 309.
[255] *Ibid.*, p. 309.
[256] *Ibid.*, p. 309.

È vero, quindi, che il fenomeno dell'incompletabilità della matematica, nel senso della non assiomatizzabilità formale, riguarda la matematica in senso oggettivo: esso non si applica cioè nel caso in cui si considerassero come valide, o 'percepite' con certezza assoluta, solo e soltanto le proposizioni per le quali si dispone di una dimostrazione matematica. Per esse, si potrebbe dunque immaginarne la coestensività con l'insieme dei teoremi di un sistema formale, e quindi l'analogia delle capacità deduttive del matematico con quelle di una macchina di Turing: quello che segue, in questo caso, dal secondo teorema di incompletezza, è solo il fatto che la consistenza del sistema d'assiomi, ovverosia, la correttezza della macchina (rispetto ad una certa classe di enunciati) non può essere accertata mediante strumenti matematici.

In altre parole, dice Gödel, "la mente umana (nel campo della matematica pura) *sarebbe* equivalente ad una macchina finita la quale, tuttavia, è incapace di comprendere completamente il proprio funzionamento"[257].

Resta il fatto, ci tiene a sottolineare tuttavia Gödel, che la 'matematica in senso oggettivo', esattamente come nel caso dell'argomento insiemistico, risulta essere incompletabile nel senso di non poter essere racchiusa entro un unico sistema d'assiomi. E c'è di più: proprio nel caso estremo in cui le capacità deduttive della mente umana fossero equivalenti a quelle di una macchina di Turing, questa stessa incompletabilità della matematica in senso oggettivo, diventerebbe "particolarmente impressionante" dal momento che "esisterebbero problemi diofantei del tipo descritto *assolutamente* indecidibili, dove l'epiteto 'assolutamente' significa che essi sarebbero indecidibili, non solo rispetto ad un particolare sistema assiomatico, bensì rispetto ad *ogni* possibile dimostrazione matematica che la mente umana è in grado di concepire"[258].

Dal momento che non si è in grado di escludere logicamente la possibilità che la ragione umana, per ciò che concerne la conoscenza matematica, sia equivalente ad una macchina di Turing mediante i teoremi di incompletezza, l'esistenza di proposizioni indecidibili in senso assoluto (già a livello, si noti, della teoria dei numeri), sarebbe una conseguenza diretta dell'indecidibilità algoritmica della matematica (intesa come la collezione delle proposizioni oggettivamente vere).

Il risultato di quest'analisi, consiste pertanto per Gödel nella seguente, inevitabile conclusione:

[257] *Ibid.*, pp. 309-310.
[258] *Ibid.*, p. 310.

> *O la matematica è incompletabile in questo senso, cioè i suoi principi evidenti non possono essere compresi in una regola finita, che equivale a dire che la mente umana* (*anche nel campo della matematica pura*) *sorpassa infinitamente le capacità di ogni macchina finita, oppure diversamente* [la mente umana è equivalente a una macchina finita e] *ci sono problemi diofantei del tipo specificato assolutamente indecidibili.*[259]

La distinzione tra la matematica 'in senso oggettivo' e la matematica 'in senso soggettivo' riflette le due componenti del discorso di Gödel, quella ontologica e quella epistemica, che sono venute delineandosi nei precedenti paragrafi. L'argomento che egli elabora, dunque, può essere visto come l'analisi delle implicazioni su entrambe le componenti dei teoremi di incompletezza: si ricava da essi la necessità logica dell'inesauribilità della matematica, intesa come collezione di proposizioni vere in senso assoluto (descrizione di fatti), e la possibilità logica della natura computazionale della mente, e quindi della natura algoritmica di quella porzione della prima che è individuata in relazione alle specifiche capacità deduttive del matematico umano.

Prima di esporre i dettagli delle conseguenze filosofiche che per Gödel possiede questo fatto, "stabilito per via matematica", vogliamo soffermarci sul problema di come quest'argomento di Gödel si inserisce nella discussione più generale del problema della natura della mente umana, che abbiamo visto emergere, anche se solo in forma allusiva, da più parti nell'opera gödeliana.

2.4.1 La mente secondo Gödel

Il testo della *Gibbs Lecture* non fu mai pubblicato da Gödel, nonostante egli avesse pensato ad inviare una revisione del manoscritto ritrovato tra le carte del *Nachlass* al *Bullettin of the American Mathematical Society*[260]. Tuttavia, su sollecitazione di Hao Wang, Gödel acconsentì a pubblicare la propria opinione riguardo alle conclusioni della *Gibbs Lecture* nel paragrafo conclusivo del capitolo X di [Wan74], intitolato "Menti e Macchine".

Il commento di Gödel di quelli che vengono da lui indicati come "i due risultati più interessanti sulle menti e le macchine dimostrati in modo rigoroso", e cioè il fatto che, come si ottiene riformulando le conclusioni relative all'inesauribilità della matematica, non può esistere un formalismo corretto per "l'intuizione matematica", e la disgiunzione con la quale abbiamo chiuso

[259] *Ibid.*, p. 310.

[260] Questo fatto, citato da Boolos nella nota introduttiva a [Göd51] (cfr. [Göd95], p. 290) risulterebbe da alcune lettere inedite.

il paragrafo precedente, è duplice. Da un lato, come osservazione specifica su quest'ultima si legge:

> Gödel pensa che Hilbert avesse ragione a rifiutare la seconda alternativa [cioè, l'equivalenza tra la mente umana e una macchina di Turing e l'esistenza di problemi diofantei assolutamente irresolubili]. Se questa fosse vera, ciò significherebbe che la ragione umana è assolutamente irrazionale per il fatto di sollevare quesiti ai quali non è in grado di rispondere, asserendo allo stesso tempo in modo enfatico che solo la ragione è in grado di darvi soluzione. La ragione umana sarebbe dunque imperfetta e, in un certo senso, persino inconsistente, in palese contraddizione con il fatto che quelle parti della matematica che sono state sviluppate in modo sistematico e completo (ad esempio, la teoria delle equazioni Diofantee di primo e secondo grado, queste ultime con due incognite) mostrano un grado di bellezza e perfezione straordinario.[261]

Non ci sono solo ragioni estrinseche tuttavia, sulla base delle quali si può considerare implausibile il fatto che la mente umana risulti equivalente ad una macchina di Turing, o che, volendo sottolineare l'altro aspetto del problema, esistano proposizioni aritmetiche assolutamente indecidibili: "Gödel nota anche che i tentativi di dimostrazione dell'equivalenza tra menti e macchine sono fallaci"[262].

Il riferimento, è ad una formulazione alternativa, con cambiamenti trascurabili, dell'ultima delle tre osservazioni sui risultati di indecidibilità del 1972[263], e che nel testo pubblicato in [Göd90a] (p. 306), viene presentata con il titolo "Un errore filosofico nel lavoro di Turing". Si tratta infatti di un esame dell'argomento contenuto nel paragrafo 9 di [Tur36] da noi discusso nel §1.2, e che, agli occhi di Gödel, contiene "un argomento che si suppone mostri come le procedure mentali non possano andare al di là delle procedure meccaniche".

Il fondamento della critica gödeliana ha origine dal fatto che l'argomento di Turing, come risulta dalla versione della critica di Gödel contenuta in [Wan74] ma non in quella in [Göd90a], "dipende dall'assunzione che una mente finita può avere solo un numero finito di stati distinguibili"[264]. Da entrambe le formulazioni, appare chiaro che Gödel considerasse quest'ipotesi indebita.

Nella versione del 1972, si legge ad esempio:

---

[261] [Wan74], p. 325.
[262] *Ibid.*, p. 325.
[263] [Göd90c].
[264] [Wan74], p. 325.

> Ciò di cui Turing non tiene affatto conto è che *la mente, nel suo uso, non è statica bensì costantemente in evoluzione*, ossia che si comprendono i termini astratti in modo sempre più preciso, mano a mano che si procede nell'utilizzarli, e che un numero sempre maggiore di termini astratti entrano nella sfera della nostra comprensione. Ci potrebbero essere metodi sistematici per rendere attuale quest'evoluzione, che potrebbe così entrare a far parte della procedura. Dunque, per quanto ad ogni stadio il numero e la precisione dei termini astratti a nostra disposizione possa essere finito (e di conseguenza anche il numero degli *stati mentali distinguibili* di Turing), potrebbe *convergere all'infinito* nel corso dell'applicazione della procedura. Si noti che qualcosa del genere pare realizzarsi nel processo di costruzione di assiomi dell'infinito sempre più forti in teoria degli insiemi.[265]

Il passo riportato in [Wan74], non differisce su questo in modo sostanziale, e quindi non contribuisce ad eliminare del tutto un certo alone di esotericità che circonda l'osservazione di Gödel.

Facciamo notare, *en passant*, che questa lettura da parte di Gödel del lavoro di Turing del 1937 ha creato un problema di interpretazione: alcuni[266] ritengono che Gödel abbia semplicenmente frainteso le intenzioni dell'analisi di Turing, la quale non è volta a mostrare come i processi mentali possano essere ridotti a procedure di carattere meccanico, bensì, in linea con quanto si è detto nel §1.2, solo che quei processi mentali *a carattere computazionale*, come l'atto del calcolare secondo regole, abbiano una natura effettiva; altri[267], non sono di questo avviso e chiamano in causa il diverso significato attribuito dai due all'effettiva calcolabilità (così come essa emerge dal passo citato in precedenza)[268].

Ora, a proposito del passo di Turing chiamato in causa, occorre notare due cose. Primo: la ragione per la quale il numero degli "stati mentali" viene considerato finito ai fini dell'esposizione, è l'osservazione che se così non fosse, "alcuni di essi risulterebbero 'arbitrariamente vicini' e verrebbero confusi"[269].

---

[265] [Göd90a], p. 306.

[266] Tra i quali Wilfried Sieg (si veda, ad esempio [MS95], p. 14, [SB96], p. 103).

[267] [Pic03], p. 42, nota 19.

[268] Gödel attacca in particolare l'assunzione, che è invece centrale per l'argomento fornito da Turing nel suo [Tur36], che il numero di stati mentali sia finito. La posizione di Gödel potrebbe allora essere quella, *pur rimanendo nell'ambito dei processi razionali di calcolo*, di disconoscere che questi soddisfino l'ipotesi di finitezza assunta da Turing, e che quindi il calcolo di una funzione numerica da parte di un 'calcolatore' umano soddisfi sempre le ipotesi **(D)**, **(F)** e **(L)** di cui si è detto al §1.2 (il contrasto emergerebbe, in questo caso, sulla correttezza del passo (1) della figura 1.1 del §1.2).

[269] [Tur36], p. 136.

Per ammissione dello stesso Turing, questa restrizione ha una motivazione analoga a quella sul numero di simboli occorrenti ai fini del computo, ed è fondata, come si è visto nel §1.2, sulla necessità che le configurazioni simboliche siano 'riconoscibili in modo immediato'. Alla base di ciò, stanno quindi, da un lato, le limitazioni sensoriali del 'calcolatore umano', dall'altro la necessità, perché l'argomento sia stringente, di individuare i passi elementari su cui si basa il computo (cioè di poter assumere la riconoscibilità di una configurazione simbolica come un passo che non è ulteriormente analizzabile).

A ciò va certamente aggiunto, in secondo luogo, che Turing non attribuisce alcuna 'valenza cognitiva' al termine "stato mentale", al punto di paventare la possibilità di sostituirlo con una "controparte più fisica"[270].

Dunque, per tornare al problema di interpretazione di cui si diceva poc'anzi, è certamente possibile che Gödel attribuisca all'argomento di Turing una valenza 'mentalista' che il matematico britannico non solo non aveva intenzione di riconoscergli, ma che volesse addirittura consapevolmente evitare. D'altra parte, è pur vero che il passo fa emergere alcuni aspetti caratteristici del pensiero di Gödel, e dunque paventa una drastica differenza di merito, relativa a come, da un punto di vista cognitivo, la conoscenza (e principalmente quella matematica) si espleti.

Sotto quest'ultimo riguardo, in particolare, non sembra cosa facile, come si diceva, penetrare il significato delle parole di Gödel in modo esauriente. La 'procedura mentale' che, rovesciando la critica a Turing, 'può andare oltre' i processi effettivi di tipo meccanico, viene caratterizzata mediante due elementi: la comprensione degli enti astratti che procede costantemente verso una sempre maggiore chiarezza, e l'intervento di enti astratti del tutto nuovi nella sfera della comprensione.

Il riferimento agli assiomi dell'infinito si chiarisce se letto attraverso l'argomento insiemistico della *Gibbs Lecture*, cioè se si guarda agli assiomi dei grandi cardinali, come Gödel faceva in quel caso, come a principi che, scaturendo in modo naturale dall'analisi del significato stesso della nozione iterativa di insieme, rappresentano un esempio soprattutto di come, sulla base di una migliore comprensione dei concetti astratti, la 'procedura mentale' abbia luogo.

Riguardo all'idea di poter rendere sistematica la procedura in questione ed il ricorso ad essa, cioè la possibilità, come si legge nel passo corrispondente in [Wan74], di "accelerare, specializzare e determinare in modo univoco" il

[270] *Ibid.*, p. 136.

processo di evoluzione della mente, Gödel sottolinea come "la definizione precisa di una procedura di questo tipo richiederebbe un approfondimento sostanziale della nostra comprensione delle operazioni della mente"[271]; e dice anche (stavolta in [Göd90a]) che il risultato di una simile analisi sarebbe "una funzione aritmetica non-ricorsiva"[272].

Ci limitiamo ad osservare come, in particolare quest'ultimo accenno, richiami alla mente l'ipotesi, avanzata da Gödel alla conferenza di Princeton nel 1946, di giungere ad un concetto assoluto di dimostrabilità in matematica, per mezzo di una qualche definizione "non costruttiva" del processo di estensione di un sistema d'assiomi, mediante quei principi che nascono dalla "contemplazione" di quelli assunti[273].

I due accenni sembrano nascere da uno spunto comune: essendo la mente costitutivamente non computazionale, ed esprimendosi questo carattere nell'elaborazione di quei principi matematici che si riferiscono a concetti astratti, la comprensione di come si instaurino le relazioni che rendono possibile quest'elaborazione produrrebbe una descrizione del fondamento, essenzialmente non costruttivo, delle capacità deduttive della ragione umana.

Altrettanto interessanti sono le osservazioni di Gödel che Wang riporta come chiosa alla critica dell'argomento di Turing:

> L'argomento di Turing diviene valido sotto due assunzioni ulteriori, che sono generalmente accettate oggigiorno, ossia: 1. Non esiste mente separata dalla materia. 2. Il cervello funziona essenzialmente come un computer digitale. (2 potrebbe essere sostituita da: 2′ Le leggi fisiche, nelle loro conseguenze osservabili, hanno un limite finito di precisione). Tuttavia, mentre Gödel ritiene che 2 sia molto probabile e 2′ praticamente certa, egli crede che che 1 sia un pregiudizio della nostra epoca, che verrà scientificamente confutato (forse per dal fatto che non ci sono sufficienti cellule nervose per portare a termine le operazioni osservabili della mente).[274]

Dunque, l'avversione di Gödel per l'argomento di Turing, dietro il quale egli riconosceva l'intento di fornire un sostegno rigoroso all'interpretazione computazionale della mente, aveva origine da una posizione anti-meccanicista.

Il punto di vista di Gödel era in particolare basato su due capisaldo: la presenza di una componente effettivamente computazionale da un lato, il

[271] [Wan74], p. 325.
[272] [Göd90a], p. 306.
[273] Si veda la discussione del passo in questione nel §2.3.2.
[274] [Wan74], p. 326.

cervello, ma la convinzione, dall'altro, che la mente non fosse interamente riducibile alla sola attività delle sue parti. È chiaro che l'analisi del ragionamento matematico in termini di applicazione di una procedura effettiva ma non meccanica, basata sulla comprensione dei concetti astratti e sulla quale si fonda l'evoluzione costante della mente, è l'espressione di questa seconda convinzione (oppure, viceversa, questa seconda convinzione è generata da una simile analisi).

Per completare il quadro, occorre tenere presente il fatto che Gödel era apparentemente disposto, proprio in relazione alla conoscenza matematica, a spingere molto in là le proprie convinzioni sulla componente 'materiale' della ragione umana.

Sempre nel testo di Wang, ma in una sezione dedicata alla caratterizzazione dei concetti generali in matematica, si legge infatti:

> Gödel ipotizza che sia necessario qualche organo sensoriale per rendere possibile la trattazione delle impressioni astratte (in opposizione alle impressioni sensoriali), dal momento che mostriamo qualche debolezza nel trattare le impressioni astratte alla quale si pone rimedio guardando ad esse in relazione alle o in occasione delle impressioni sensoriali. Un tale organo sensoriale deve essere strettamente legato al centro neurale per il linguaggio.[275]

In questo caso, dunque, quell'analogia tra percezione sensibile e intuizione matematica che abbiamo ritrovato altrove nell'opera gödeliana, prosegue e si approfondisce.

Accanto agli organi di senso, infatti, Gödel ipotizza un apparato sensoriale preposto alla percezione degli enti astratti e che è quindi all'origine della 'percezione con certezza matematica'. Tuttavia, come si è visto, ciò non significa che per Gödel la conoscenza matematica, così come la conoscenza in generale, sia interamente riconducibile ad una percezione sensoriale e quindi, per vedere il problema da un punto di vista più vicino all'impostazione della presente sezione, che la mente sia riducibile al solo cervello.

Interviene dunque qualcosa di immateriale, ed interviene in particolare nella percezione degli enti astratti, la quale appare in qualche modo connessa all'analisi e alla comprensione del loro 'significato' (così come l'organo sensoriale che la governa, è connesso all'area cerebrale deputata all'attività linguistica).

[275] *Ibid.*, p. 85.

2.4.2 Contro il materialismo

Prima di tornare alla *Gibbs Lecture* e all'analisi gödeliana delle implicazioni dei teoremi di incompletezza, è forse opportuno ricapitolare alcuni punti essenziali.

Esaminando la questione dell'indecidibilità alla luce delle varie acquisizioni, tecniche e concettuali, avvenute nel corso degli anni '30 del '900, Gödel (in [Göd38c]) ricava il seguente risultato:

- La teoria delle proposizioni diofantee di tipo Π_2^0 costituisce un limite per la meccanizzazione della matematica, ovvero per le potenzialità matematiche delle macchine di Turing.

Da questo, e dall'analisi rigorosa dei teoremi di incompletezza di [Göd51] che si è visto poc'anzi, segue che:

- Per quanto sia logicamente possibile che la mente umana equivalga ad una macchina di Turing, se ciò fosse il caso si avrebbe che: 1. la correttezza del ragionamento matematico non potrebbe essere stabilita mediante una dimostrazione matematica, e 2. la stessa porzione della matematica (problemi diofantei di tipo Π_2^0) costituirebbe un limite assolutamente invalicabile per le potenzialità in campo matematico della ragione umana.

La situazione è riassunta dalle tre alternative che discendono dalla conclusione disgiuntiva di [Göd51]:

1. La mente umana sorpassa infinitamente ogni macchina di Turing (ossia, in altri termini, la matematica soggettiva coincide, almeno per ciò che concerne l'aritmetica, con quella oggettiva) e non esistono problemi matematici diofantei (del tipo isolato da Gödel) indecidibili in modo assoluto.

2. La mente umana equivale ad una macchina di Turing e alcune tra le proposizioni formalmente indecidibili sono anche tali in modo assoluto (per principio non possono essere cioè decise per mezzo di nessuna prova che la mente umana è in grado di elaborare).

3. La mente umana sorpassa infinitamente ogni macchina di Turing e alcune tra le proposizioni formalmente indecidibili sono tali in modo assoluto.

Molti luoghi dell'opera gödeliana possono essere visti come l'esame di quelle ragioni che rendono implausibile almeno la seconda conclusione: considerazioni di carattere 'estetico' o legate agli sviluppi della disciplina (la bellezza e la precisione citate, come si è visto, in [Wan74], di quelle parti della teoria delle equazioni diofantee sviluppate in modo completo); suggestioni di tipo specificamente filosofico-epistemico (è improbabile che la ragione ponga problemi che non è in grado di risolvere); convinzioni di tipo cognitivo (parte della ragione umana ha un carattere immateriale, dunque la mente umana è inerentemente, cioè qualitativamente diversa da una macchina).

Questo fatto viene riassunto da Gödel nella conferenza del 1951, in quella che costituisce, in contrapposizione con la precedente 'parte matematica', la 'parte filosofica' dell'intervento, ove egli sottolinea come entrambe le alternative (e dunque anche la terza, che deriva da una combinazione delle due) "sono decisamente in opposizione alla filosofia materialistica"[276].

Cosa Gödel intenda con quest'affermazione, viene spiegato nel modo seguente:

> Infatti, se vale la prima alternativa, ciò pare implicare che l'operare della mente umana non può essere ridotto all'operare del cervello, che ad ogni apparenza è una macchina finita con un numero finito di parti, ossia i neuroni e le rispettive connessioni. Quindi si è apparentemente spinti ad adottare qualche punto di vista di tipo vitalistico. D'altro canto, la seconda alternativa, nella quale esistono proposizioni matematiche assolutamente indecidibili, sembra confutare la posizione secondo la quale la matematica è solo una nostra creazione; perché il creatore conosce necessariamente tutte le proprietà delle proprie creature dal momento che esse non possono averne altre eccetto quelle che egli ha dato loro. Questa alternativa sembra quindi implicare che gli oggetti ed i fatti matematici (o almeno *qualcosa* in essi) possiedono un'esistenza oggettiva ed indipendente dai nostri atti mentali e dalle nostre decisioni, ovverosia, qualche forma di Platonismo o "realismo" rispetto agli oggetti della matematica.[277]

Alla luce di quanto si è detto nel ricapitolare le conclusioni raggiunte, la situazione che si viene così a determinare è davvero curiosa: sappiamo dai commenti contenuti nel testo di Wang, che Gödel era convinto del fatto che "Hilbert avesse ragione" a ritenere che non esistono proposizioni assolutamente indecidibili (sicuramente non in campo dell'aritmetica, e presumibilmente, come abbiamo ipotizzato in coincidenza col venir meno delle ipotesi relative all'esistenza di proposizioni assolutamente indecidibili nella teoria

[276] [Göd51], p. 311.
[277] *Ibid.*, pp. 311-312.

degli insiemi, anche a livelli superiori); da quanto Gödel dice in questo passo, quell' 'ottimismo razionalista' che egli era disposto conseguentemente a sottoscrivere, si lega in modo naturale al punto di vista secondo il quale gli enti matematici sono il prodotto di un atto creativo (confermando così quanto si è già detto relativamente all'indipendenza tra un'assunzione ontologica di tipo realista rispetto agli enti matematici e la convinzione epistemica sulla resolubilità di ogni questione ad essi relativa); questo legame permette quindi retoricamente a Gödel di argomentare a favore di una qualche forma di platonismo e contro ogni forma di costruttivismo (il che significa, in ultima analisi, che egli utilizza la sua posizione epistemica a sostegno della propria visione ontologica).

Quest'argomento, così come emerge dalla conferenza del 1951, si compone di due parti.

In positivo, Gödel sottolinea come il platonismo in matematica abbia trovato conferme indipendenti dalla 'soluzione' della conclusione raggiunta, rispetto alle quali, cioè, è irrilevante quale delle alternative valga. Un fatto che agli occhi di Gödel emerge chiaramente considerando che: (i), nonostante il "grado insormontabile di esattezza" raggiunto dagli sviluppi nel campo dei fondamenti della matematica, nessun aiuto è arrivato da ciò per la soluzione dei problemi matematici, a riprova del fatto che, secondo Gödel, le difficoltà incontrate non sono dovute alla "mancanza di una consapevolezza chiara rispetto a ciò che si è creato", bensì a qualcosa di più profondo; (ii) "l'attività del matematico mostra ben poca di quella libertà di cui un creatore dovrebbe godere", come dimostra il fatto che il matematico non sia nella posizione di "generare la validità dei teoremi" relativi a quegli enti che sarebbero il frutto della propria attività immaginativa; infine, (iii), a dispetto del fatto che ogni ente matematico avrebbe origine in un atto creativo distinto, e quindi questo in particolare dovrebbe valere per il concetto di insieme e quello di numero, "al fine di dimostrare certe proposizioni sugli interi, risulta necessario il concetto di insieme", dal che discenderebbe che "per scoprire quali proprietà *noi* abbiamo dato a certi oggetti, siamo costretti a crearne degli altri"[278].

Per ciò che riguarda invece la *pars destruens*, la critica al costruttivismo, questa viene riassunta da Gödel nella "confutazione" della "più precisa, e allo stesso tempo la più radicale, formulazione che è stata data finora" di quella posizione.

[278] *Ibid.*, p. 314.

Si tratta cioè del punto di vista secondo il quale:

1. la matematica è il frutto di convenzioni linguistiche, ovverosia, di regole puramente sintattiche (le quali non hanno alcun contenuto di tipo fattuale);

2. le proposizioni matematiche, ed in particolare i teoremi, sono delle tautologie (e quindi sono prive di contenuto), cioè sono vere esclusivamente in virtù delle definizioni dei termini (che, in accordo a 1, hanno un'origine meramente convenzionale).

La confutazione di cui parla Gödel, altro non è che il frutto dei teoremi di incompletezza (ed è strettamente legata alla riflessioni di Gödel sul significato del risultato per il finitismo hilbertiano[279]). Da essi discende infatti che nessuna dimostrazione di consistenza per un simile insieme di regole, che non è altro se non una prova del loro carattere tautologico[280], può fare a meno del concorso di certi enti astratti, dove ciò indica "concetti che non si riferiscono a oggetti sensibili"[281] e che quindi non hanno di certo una natura sintattica[282].

Come conseguenza di ciò che egli legge come una "confutazione" della forma più radicale dell'idea che gli enti matematici siano il frutto di un atto creativo, Gödel può così esporre le proprie convinzioni, a cominciare da quella

---

[279] D'altra parte è lo stesso Gödel (cfr. *Ibid.*, p. 315, nota 23) a sottolineare come esista una relazione precisa (una sostanziale equivalenza) tra la verifica di un simile punto di vista convenzionalista, o nominalista, e la realizzazione del programma di Hilbert.

[280] Una giustificazione per questa equivalenza la si ritrova proprio alla nota 23 di p. 315. Banalmente infatti, la tautologicità degli assiomi e delle regole (cioè il fatto che da essi non discende la verità o la falsità di alcuna proposizione fattuale) implica la consistenza, considerando che "un'inconsistenza (nella logica classica che è in questo caso sotto esame) implicherebbe ogni proposizione fattuale". Viceversa, così come il programma formalista riteneva che la consistenza fosse un criterio essenziale per la riduzione della matematica a regole di tipo puramente sintattico, "la dimostrazione di consistenza fornisce la sicurezza che queste regole sono vuote di contenuto fin tanto che esse non implicano alcuna proposizione di tipo fattuale".

[281] *Ibid.*, p. 318.

[282] L'idea cruciale di questa confutazione è al centro delle varie versioni di un articolo di Gödel dedicato alla logica di Carnap, sul quale Gödel lavorò dal 1953 al 1959 inutilmente (senza cioè raggiungere una formulazione che lo soddisfacesse e lo inducesse ad accettare l'offerta di P. A. Schilpp di pubblicare l'articolo nel volume dedicato a Carnap della serie *The Library of Living Philosophers*). Alcune versioni del lavoro sono pubblicate nel terzo volume dei *Collected Works* (cfr. [Göd95], pp. 334-362).

secondo la quale la realtà matematica sia del tutto indipendente dalla ragione umana e dalle sue inclinazioni:

> Mi pare tuttavia che nonostante tutto, un ingrediente di questa teoria sbagliata della verità matematica sia perfettamente corretto e chiarisca completamente la vera natura della matematica. È infatti corretto che una proposizione matematica non dice nulla riguardo alla realtà fisica o psichica esistente nello spazio e nel tempo, perchè è vera già in virtù dei termini che occorrono in essa, indipendentemente dal mondo delle cose reali. Quello che è sbagliato, tuttavia, è il fatto che il significato dei termini (cioè i concetti che essi denotano) sia considerato come qualcosa costituito dall'uomo e consistente in mere convenzioni semantiche. La verità, credo è che i concetti formano una realtà oggettiva a sé stante, che noi non possiamo creare o cambiare ma solo percepire e descrivere.[283]

Il passo chiarisce anche, quindi, l'avversione di Gödel nei confronti di due varianti della posizione realista, e cioè quella 'empirista' (gli enti matematici sono in un qualche senso parte della realtà sensibile) e quella 'psicologista' (gli enti matematici trovano posto ed hanno un'effettiva esistenza tra i fenomeni psichici), e delinea in modo chiaro l'elemento 'platonista' del realismo gödeliano: il mondo della matematica è separato sia dalla realtà fisica che da quella psichica.

Il confronto con la posizione nominalistico-convenzionalista si sposta, di conseguenza, nel campo della natura delle proposizioni che compongono il 'discorso' della matematica e del problema del significato dei termini che vi occorrono.

Gödel definisce a tal fine le "proposizioni analitiche", come proposizioni vere "in relazione ai concetti che occorrono in esse", e distingue queste ultime dalle proposizioni tautologiche (che pure in un certo senso sono analitiche, ma solo in quanto "vere in relazione alle definizioni" dei termini). Le proposizioni matematiche, che egli identifica con il primo tipo, sono sì vere in funzione del "significato" dei termini che vi occorrono, come anche altrove nell'opera gödeliana si può leggere, ma non nel senso che dipendono da certe espressioni del linguaggio (che, dopo tutto, possono pur avere una natura convenzionale), bensì dai concetti che ne costituiscono la realizzazione.

Un fatto che risulta chiaro, e non poteva davvero mancare un simile riferimento, dagli assiomi dell'infinito della teoria degli insiemi:

> È vero che questi assiomi sono validi in virtù del significato del termine "insieme" - si potrebbe persino dire che essi esprimono il significato stesso

[283] *Ibid.*, p. 320.

> del termine "insieme" - e quindi li si potrebbero correttamente chiamare analitici; tuttavia, il termine "tautologico", cioè privo di contenuto, per essi è completamente fuori luogo, dal momento che anche l'asserzione circa l'esistenza di un concetto di insieme che soddisfa questi assiomi (o l'asserzione di consistenza di questi assiomi) è così distante dall'essere vuota che non è possibile dimostrarla senza fare ricorso nuovamente al concetto di insieme, o a qualche altro concetto astratto di natura simile.[284]

È dunque una certa persistenza dei concetti, ed il loro ruolo di nozioni primitive che corrispondono (semanticamente) a certi termini linguistici, che rende palese il carattere non tautologico delle proposizioni matematiche.

Sulla loro natura analitica, in relazione alla realtà indipendente dei concetti di cui queste trattano, Gödel aggiunge:

> Questo concetto di analiticità è così distante dal significare "privo di contenuto" che è perfettamente possibile che una proposizione analitica risulti essere indecidibile (oppure decidibile solo con una certa probabilità). Perché la nostra conoscenza del mondo dei concetti potrebbe essere altrettanto limitata e incompleta come quella del mondo delle cose. È certamente innegabile che questa conoscenza sia talvolta non solo incompleta, ma persino indistinta. Questo si verifica nel caso dei paradossi della teoria degli insiemi, che sono sovente presentati come una confutazione del Platonismo, ma, a mio parere, ingiustamente. Le nostre percezioni visive contraddicono in qualche caso quelle tattili, ad esempio nel caso di un bastone immerso nell'acqua, ma nessuno sano di mente vorrà concludere da ciò che il mondo esterno non esiste.[285]

L'importanza della critica al convenzionalismo nella *Gibbs Lecture* consiste proprio nel fatto che essa introduce in modo chiaro il tema del ruolo dell'elemento simbolico-linguistico (le proposizioni) nella conoscenza matematica. Questo fatto risulta ancora più importante se si tiene conto di quei luoghi nell'opera di Gödel, nei quali egli si riferisce alla comprensione dei concetti astratti come comprensione del loro 'significato', un uso che si è visto in quale senso lo si può chiarire. Ma la disamina da parte di Gödel della natura delle proposizioni matematiche ci permette anche di riassumere complessivamente il quadro dell'epistemologia gödeliana in modo più compiuto e secondo lo schema tradizionale del problema della conoscenza.

Vi sono tre componenenti: la mente (livello cognitivo), le proposizioni (livello linguistico o simbolico) e gli enti matematici, gli oggetti (livello ontologico). La relazione fondamentale, quella *conoscitiva*, che si stabilisce tra il

[284] *Ibid.*, p. 321.
[285] *Ibid.*, p. 321.

livello cognitivo e quello ontologico, passa attraverso il livello sintattico, avendo un passaggio fondamentale nella comprensione del significato dei termini delle proposizioni matematiche, e si fonda sul rapporto semantico tra questo ed il livello ontologico: poiché i termini *si riferiscono a* i concetti (e quindi non c'è arbitrarietà o convenzionalità linguistica dietro il loro significato), comprendere il significato *è* comprendere l'oggetto a cui esso è diretto.

È interessante a questo punto notare quali sono le peculiarità con le quali questo schema si dà in Gödel. La prima, è certamente il fatto che l'oggetto dell'atto conoscitivo è un concetto astratto (per quanto si sia visto come in più occasioni Gödel avvicini proprio da un punto di vista gnoseologico la conoscenza sensibile a quella matematica), e che deriva dall'approccio platonista.

Un altro aspetto di rilievo, che emerge dalla critica dell'argomento di Turing, è che questo schema descrive solo *in parte* la conoscenza matematica, dal momento che essa si caratterizza, oltre che per il momento di approfondimento della comprensione degli enti astratti, anche per l'intervento di *nuovi* concetti astratti nella sfera conoscitiva. Dunque, la relazione tra il livello cognitivo e quello ontologico si articola su due direzioni, e beneficia, per così dire, anche di un processo di *feedback*[286].

Ma vi è anche un terzo elemento di interesse. Il fatto che la conoscenza matematica, la conoscenza dei concetti astratti cioè, può rivelarsi solo parziale e incompleta, a causa, presumibilmente, delle limitazioni a cui è soggetto l'organo sensoriale ad essa preposto. Ciò ci consente di guardare alla questione delle proposizioni indecidibili in senso assoluto, della quale ci siamo a lungo occupati, come ad un aspetto di rottura meno drammatico di quel che poteva apparire in un primo momento. Perché se è vero che l'allusione a nuovi principi per risolvere questioni indecise rende per Gödel probabile la decisione di certe proposizioni indipendenti (in particolare, di quelle insiemistiche), è pur vero anche che ciò non esclude che talune, rivelandosi così intrinsecamente legate alle limitazioni delle nostre facoltà intellettive,

[286]È bene ribadire che, per quanto si è visto fin qui, le riflessioni di Gödel 'autorizzano' l'adozione di una tale ricostruzione schematica per la sola conoscenza *matematica*. Se una simile articolazione possa reggere, ed in quale forma, per la conoscenza in generale, è un fatto che potrebbe forse trovare risposta nell'analisi dei lavori gödeliani di fisica (ossia, [Göd49a], [Göd52], [Göd49b], [Göd49c], [Göd46a], [Göd46b]) ed in particolare quelli, tra essi (gli ultimi tre), che cercano di evidenziare certe connessioni filosofiche della teoria della relatività.

descrivano fatti del mondo dei concetti che sono destinati a rimanere a noi inaccessibili.

Si ha così che alla fine, delle tre alternative logiche che sono il frutto l'analisi rigorosa della valenza matematica dei teoremi di incompletezza, proprio la terza, se interpretata in senso più generale, rappresenta meglio, in modo forse inaspettato, il punto di vista di Gödel in tutte le sue sfumature: è vero che la mente umana sorpassa le capacità di una macchina di Turing e che quindi la classe dei problemi diofantei individuata non costituisce un limite per le potenzialità matematiche della ragione, ma è anche possibile che, a causa delle limitazioni a cui potrebbero essere soggette le nostre capacità di percepire la realtà matematica indipendente, esistano proposizioni che descrivono parti di questa realtà che risultino indecidibili anche in modo assoluto (cioè, in relazione a queste limitazioni).

2.5 Una nota conclusiva sulla filosofia di Gödel

L'opera di analisi e rielaborazione a cui si è sottoposto parte del materiale ascrivibile a vario titolo a Gödel, ci pare faccia emergere in modo evidente tratti di quella che appare una vera e propria dottrina filosofica.

Utilizzando i teoremi di incompletezza e l'analisi dei vari aspetti del loro significato come chiave di lettura, si giunge ad accumulare una serie di osservazioni ricorrenti e concordanze testuali che appaiono un indizio preciso sia, da un lato, dell'esistenza di una cornice concettuale unitaria entro la quale si inquadrano le riflessioni di Gödel, che, dall'altro, della multiformità e delle ramificazioni del punto di vista gödeliano.

Un esempio lampante, è il caso dei 'nuovi assiomi' per la matematica che, come si sarà notato, costituiscono un elemento in relazione al quale confluiscono riflessioni di vario genere, nate cioè in relazione a problematiche diverse fra loro: ad essi Gödel fa riferimento in un contesto di tipo fondazionale, in quanto costituiscono il punto di arrivo naturale di quel percorso di apertura ai concetti astratti (ovvero di estensione del punto di vista finitista) reso necessario dalla scoperta dell'esistenza di proposizioni formalmente indecidibili; a questi stessi principi Gödel si rivolge come ad un modello, tanto per indicare l'oggettività della realtà matematica quanto per spiegare la natura analitica della conoscenza matematica; infine, gli assiomi dell'infinito intervengo in modo decisivo nella disamina dell'elemento discriminante della ragione umana, di ciò che la rende *essenzialmente* diversa da un dispositivo meccanico.

Un quadro del genere, in cui alla ricchezza di tematiche (fondazionali, epistemiche, cognitive) si somma uno sfondo unitario, per quanto talvolta solo intuito e non esposto in modo diffuso e sistematico, è alla base dell'interesse delle ricerche dedicate ad una sua chiarificazione.

In un manoscritto che risale all'inizio degli anni '60 ritrovato tra le carte del *Nachlass*[287], Gödel propone un criterio sulla base del quale ordinare le varie *Weltanschauungen*, tutte le possibili 'visioni filosofiche del mondo', e grazie al quale poter descrivere parimenti in termini filosofici anche lo sviluppo dell'indagine sui fondamenti della matematica:

> Credo che il principio più fruttuoso per ottenere una visione complessiva delle possibili visioni del mondo sarà quello di suddividerle in accordo al grado e al modo della loro affinità alla o, rispettivamente, distanza dalla metafisica (o religione). In questo modo si ottiene immediatamente una divisione in due gruppi: scettiscismo, materialismo e positivismo stanno da un lato, spiritualismo, idealismo e teologia dall'altro. A un tempo, si possono rilevare anche i gradi diversi in questa successione, dato che lo scetticismo e più distante dalla teologia di quanto non lo sia il materialismo, mentre d'altro canto l'idealismo, ad esempio, nella sua versione panteistica, è una forma indebolita di teologia in senso proprio.
>
> Questo schema si dimostra utile, tuttavia, per l'analisi delle dottrine filosofiche ammissibili in contesti speciali, per il fatto che le si ordinano in questo modo oppure, nei casi incerti, se ne ricercano gli elementi materialistici o spiritualistici.[288]

Sarebbe ovviamente decisamente in accordo con lo spirito "Gödel secondo Gödel" se si riuscisse ad inquadrare gli aspetti salienti della filosofia gödeliana, almeno per ciò che riguarda lo 'speciale contesto' della filosofia della matematica, secondo lo schema qui proposto.

A proposito della filosofia di Gödel, molto si è detto, e si dice tuttora, del 'platonismo'. Molto meno, ci pare, si dice di quella componente episte-

[287] Il testo in questione, [Göd61], è stato pubblicato in [Göd95], pp. 374-387. L'attribuzione cronologica, si basa sul fatto che le pagine manoscritte siano state trovate in una busta dell'American Philosophical Society datata 13 dicembre 1961. Da qui, anche l'ipotesi che potesse trattarsi di un testo preparatorio per la confererenza dei nuovi membri (Gödel era stato eletto nell'aprile del 1961 ed aveva firmato il libro degli afferenti nel novembre dello stesso anno) che era d'uso all'interno dell'associazione. Non esiste alcuna prova del fatto che Gödel abbia accettato l'invito a tenere l'intervento in questione.

[288] [Göd61], p. 374.

mica dell'impostazione filosofica gödeliana, che nel corso del presente lavoro si è cercato di evidenziare, la quale, seppur strettamente legata alla prima componente, è anche, come si è visto, del tutto indipendente da essa, nel senso che non ne costituisce una banale conseguenza, e per tale motivo è un aspetto che appare meritorio di una menzione a parte.

Anche per questo motivo, è difficile sfuggire all'impressione che, rispetto allo schema proposto da Gödel, nel caso dell'impostazione filosofica gödeliana si sia di fronte proprio ad uno di quei casi incerti, frutto della commistione di elementi diversi fra loro.

Si prenda, in primo luogo, il realismo gödeliano rispetto agli enti matematici. Come si è visto, Gödel è convinto dell'esistenza di un 'mondo dei concetti', separato da quello degli oggetti reali, che è all'origine della sua avversione ad un realismo di tipo empirista, che collochi cioè la realtà degli enti matematici accanto, e non in contrapposizione, a quella degli oggetti sensibili. Tuttavia, la propria convinzione nell'oggettività della realtà matematica è così forte da spingere Gödel a vedere nel ricorso a procedure di tipo induttivo uno sbocco naturale di quest'assunzione, adombrando così il superamento di uno dei tabù della concezione classica, e rigorosamente 'deduttivista', della matematica, e quindi ad aprire le porte, per contrapposizione, ad un empirismo di tipo realista[289].

Un caso analogo si verifica per ciò che riguarda la concezione gödeliana della mente. Qui, l'elemento di tipo materialistico, che pure sussiste in misura considerevole come dimostra non tanto l'accostamento del cervello ad una macchina, ma soprattutto l'ipotesi di un organo sensoriale deputato alla 'percezione' matematica o, più in generale, a quella dei concetti, convive con l'idea dell'esistenza di un aspetto immateriale, caratteristico della mente umana, che si espleta proprio nell'elaborazione di nuovi principi matematici in relazione all'intuizione degli enti astratti.

Questi ed altri elementi di commistione, devono rappresentare un invito alla prudenza proprio per l'inquadramento della posizione di Gödel entro categorie filosofiche tradizionali e soprattutto in quelle tradizionalmente applicate al dibattito sui fondamenti della matematica. Non è improbabile, ad esempio, che da un'analisi più attenta di quel materiale che dà testimonianza dell'indagine gödeliana intorno al costruttivismo e alle sue varie accezioni, emergano altri aspetti di quest'approccio fondazionale che Gödel avrebbe volentieri inserito nel quadro della propria concezione (e che si aggiungerebbero

[289]Si veda al proposito, [Göd51], p. 313.

così alla convinzione che non esistono proposizioni matematiche assolutamente indecidibili, che, nella *Gibbs Lecture*, Gödel lega proprio all'idea che gli enti matematici siano frutto di un atto creativo).

Tuttavia, questi stessi aspetti che fanno emergere una certa complessità dello sfondo concettuale, rendono legittimo chiedersi quali siano i papabili precursori, se ve ne sono di riconoscibili, della prospettiva gödeliana.

Da questo punto di vista, gli scritti di Gödel che si sono analizzati nel corso del presente lavoro offrono un aiuto limitato. Ci si ricorderà, forse, di quel passo del supplemento alla seconda edizione dell'articolo dedicato all'Ipotesi del Continuo di Cantor in cui, nell'esporre l'idea relativa all'esistenza di un "dato di secondo tipo", ossia di tipo qualitativamente diverso dal dato derivante dalla sensazione, che, come è presente nel caso della conoscenza degli oggetti fisici, è all'origine dell'intuizione degli enti matematici, Gödel chiami in causa Kant per sostenere, in opposizione al filosofo di Königsberg, la natura oggettiva di questo tipo di dato.

Nel manoscritto del 1961 citato in precedenza, si assiste ad un fatto del tutto simile, laddove viene proposta una similitudine con Kant per osservare poi, immediatamente dopo, la necessità di un rovesciamento della posizione kantiana. In più, questo peculiare approccio alla filosofia kantiana viene in qualche modo reso sistematico.

Lo spunto è offerto dal riferimento a nuovi assiomi, logicamente indipendenti da quelli precedentemente assunti e dai quali dipende la soluzione di problemi irresolubili rispetto ai primi, che, sottolinea Gödel, "concorda in principio con la concezione Kantiana della matematica":

> Le affermazioni rilevanti di Kant sono, è vero, scorrette se prese in senso letterale, dal momento che Kant sostiene che nella derivazione di teoremi geometrici si ha bisogno di nuove intuizioni geometriche, e che quindi una dimostrazione [di essi] di tipo puramente logico da un numero finito di assiomi è impossibile. Questo è dimostrabilmente falso. Tuttavia, se in questa proposizione si sostituisce il termine "geometrico" con "matematico" o "insiemistico", allora essa diviene un'affermazione dimostrabilmente vera. Credo che sia un carattere generale di molte affermazioni di Kant quello di essere false se prese alla lettera ma di contenere verità profonde in un senso più generale.[290]

Secondo quanto Gödel dice, dunque, la dottrina kantiana può in certa

[290] *Ibid.*, p. 384.

misura rappresentare una sorta di specchio attraverso il quale individuare, per rovesciamento, certi tratti del punto di vista gödeliano[291].

In questo quadro di 'rilettura' delle convinzioni kantiane si inserisce anche quello che è forse l'aspetto più rilevante nella concezione gödeliana della conoscenza matematica e che, ancora una volta, ci consente di sottolineare la rilevanza di quell'elemento epistemico a cui si è più volte accennato, che emerge in stretta connessione con il riferimento ai 'nuovi assiomi' per la matematica. Se i concetti astratti, infatti, sono un costituente dell'oggetto della conoscenza in generale (in quanto, in accordo con Kant, essi intervengono anche nell'idea di oggetto dell'esperienza sensibile), e siccome questa componente è dotata di una realtà indipendente dal soggetto, l'idea di Gödel è che, mediante la scoperta di nuovi principi, si possa esaurire progressivamente proprio la conoscenza dell'elemento *oggettivo* (i concetti astratti) del *fenomeno*. È possibile riscontrare in quest'idea un'opposizione alla dottrina kantiana, ad esempio per ciò che concerne la posizione secondo la quale la componente oggettiva dell'oggetto di conoscenza, il noumeno, è inconoscibile? È possibile, se il noumeno è effettivamente pensato come ciò che sta, per essenza, oltre le capacità conoscitive del soggetto; un simile confronto è meno giustificato, se il noumeno è un'idea regolatrice della conoscenza, e dunque ne rappresenta l'elemento limite, ciò a cui essa tende (una posizione che non solo è compatibile con quella gödeliana, ma che, per molti aspetti, potrebbe essere identificata con essa).

È in relazione a questo specifico aspetto, la comprensione dei concetti astratti, che si ritrova, nel manoscritto del 1961, l'allusione a Husserl e all'approccio fenomenologico. In questo senso, occorre anche tenere presente che il richiamo alla filosofia husserliana si lega al problema di trovare un "metodo sistematico" per quella "chiarificazione del significato" che è alla base dell'approfondimento della conoscenza degli enti astratti così fondamentale, per Gödel, ai fini di un incremento delle conoscenze matematiche. Il riferimento ad Husserl, dunque, sembra essere in parte indipendente dalla questione di

[291]È possibile che a tale proposito giovi l'analisi dei contributi gödeliani di fisica, di cui si è detto alla nota 286, che aspirano ad approfondire il nesso tra certe implicazioni della teoria einsteiniana della relatività e la filosofia kantiana. In particolare, una simile disamina potrebbe far meglio apprezzare la reale conoscenza e l'effettiva comprensione, da parte del logico austriaco, dell'opera del filosofo di Königsberg (e condurre quindi ad un giudizio sulla correttezza dell'interpretazione di essa che Gödel fornisce nel contributo all'attenzione nel testo).

un'analogia di fondo tra la filosofia di Gödel e quella del filosofo tedesco[292].

Non c'è dubbio però che Gödel, vedendo alla base della comprensione dei concetti astratti un "direzionare la nostra attenzione in un certo modo, cioè sui nostri atti relativi all'uso di questi concetti, sulle nostre capacità di eseguire questi atti, ecc.", dovette scorgere, nella fenomenologia husserliana, i presupposti per dare vita ad "una procedura o una tecnica" mediante la quale raggiungere "un nuovo stato di coscienza nel quale si descrive in dettaglio i concetti alla base del nostro pensiero, o si colgono altri concetti fondamentali a noi finora sconosciuti"[293].

Soprattutto, egli vide però in Husserl e nella fenomenologia quella versione 'riveduta e corretta' della filosofia kantiana (che, dopotutto, secondo Gödel "si fonda sull'idea della fenomenologia"[294]), sulla quale versione costruire una concezione che, in modo trasversale, rispetto allo schema considerato, "eviti sia i salti mortali dell'idealismo in una nuova metafisica così come il rifiuto positivistico di ogni metafisica"[295].

In sede di valutazione complessiva, e quindi al di là del problema di isolare i tratti salienti della concezione di Gödel, ciò ci pare conduca all'idea che non c'è ingenuità nelle considerazioni di carattere filosofico presenti nell'opera gödeliana, e che quindi il suo inquadramento in uno schema generale che non sia adatto, come almeno nell'intento vuole essere quello da lui stesso proposto, alla trattazione della miscela di suggestioni molto diverse tra loro, svilirebbe ingiustamente la portata della *Weltanschauung* del logico austriaco[296].

---

[292]Un'indubbia eco di tematiche husserliane nei testi di Gödel (riferimento diretto al filosofo tedesco citato nel testo a parte), ha stimolato molte speculazioni su possibili relazioni strette tra la filosofia della matematica gödeliana e l'approccio fenomenologico. Due contributi recenti sul tema sono [Tie98] e [vHK03]. Tuttavia, a nostro avviso, i suddetti lavori si preoccupano poco di ricostruire, prim'ancora della connessione suddetta, l'articolazione della visione di Gödel sulla base dei suoi scritti (un aspetto che, nel presente lavoro, si è invece deciso di privilegiare).

[293]*Ibid.*, p. 382.

[294]*Ibid.*, p. 384.

[295]*Ibid.*, p. 386.

[296]Il problema, che la pur ingente documentazione del *Nachlass* fin qui pubblicata lascia sostanzialmente aperto, di determinare se e come siano avvenuti contatti tra Gödel e filosofi acclarati, non deve sorprendere. Soprattutto perché ben più clamorose appaiono certe 'lacune' rispetto a personalità riconosciute del mondo matematico che, parrebbe logico aspettarsi, Gödel poteva avere tutto l'interesse a contattare. Tra queste, spicca la figura di Hermann Weyl il quale, oltretutto, fu per lunghi anni collega di Gödel all'Insitute for Advanced Studies avendo accettato un posto da docente a Princeton nel 1933 e avendolo mantenuto fino al 1951.

2.5.1 Il sogno leibniziano del platonista

Molto di quanto si è detto fin qui, va letto ovviamente come il tentativo di far emergere e, in un certa misura, chiarire, alcuni aspetti 'filosofici' del pensiero di Gödel. Vorremmo quindi concludere questa parte del lavoro con una breve considerazione relativa alla visione da parte di Gödel della filosofia.

Si tende comunemente a suddividere l'opera di Gödel in due parti: una prima alla quale appartengono tutti i lavori di carattere tecnico, che è cronologicamente situata grossomodo nel decennio 1930-40; una seconda, il cui inizio lo si può fare per convenienza coincidere con il trasferimento definitivo di Gödel negli Stati Uniti all'inizio del 1940, che contiene quelle ricerche che risentono del progressivo sopravvento che le indagini intorno a problematiche di natura più squisitamente filosofica, assunsero sulle questioni di tipo matematico[297].

È ovvio che il giudizio sui contributi di Gödel alla logica e alla matematica da un lato, e alla filosofia (della matematica in particolare) dall'altro, non possono che essere molto diversi tra loro.

Nell'ottica gödeliana, una simile discrepanza, così come una certa difficoltà del Nostro a misurarsi con tematiche di tipo filosofico[298], troverebbe probabilmente spiegazione nei differenti gradi di sviluppo delle due discipline.

Ci sono indizi che Gödel ritenesse infatti come una certa inconclusività del discorso filosofico, fosse la naturale conseguenza dello sviluppo non ancora completato delle capacità raziocinanti in questo settore del sapere, e che un miglioramento decisivo dello stato della disciplina sarebbe giunto attraverso un incremento in sistematicità dell'analisi concettuale, lungo quelle linee di ricerca che avevano condotto Leibniz ad individuare la propria *Characteristica universalis*[299].

Dato il ruolo cruciale che una simile analisi svolge in ambito matematico, c'è da aspettarsi che Gödel vedesse in un simile sviluppo un passaggio essenziale per l'incremento decisivo della conoscenza matematica secondo il proprio 'programma' per i nuovi assiomi.

[297] Significativo, a tale proposito, come si è già avuto modo di notare, è l'abbandono da parte di Gödel delle ricerche su una dimostrazione di indipendenza dell'Ipotesi del Continuo all'inizio degli anni '40.

[298] Come esempio che ben riassume queste difficoltà, valga la genesi tormentatissima del lavoro dedicato alla filosofia di Carnap, la ricerca di una stesura soddisfacente del quale, infruttuosa, impegnò Gödel per ben sei anni tra il 1953 e il 1959.

[299] Questo fatto viene esplicitamente sottolineato in chiusura del lavoro dedicato alla logica matematica di Russell ([Göd44], pp. 140-141).

Tuttavia, è proprio la filosofia, ed il punto di vista di Gödel sulla matematica in particolare, pare chiaro, che ne avrebbe giovato, acquisendo quel rigore che, leibnizianamente, avrebbe tramutato i tentativi gödeliani di produrre argomenti sulla base dei quali confutare i punti di vista avversi al proprio, in dimostrazioni stringenti ed ineludibili:

> Ho l'impressione che a seguito di una chiarificazione dei concetti in questione sarà possibile condurre questo genere di discussioni con il rigore matematico, e che il risultato sarà che (sotto certe ipotesi che difficilmente potrebbero essere rifiutate, ed in particolare l'assunzione che esiste qualcosa come una conoscenza matematica) il punto di vista Platonista è l'unico sostenibile.[300]

[300][Göd51], pp. 322-323.

PARTE II

LOGICA E MECCANICISMO

Capitolo 3

Dalla parte delle macchine: Turing e i fondamenti dell'Intelligenza Artificiale

Il nome di Alan Mathison Turing è inevitabilmente legato, in un modo o nell'altro, all'idea di macchina. In un senso, perchè il suo contributo alla logica matematica più celebre, quell'articolo sui "numeri computabili" al quale ci siamo più volte richiamati nel corso della prima parte del presente lavoro, contiene la definizione del modello astratto di dispositivo meccanico, meglio noto come 'macchina di Turing', che è stato tra i concetti fondamentali per la nascita della teoria della ricorsività, e che ha rappresentato, poi, la nozione teorica di riferimento per la *computer science*. In secondo luogo, perché a partire dagli anni '40 del '900, egli fu impegnato in una serie di progetti ai quali si deve la nascita di alcuni tra i primi esempi dei calcolatori digitali moderni.

I contributi scritti che sono la testimonianza di questo impegno, offrono molteplici spunti di riflessione. Questo è già vero per quel che riguarda il lavoro del 1937, come appare chiaro sulla base dell'analisi di cui esso è stato oggetto in precedenza. Ma ciò risulta in modo persino più evidente, se quello stesso contributo viene letto alla luce degli scritti più tardi, nei quali il problema della costruzione (nel senso letterale del termine) di certe macchine calcolatrici, si intreccia con una serie di considerazioni relative alle capacità che certi dispositivi meccanici sono in grado di sviluppare, sfociando nell'idea di implementare su di essi una qualche forma di 'intelligenza'.

È sulla base di questo progetto che a Turing si ascrive comunemente una concezione meccanicista della mente umana.

Con il presente capitolo ci proponiamo di offrire un'analisi dei lavori più rappresentativi dell'intento di Turing, alla luce della quale riprendere il tema del rapporto tra menti e macchine attraverso un confronto tra la posizione del matematico britannico e la prospettiva gödeliana.

3.1 L'indagine sulle Logiche Ordinali

Tra il 29 settembre 1936 ed 18 luglio del 1939, Turing svolse un lungo soggiorno presso l'Institute for Advanced Studies di Princeton. Si trattava, retrospettivamente parlando, di un esito scontato per l'autore di uno dei testi fondamentali per la 'confluenza di idee' (come all'evento si è riferito Robin Gandy) che, intorno al 1936, aveva portato all'atto di origine della teoria della ricorsività, quello di ritrovarsi in mezzo a quanti, come Church e Kleene, avevano avuto un ruolo parimenti essenziale per lo stesso settore di ricerca.

È in questo quadro che si inserisce la Tesi di Dottorato di Turing dedicata alle Logiche Ordinali. E fu dietro suggerimento di Church, che il matematico britannico decise di occuparsi di certe implicazioni dei teoremi di Gödel[301].

Per quanto il legame di merito con gli scritti più tardi possa apparire labile, è in questa sede, attraverso una riflessione sulla conoscenza matematica, che la problematica della natura della mente umana comincia a profilarsi.

Prima di dedicarci a questo tema, che è legato al nucleo concettuale del-

[301]La coincidenza che vede il lavoro di Turing, che è direttamente legato ai teoremi di incompletezza, esser stato elaborato a Princeton, potrebbe ingenerare la curiosità sul fatto se vi sia stato un contatto diretto, in qualche forma, con lo stesso Gödel (il quale, come si sarà potuto apprezzare già da alcuni riferimenti contenuti nella prima parte, aveva stabilito un legame proficuo con il gruppo logico dell'Institute for Advanced Studies, che costituirà la premessa per il suo trasferimento definitivo a Princeton a partire dal gennaio 1940). Stando alle fonti note, tuttavia, Turing non ebbe occasione di incontrare Gödel (né intrattenne con lui alcuno scambio epistolare), nonostante i soggiorni negli Stati Uniti dei due si siano sovrapposti per un breve periodo. Gödel infatti, che era già stato a Princeton due volte, una prima dal 30 settembre 1933 al 26 maggio 1934 ed una seconda volta nel settembre 1935, visita, quest'ultima, interrotta poi per motivi di salute alla fine di novembre, tornò all'IAS (prima del suo trasferimento definitivo), una terza volta nel novembre 1938 per esporre il risultato di consistenza dell'Ipotesi del Continuo con la teoria degli insiemi. Tuttavia, già nel febbraio del 1939 egli si spostò per tutto il semestre primaverile alla Notre Dame University per poi ripartire per l'Europa nel giugno dello stesso anno.

l'intero capitolo, ci pare opportuno presentare brevemente, in primo luogo, i risultati principali acquisiti da Turing con la propria Tesi.

3.1.1 L'idea, i risultati

La genesi del concetto di Logica Ordinale, in modo che ne emerga la relazione con i teoremi di incompletezza e persino con alcune delle considerazioni di Gödel a proposito di quei risultati, può essere esemplificata, in modo informale, come segue.

Supponiamo di considerare la seguente successione $\langle \mathsf{S}_i \rangle_{i \in \mathbf{N}}$ di sistemi formali:

$$\begin{aligned} \mathsf{S}_0 &:= \mathsf{PA} \\ \mathsf{S}_{n+1} &:= \mathsf{S}_n \cup \{Con_n\} \end{aligned}$$

(dove Con_n indica l'enunciato che esprime la consistenza del sistema S_n).

Guardando ad essa da un punto di vista intutivo, si osserva quanto segue: il sistema di partenza, l'aritmetica di Peano, viene vista comunemente come la teoria i cui assiomi racchiudono le proprietà basilari dei numeri naturali; essa viene quindi considerata una teoria intuitivamente *corretta*, i cui teoremi sono proposizioni numeriche *vere*; la definizione permette inoltre, dato un sistema S_n della successione, di costruirne in modo effettivo il suo successore: ciò significa che c'è un *metodo* (aggiungere la proposizione che esprime la consistenza del sistema) sulla base del quale, dato S_n si ottiene il nuovo sistema S_{n+1}; la relazione tra un sistema formale della successione individuata ed il proprio successore, è poi tale che: (i) l'insieme dei teoremi del primo è contenuto propriamente (a causa del secondo teorema di Gödel) nell'insieme dei teoremi del secondo, e, (ii), la correttezza del primo implica la correttezza del successore (se S_n è corretto, ossia dimostra solo proposizioni vere, allora la proposizione Con_n, in particolare, è vera e quindi S_{n+1} è anch'esso corretto).

Se tralasciamo per il momento il fatto che egli si riferisce a successioni transfinite di sistemi formali, queste sono le circostanze generali alle quali, nota Turing, è possibile dare vita a Logiche Ordinali di un qualche rilievo.

Il modo con il quale il concetto intuitivo viene poi effettivamente reso nella tesi del matematico britannico, risente tuttavia dalla scelta del formalismo di riferimento (la quale è, a sua volta, motivata dal ruolo di supervisore del lavoro ricoperto da Church).

Turing opera infatti entro la cornice del "calcolo delle conversioni", oggi meglio noto come λ-calcolo, nel quale la relazione fondamentale tra le espressioni, quella di β-riduzione (in un passo), è definita dalla seguente inferenza:

$$(\lambda x.M)N \rightarrow_\beta M[x := N]$$

(dove M, N sono termini del linguaggio che contiene, oltre alle funzioni di astrazione - $\lambda x.M$ - e applicazione - $(M)N$ - un'infinità numerabile di variabili, e $M[x := N]$ indica il risultato della sostituzione di N ad x in M).

È a partire da questa relazione, che si definisce quella di 'conversione' tra espressioni del linguaggio (ben formate[302]), alla quale si deve il nome d'origine del calcolo: si tratta infatti della chiusura riflessiva, transitiva e simmetrica della β-riduzione in un passo (conveniamo di scrivere, con Turing, '$M\mathrm{conv}N$' per 'M si converte a N', nel senso indicato).

Un secondo, più decisivo, elemento di differenza tra il lavoro di Turing e i termini dell'esempio dato in apertura di paragrafo, consiste in una restrizione sulla forma logica dei teoremi che vengono considerati rilevanti ai fini dell'indagine, e rispetto ai quali sono date le definizioni successive.

Nel §3 di [Tur39][303], infatti, Turing isola sostanzialmente due tipi di proposizioni alle quali egli intende confinare la propria attenzione: (i) enunciati della forma $\forall x(f(x) = 0)$ e, (ii), formule chiuse del tipo $\forall x \exists y(g(x, y) = 0)$, con f e g funzioni primitive ricorsive (si tratta quindi di enunciati che sono, nella terminologia contemporanea, formule Π^0_1 e Π^0_2, rispettivamente). Problemi di questo tipo sono ciò che Turing chiama i "teoremi aritmetici"[304].

A parte il fatto ovvio che le classi di formule in questione sono particolarmente legate ai risultati di Gödel, ai quali il lavoro è debitore dello spunto iniziale, la motivazione che viene addotta per giustificare la restrizione ai soli "teoremi aritmetici" pare, stando al testo, sia da ritrovarsi unicamente nell'affermazione di Turing secondo la quale "un certo numero di problemi irrisolti, come il problema della verità dell'ultimo teorema di Fermat, sono aritmetici [ed] [e]sistono anche problemi dell'analisi che sono [dello stesso ti-

[302]Come è noto, nel caso del λ-calcolo (non tipato) le espressioni ben formate, sono quelle che appartengono alla collezione definita induttivamente a partire dalle variabili, per chiusura sotto astrazione e applicazione.

[303][Tur39], pp. 162-166. *NdA*: i riferimenti a [Tur39] sono relativi alla riedizione di quest'ultimo in [Dav65] come risulta dalla voce bibliografica.

[304]La terminologia usata da Turing, si discosta, su questo punto, in modo evidente da quella che è oggi in uso, dal momento che con 'formule aritmetiche' si suole indicare quelle della classe $\Pi^0_\infty := \bigcup_{n \in \mathbf{N}} \Pi^0_n$, ossia formule ottenute da una matrice primitiva ricorsiva mediante un numero finito arbitrario di quantificazioni alternate (con un quantificatore universale in testa).

po]"[305]. L'esempio che, di questi ultimi, Turing fornisce è quello dell'Ipotesi di Riemann.

I "teoremi aritmetici" sono dunque una classe di problemi matematicamente rilevanti, e della quale fanno parte alcune delle questioni indecise più importanti.

La trattazione di queste classi di enunciati in rapporto al formalismo scelto, è resa possibile dalla seguente osservazione[306]: ogni "teorema aritmetico" equivale al fatto che una certa espressione A del 'calcolo delle conversioni' è tale che, per ogni numerale $\overline{n}$[307], si abbia:

$$(A)\overline{n}\mathrm{conv}\overline{2}$$

(proprietà che Turing esprime come il fatto che A è "doppia", o "duale" e che indicheremo in seguito con $\Delta(A)$).

È possibile, anzi, associare ad ogni "teorema aritmetico" un'espressione del linguaggio del λ-calcolo tale che la verità del primo, coincida con la dualità della seconda.

Attraverso quest'equivalenza, si giunge così all'introduzione del concetto fondamentale di *logica*, con il quale Turing intende un'espressione L del linguaggio tale che, per ogni altra espressione A:

$$(L)A\mathrm{conv}\overline{2} \Rightarrow \Delta(A)$$

Una logica, non è dunque altro che un metodo per decidere se un'espressione del linguaggio è duale, e quindi, per la definizione di dualità succitata, per determinare se il corrispondente "teorema aritmetico" è vero.

Data una logica L, con *estensione di* L, Turing indica inoltre la classe di quelle espressioni A per le quali si ha $(L)A\mathrm{conv}\overline{2}$. Una classe di logiche è quindi *completa*, se la riunione delle estensioni delle logiche che la compongono contiene tutte le espressioni che sono duali.

Quanto detto, può essere espresso in termini equivalenti come segue: una logica è un insieme ricorsivamente enumerabile $L \subseteq \mathbf{N}$, i cui elementi sono i

[305] *Ibid.*, p. 165.
[306] *Ibid.*, pp. 164-165.
[307] Come si è soliti fare nel caso di un formalismo per il λ-calcolo, vale:

$$\overline{n} := \lambda fx.\underbrace{f(f(\ldots f}_{n-\text{volte}}(x)\ldots))$$

gödeliani di enunciati Π^0_1 o Π^0_2, tale che:

$$\ulcorner \phi \urcorner \in L \Rightarrow \phi \text{ è vero}$$

Una collezione di logiche è completa, se la riunione insiemistica applicata ad essa equivale alla classe degli enunciati veri di tipo Π^0_1 o Π^0_2.

Ciò ci consente di riconnetterci all'esemplificazione dell'idea di base dalla quale abbiamo preso le mosse, dal momento che la nozione di "logica" non è altro che un modo per caratterizzare l'idea di un sistema formale d'assiomi, potendo noi vedere quest'ultimi come insiemi ricorsivamente enumerabili di (codici di) enunciati ottenuti dagli assiomi (o meglio, dai loro gödeliani) per chiusura sotto la nozione di derivazione.

Ma ora possiamo anche considerare quell'estensione dell'idea intuitiva che avevamo inizialmente lasciato da parte: è possibile costruire una successione transfinita di sistemi formali?

Perché l'eventualità sia realizzabile è necessario che si sia in possesso di una descrizione costruttiva dei numeri ordinali, un fatto essenziale affinché, considerando il 'sistema limite' di una successione di teorie $\mathsf{S}_{a_1}, \mathsf{S}_{a_2}, \mathsf{S}_{a_3}, \ldots$ (il quale, in modo del tutto naturale, verrà definito come la riunione degli S_{a_i}, considerati estensionalmente), si abbia ancora a che fare con un insieme ricorsivamente enumerabile di teoremi (che risulti quindi esprimibile come una logica).

Turing affronta la questione della rappresentazione degli ordinali nel §7, prospettando due soluzioni.

La prima, consiste nel restringere l'attenzione agli ordinali ricorsivi: si considerano cioè le relazioni R di buon ordine definite su insiemi di numeri naturali che siano ricorsive; si chiamano *ricorsivi* gli ordinali $\alpha \in \Omega$ per i quali esiste un buon ordine ricorsivo R che sia isomorfo ad essi.

Assumiamo di indicare, nel seguito, con $\mathcal{W} := \langle D_{\mathcal{W}}, |\ |_{\mathcal{W}} \rangle$ il sistema di notazioni per ordinali composto dall'insieme $D_{\mathcal{W}} \subseteq \mathbf{N}$ degli indici (delle funzioni caratteristiche) di buoni ordini ricorsivi, e dalla funzione di rappresentazione $|\ |_{\mathcal{W}}$ che associa ad ognuno di essi l'ordinale al quale sono isomorfi.

La seconda via, passa invece attraverso le notazioni per gli ordinali introdotte da Church e Kleene[308].

[308] [CK36]. Come è noto, la modifica della definizione originaria operata da Kleene in [Kle38], che prevede la definizione simultanea del sistema di notazioni $\mathcal{O}$ ed una relazione d'ordine parziale $<_{\mathcal{O}}$ su di esso, sarebbe la versione più conveniente. Essa consente infatti, sfruttando il teorema di recursione, di definire alcune operazioni ricorsive parziali tra

Il sistema di notazioni $\mathcal{O} := \langle D_{\mathcal{O}}, |\ |_{\mathcal{O}} \rangle$ in questione, prevede la definizione induttiva dell'insieme $D_{\mathcal{O}} \subseteq \mathbf{N}$, il quale soddisfa le seguenti condizioni (rispettivamente la clausola per l'elemento iniziale, per la notazione 'successore' e per la notazione 'limite'): (i) $1 \in D_{\mathcal{O}}$, (ii) $\forall n \in \mathbf{N}[n \in D_{\mathcal{O}} \Rightarrow 2^n \in D_{\mathcal{O}}]$, (iii) $\forall n \in \mathbf{N}[\{e\}(n) \in \mathcal{O}\ \&\ \{e\}(n) < \{e\}(n+1) \Rightarrow 3 \cdot 5^e \in \mathcal{O}]$[309]. Per quel che riguarda invece la funzione di rappresentazione $|\ |_{\mathcal{O}}$, che associa ad ogni elemento di $D_{\mathcal{O}}$ un ordinale, essa è conseguentemente definita dalle clausole: (i)$|1|_{\mathcal{O}} = 0$, (ii) $|2^n|_{\mathcal{O}} = |n|_{\mathcal{O}} + 1$, (iii) $|3 \cdot 5^e|_{\mathcal{O}} = \sup_n(|\{e\}(n)|_{\mathcal{O}})$.

Sappiamo oggi, che le due soluzioni indicate da Turing per trattare gli ordinali in modo costruttivo sono equivalenti: ogni ordinale per il quale esiste una notazione in $\mathcal{O}$ è ricorsivo e, viceversa, ogni ordinale ricorsivo è rappresentabile in $\mathcal{O}$[310].

Ad ogni modo, per tornare al testo, la possibilità di trattare gli ordinali costruttivamente, considerato che da entrambe le soluzioni prospettate discende l'esistenza di espressioni ben formate del λ-calcolo, delle *formule ordinali*, che rappresentano numeri ordinali, consente al matematico britannico di introdurre infine, nel §8, le logiche ordinali: con questo termine si indica un'espressione Λ tale che, per ogni formula ordinale O, $(\Lambda)O$ è una logica.

Ciò significa quindi, che una logica ordinale non è altro che una funzione ricorsiva parziale la quale, in corrispondenza ad ogni ordinale costruttivo, dà

---

notazioni, ad esempio di somma e prodotto, che soddisfano una condizione di adeguatezza rispetto alle corrispondenti operazioni tra gli ordinali rappresentati. Un secondo vantaggio, è dato dal fatto che il sistema di notazioni in questione è *universale*: si dimostra che dato un qualsiasi altro sistema di notazioni $\mathcal{S}$ (indicato con $D_{\mathcal{S}}$ l'insieme delle notazioni), esiste una funzione ricorsiva parziale e iniettiva f, che fa corrispondere a $D_{\mathcal{S}}$ un sottoinsieme dell'insieme di notazioni $D_{\mathcal{O}}$ di $\mathcal{O}$, ed è tale che (se $|\ |_{\mathcal{S}}$ e $|\ |_{\mathcal{O}}$ sono le funzioni di rappresentazione dei due sistemi), $n \in D_{\mathcal{S}} \Rightarrow |n|_{\mathcal{S}} \leq |f(n)|_{\mathcal{O}}$ (dal ché segue che $\mathcal{O}$ ha almeno una notazione per ogni ordinale rappresentato in $\mathcal{S}$). Per una dimostrazione di entrambi i fatti, si veda ad esempio [Rog67], pp. 209-210.

[309] Quest'ultima condizione asserisce l'esistenza di un elemento di $D_{\mathcal{O}}$ in relazione a (ed in funzione di) ogni successione strettamente crescente e generata in modo effettivo di elementi di $D_{\mathcal{O}}$.

[310] Il risultato chiave in questo senso è dovuto a Werner Markwald ([Mar54]) e, indipendentemente, a Clifford Spector ([Spe55]). Una dimostrazione dell'equivalenza indicata nel testo, può essere trovata in [Rog67], pp. 211-212. Facciamo notare, tuttavia, che occorre qualche piccolo accorgimento per 'aggiornare' del tutto la trattazione di Turing: ciò è dovuto al fatto che se esiste una funzione ricorsiva che preserva l'ordine tra $D_{\mathcal{O}}$ e $D_{\mathcal{W}}$, il viceversa vale invece soltanto rispetto ad un sottoinsieme proprio di quest'ultimo insieme (si rimanda all'esauriente operazione di aggiornamento di [Tur39], operata da Feferman in [Fef88], per tutti i dettagli tecnici).

una logica, ossia un insieme ricorsivamente enumerabile con le caratteristiche indicate in precedenza.

Come si sarà notato dalla definizione del sistema $\mathcal{O}$ di notazioni, la funzione di rappresentazione, che associa agli elementi di $D_{\mathcal{O}}$ un ordinale di un certo segmento iniziale della classe Ω, non è iniettiva[311].

Ciò è indice di un fatto del tutto generale: per ogni ordinale che sia rappresentabile costruttivamente, ci possono essere più notazioni.

Una simile osservazione è all'origine dell'introduzione[312], da parte di Turing, della proprietà di *invarianza* di una logica ordinale: una logica ordinale Λ è *invariante fino all'ordinale* α se, per ogni $a, a' \in D_X$ (con $X = \mathcal{O}, \mathcal{W}$) tali che $|a|_X = |a'|_X < \alpha$, le estensioni di $(\Lambda)a$ e $(\Lambda)a'$, cioè la collezione delle espressioni 'dimostrabilmente' duali in esse (o, equivalentemente, dei "teoremi aritmetici" veri), sono uguali[313].

Dunque, per una logica ordinale essere invariante significa, in un certo senso, essere indipendente dalla scelta di una tra le varie possibili notazioni costruttive per ordinali. Considerato che il problema della completezza rispetto ad una certa logica ordinale Λ, non è altro che, nella sua forma più forte, il problema di determinare l'esistenza di una certa notazione a tale che l'estensione di $(\Lambda)a$ contiene tutti gli enunciati veri di tipo Π^0_2, si comprenderà l'importanza della proprietà di invarianza per lo scopo che il contributo di Turing si prefigge: studiare la questione della completezza sintattica, nel caso generalizzato delle successioni transfinite (di tipo costruttivo) di sistemi formali.

Prima di passare ad illustrare i risultati generali sulle logiche ordinali, Turing si sofferma su due esempi.

L'argomento che Turing fornisce[314] per mostrare come, in corrispondenza ad una collezione di sistemi formali che soddisfano le condizioni indicate in

[311]Di questo fatto ci si può facilmente convincere considerando l'insieme degli ordinali finiti $n_{\mathcal{O}}$ in $\mathcal{O}$ (cioè gli elementi della successione $1, 2, 2^2, 2^{2^2}, \ldots$). Assumendo di indicare con e, e' i codici, rispettivamente, della funzione $n \mapsto n_{\mathcal{O}}$ (che associa ad $n \in \mathbf{N}$ il suo rappresentante in $\mathcal{O}$), e della funzione $n \mapsto 2^{n_{\mathcal{O}}}$ (che gli associa il successore del rappresentante), si osserva che $a = 3 \cdot 5^e$ e $b = 3 \cdot 5^{e'}$ sono tali che: (i) $a, b \in D_{\mathcal{O}}$, (ii) $a \neq b$ e (iii) $|a|_{\mathcal{O}} = |b|_{\mathcal{O}} = \omega$.

[312][Tur39], p. 194.

[313]Una logica è quindi *invariante*, se lo è fino ad ogni ordinale per il quale esiste una notazione, ossia se lo è fino all'ordinale ω^{CK} con il quale si è soliti indicare il più grande ordinale per il quale esiste una notazione nel sistema universale di Kleene.

[314][Tur39], pp. 184-187.

apertura di paragrafo, sia possibile definire una logica ordinale[315], permette, a noi, di riconciliare la nostra illustrazione intuitiva con la trattazione effettiva di Turing, al matematico britannico di presentare esempi di logiche ordinali nella forma di collezioni di sistemi formali nel senso usuale del termine (anziché di procedure λ-definibili di enumerazione).

Il caso principale, essendo quello discusso più nel dettaglio[316], è l'esempio della logica ordinale $\Lambda_P := \langle \mathsf{S}_a \mid a \in \mathcal{O} \rangle$, definita a partire dal sistema P ispirato al formalismo dei *Principia mathematica* e utilizzato da Gödel in [Göd31b].

La collezione di teorie assiomatiche indicate da Turing, può essere descritta mediante le seguenti clausole:

$$\begin{array}{rcl} \mathsf{S}_1 & := & P \\ \mathsf{S}_{2^a} & := & \mathsf{S}_a \cup \{L - Ref_a\} \\ \mathsf{S}_{3 \cdot 5^e} & := & \bigcup\limits_{n<\omega} \mathsf{S}_{\{e\}(n)} \end{array}$$

dove $L - Ref_a$ indica il cosiddetto *principio di riflessione locale* per S_a, ossia lo schema:

$$Teor_{\mathsf{S}_a}(\ulcorner \phi \urcorner) \rightarrow \phi$$

ristretto a formule ϕ di tipo Π^0_2.

Si noti che l'enunciato Con_{S}, ossia $\neg Teor_{\mathsf{S}}(\ulcorner 0 = 1 \urcorner)$, equivale logicamente all'enunciato $Teor_{\mathsf{S}}(\ulcorner 0 = 1 \urcorner) \rightarrow 0 = 1$, cioè ad un'istanza di $L - Ref_{\mathsf{S}}$. Il principio mediante il quale si costruisce Λ_P, comprende dunque, come caso particolare, quello dell'indecidibilità gödeliana che abbiamo utilizzato per illustrare intuitivamente il concetto di logica ordinale.

Il secondo esempio che Turing offre, è la collezione $\Lambda_H := \langle \mathsf{S}_a \mid a \in D_{\mathcal{W}} \rangle$ (per un'introduzione della quale, vale la pena notare, egli rimanda alla nota 48a di [Göd31b][317]), è costituito da sistemi tipati S_a i quali, per $a \in D_{\mathcal{W}}$, con-

[315] In generale, si richiede l'esistenza di una collezione $\mathcal{C}$ di sistemi formali tale che: (i) di essa faccia parte un sistema C_0 nel linguaggio del quale siano esprimibili tutti i "teoremi aritmetici", e che sia corretto per essi; (ii) ci sia un metodo effettivo, sotto il quale $\mathcal{C}$ è chiusa, per passare da un sistema C di $\mathcal{C}$ ad un'estensione C' in modo che si preservi, nel passaggio, la correttezza; (iii) $\mathcal{C}$ sia chiusa sotto la riunione (effettiva) di sistemi che ne fanno parte.

[316] *Ibid*, pp. 187-190.

[317] Si tratta della nota nella quale Gödel sottolinea come "la vera ragione dell'incompletezza" risieda nella possibilità di estendere ogni sistema formale con le variabili di tipo immediatamente superiore a quelle che il sistema possiede. Nota, della quale si è parlato diffusamente nella prima parte del lavoro e che si è citato per intero nel §1.3.

tengono un'infinità numerabile di variabili individuali $x_0^n, x_1^n, \ldots$, (per ogni n che sia un elemento del dominio del buon ordine indiciato da a' tale che $|a'|_{\mathcal{W}} = \omega + |a|_{\mathcal{W}}$), le cui formule atomiche sono della forma $x^m \in x^n$ per $m <_{a'} n$.

Si può presumere, per quanto Turing non lo indichi esplicitamente nel testo, che i sistemi di Λ_H possiedano una versione stratificata dell'assioma di comprensione[318].

È rispetto a Λ_P e Λ_H in particolare, che Turing valuta i risultati ottenuti, per quanto essi abbiano un carattere assolutamente generale[319].

Sono cinque le proposizioni, tra quelle menzionate nel lavoro, che meritano particolare attenzione. Queste, possono essere riformulate come segue[320]:

1. Se una logica ordinale Λ è invariante, allora non può essere *strettamente crescente*, ossia *non* vale che, se a, a' sono notazioni per ordinali tali che $a < a'$, allora $(\Lambda)a \subset (\Lambda)a'$.

2. Se Λ è (una logica ordinale basata su $\mathcal{W}$ e) invariante fino all'ordinale α, allora per ogni $a \in D_{\mathcal{W}}$ che rappresenti un ordinale $\beta < \alpha$, $(\Lambda)a \subseteq \bigcup_{n_W < \omega_W} (\Lambda)n_W$ (dove con n_W indichiamo gli indici di ordinali finiti in $\mathcal{W}$ e con ω_W la corrispondente notazione limite).

3. Se Λ è (basata su $\mathcal{O}$ e) invariante fino ad α, allora per ogni $a \in D_{\mathcal{O}}$ che rappresenti un $\beta < \alpha$, $(\Lambda)a \subseteq \bigcup_{a' < \omega_O^2} (\Lambda)a'$ (dove ω_O^2 è ovviamente un elemento di $D_{\mathcal{O}}$ tale che $|\omega_O^2|_{\mathcal{O}} = \omega^2$).

4. Λ_H è invariante.

5. Λ_P è completa per gli enunciati di tipo Π_1^0[321].

[318] Si veda [Fef88], p. 125. La sintassi dei sistemi che compongono Λ_H non viene descritta da Turing. Feferman nota come la scelta di $|a'|_{\mathcal{W}} = \omega + |a|_{\mathcal{W}}$ sia probabilmente motivata dal fatto che, in questo modo, risulta che S_1 di Λ_H equivale alla teoria dei tipi finiti P.

[319] A dire il vero, nel paragrafo conclusivo del lavoro (§12) Turing introduce anche le logiche ordinali Λ_{G^i} (con $i = 1, 2, 3$), tutte basate sul calcolo delle sequenze di Gentzen. Tuttavia, la definizione della sintassi lascia, in questo caso, decisamente a desiderare, rendendo difficile una resa soddisfacente di queste logiche (si comprende comunque che l'idea di fondo è quella di incorporare, nei sistemi che compongono queste logiche ordinali, regole di definizione per recursione e di dimostrazione per induzione su a, per ogni $a \in W$).

[320] Per i dettagli della dimostrazione di ognuna di esse, si rimanda a [Fef88], pp. 126-128.

[321] A questi risultati bisognerebbe aggiungere, per completare il quadro, la congettura di Turing (cfr. [Tur39], p. 203) che Λ_P sia anche Π_2^0-completa. Questa congettura verrà tuttavia confutata da Feferman nel suo [Fef62].

Il significato di 2 e 3 (che sono raffinamenti di 1) dovrebbe essere chiaro: se una logica ordinale è invariante fino ad un certo ordinale costruttivo (supponiamo anche molto grande) α, la collezione degli enunciati veri di tipo Π_2^0 in essa dimostrabili in relazione ad un certo altro ordinale, per quanto vicino ad α esso sia, non può estendere l'insieme degli enunciati di quel genere che sono dimostrabili rispetto ad un ordinale fissato, che può risultare ad esso minore e che dipende dalla scelta del sistema di notazioni.

In particolare, allora, non ci può essere una logica ordinale invariante e completa. Segue quindi, da 4 e 5, che Λ_H è incompleta e che Λ_P non è invariante.

Questo, spiega lo scarso entusiasmo che Turing dimostra per i risultati ottenuti (egli chiama 2 e 3 "Teoremi di Incompletezza"), ed in modo particolare per l'ultimo della lista, rispetto al quale chiosa:

> Questo teorema di completezza, come al solito, non è di nessun aiuto. Per quanto esso mostri, ad esempio, che è possibile dimostrare l'ultimo teorema di Fermat mediante Λ_P (se esso è vero), la verità del teorema sarebbe assunta considerando una certa [espressione] come una formula ordinale.[322]

Da un punto di vista generale, l'insoddisfazione da Turing può essere spiegata come segue: il teorema di Π_1^0-completezza per una logica ordinale Λ che sia basata sul sistema di notazioni $\mathcal{O}$, asserisce che, dato un certo enunciato ϕ di tipo Π_1^0, esiste un $a \in D_{\mathcal{O}}$ tale che $\ulcorner\phi\urcorner \in (\Lambda)a$[323]. La dimostrabilità di ϕ, e dunque la sua verità, assumendo la consistenza della logica ordinale[324], dipendono dalla capacità di asserire, per una certa espressione a, che essa è una notazione per ordinale. In particolare, se a è una notazione per un ordinale limite, e quindi se esso è definito in relazione ad una successione di notazioni $a_0, a_1, \ldots$, si ha che:

$$a \in D_{\mathcal{O}} \Leftrightarrow \forall x[a_x \in D_{\mathcal{O}} \wedge \forall y, z(y < z \to a_y < a_z)]$$

Si ha quindi che la verità di ϕ dipende dal saper riconoscere come vero un enunciato che può essere almeno altrettanto complicato di ϕ[325]. Dunque,

[322][Tur39], p. 207

[323]Poiché questa è la forma di ogni teorema di completezza per una logica ordinale, a ciò si può ricondurre il "come al solito" del passo di Turing.

[324]Che equivale, come è noto, alla Π_1^0-validità.

[325]Verrà dimostrato poi da Kleene (in [Kle55]), che $a \in \mathcal{O}$ non è un enunciato aritmetico (ossia non equivale ad alcuna formula in forma prenessa ottenuta da una matrice ricorsiva con un numero finito, ma illimitatatamente grande di quantificatori numerici). Dunque, la

per quanto *sappiamo* che per ogni enunciato vero di quella forma c'è una dimostrazione di esso ad un qualche livello della successione Λ_P, non si è affatto con ciò ridotto le difficoltà del problema di *determinarne* la verità.

La lettura del risultato di completezza per Λ_P, prelude al bilancio sul valore dell'intera indagine sulle logiche ordinali, al quale, seguendo Turing, vogliamo adesso rivolgere la nostra attenzione.

3.1.2 Intuizione e Ingegnosità

Come si diceva nel paragrafo precedente cercando di illustrare l'idea alla base delle logiche ordinali, il lavoro di Turing è fortemente indebitato, nella sua origine, nei confronti di una certa lettura dei teoremi di incompletezza. Questo fatto, viene esplicitamente riconosciuto in apertura di [Tur39], laddove si legge:

> I ben noti teoremi di Gödel [...] mostrano che ogni sistema logico è in un certo senso incompleto, ma allo stesso tempo essi indicano il modo in cui da un sistema logico L, un sistema L' più completo del primo può essere ottenuto. Ripetendo il processo si ottiene una successione [di sistemi] $L, L_1 = L', L_2 = L'_1, \ldots$ ognuno dei quali è più completo del precedente.[326]

I risultati di Gödel, dunque, non solo indicano l'esistenza di un enunciato indecidibile in un dato sistema formale, e quindi ne mostrano l'incompletezza sintattica, ma, guardando al medesimo fatto da un punto di vista diverso, essi rivelano anche il modo con cui quello stesso sistema può essere completato (ad esempio, aggiungendovi l'enunciato che ne esprime la consistenza). È in questo secondo senso, in particolare, che i teoremi di incompletezza forniscono lo spunto per l'idea intuitiva alla base delle logiche ordinali.

L'incompletezza e l'incompletabilità di ogni sistema consistente e ricorsivamente assiomatizzabile, ossia i risultati di Gödel nel loro senso più diretto, non sono tuttavia fatti eludibili.

Questa consapevolezza, conduceva Turing a formulare le proprie aspettative nei confronti dell'indagine intorno alle logiche ordinali, nel modo seguente:

---

verità di un enunciato Π^0_1 dipende in realtà dalla verità di una proposizione estremamente più complicata di esso. Feferman nota, più 'localmente', come il giudizio di Turing possa essere riferito anche alla particolare tecnica di prova usata per dimostrare il teorema di completezza (si veda [Fef88], p. 128).

[326][Tur39], p. 155.

> Si può sperare di ottenere un sistema di inferenze logiche soddisfacente (per la prova di teoremi aritmetici) mediante qualche logica ordinale. Il teorema di Gödel mostra che tale sistema non può essere interamente meccanico; ma con una logica ordinale completa si dovrebbe essere in grado di confinare interamente i passi non meccanici alla verifica che certe particolari [espressioni] sono formule ordinali.[327]

La questione viene ripresa dal matematico britannico nel §11 del lavoro, immediatamente dopo il commento al teorema di completezza per la logica ordinale Λ_P sul quale abbiamo chiuso il paragrafo precedente.

Come si è anticipato, la discussione di Turing prende le mosse da un'analisi del ragionamento matematico. Questo, viene visto come il prodotto di due facoltà distinte: l'*intuizione* e l'*ingegnosità*.

"L'attività dell'intuizione", dice Turing, "consiste nell'asserire giudizi spontanei che non sono il risultato di una catena consapevole di ragionamenti"[328]. Ciò non significa che si tratti di un esercizio che ha un fondamento effimero, considerato che i giudizi che ne sono il prodotto "risultano essere spesso, seppur non invariabilmente, corretti". Un fatto, questo, che si rivela nel momento in cui risulta possibile giustificare tali giudizi mediante un argomento matematico, "che coinvolge anch'esso giudizi intuitivi", ma i quali risultano essere "meno esposti alla critica" di quelli originari.

È proprio in questa fase, quella del sostegno deduttivo all'intuizione, che interviene la seconda facoltà: "[l]'esercizio dell'ingegnosità in matematica consiste nell'aiutare l'intuizione mediante un opportuno ordinamento delle proposizioni, e forse di figure geometriche o disegni. È inteso che quando questi elementi vengono sistemati particolarmente bene, non si può dubitare seriamente della validità dei passaggi intuitivi richiesti"[329].

L'attenzione di Turing è chiaramente rivolta al ragionamento matematico così come esso si dà, informalmente, nella prassi della disciplina. Egli sottolinea tuttavia, come l'introduzione del formalismo della logica influisca su entrambe le facoltà individuate: sull'intuizione, dal momento che "la necessità di ricorrere [ad essa] viene largamente ridotta" (presumibilmente alla giustificazione dei soli assiomi) "dal dare vita a regole per eseguire inferenze che sono sempre intuitivamente valide"; sull'ingegnosità, perché questa "acquisisce una forma più definita", potendo essere legata alla scelta, tra le varie

[327] *Ibid.*, p. 194.
[328] *Ibid.*, pp. 208-209.
[329] *Ibid.*, p. 209.

possibilità ammesse dalle regole di inferenza, "di quali passi risultino essere i più proficui per l'obiettivo di dimostrare una certa proposizione"[330].

Quello che si è tentato di fare in campo fondazionale, Turing lo descrive proprio come il tentativo di portare agli estremi il ricorso al formalismo per inquadrare con maggior precisione le due facoltà alla base del ragionamento matematico:

> In epoca pre-gödeliana si è pensato, da parte di alcuni, che fosse possibile probabilmente portare avanti questo programma al punto da sostituire tutti i giudizi matematici intuitivi da un numero finito di queste regole. La necessità dell'intuizione sarebbe stata, allora, interamente eliminata.[331]

Non è difficile riconoscere, dietro questa formula, una delle aspirazioni fondamentali del programma finitista di David Hilbert.

L'indagine sulle logiche ordinali, come emerge dalle aspettative che nel testo Turing vi ripone, nasce piuttosto dalla consapevolezza che questo scopo, l'eliminabilità del concorso dell'intuizione, non è più perseguibile, nonostante che l'obiettivo che Turing si prefigge, ridurre l'intuizione al giudizio che una qualche espressione è una notazione per un numero ordinale, e la costruzione adottata, reiterare l'aggiunta ad un formalismo di un enunciato in esso indecidibile, possano apparire come due parti di un progetto analogo: "[s]i è cercato di vedere fino a che punto è possibile eliminare l'intuizione e lasciare solo l'ingegnosità".

A ben vedere, dice Turing, questa stessa costruzione alla base delle logiche ordinali sembra far trasparire che si sia piuttosto fatto il contrario, che si sia cioè "eliminato non l'intuizione ma l'ingegnosità", per aver assunto, come si è soliti fare nelle "discussioni metamatematiche", che di ingegnosità, cioè della capacità di costruire dimostrazioni matematiche stringenti, ve ne sia a disposizione "una dotazione illimitata": questa assunzione sta dietro l'idea di considerare i sistemi logici 'in estensione', ossia come "un metodo per enumerare le proposizioni dimostrabili", che trasforma così le dimostrazioni "in ricerche attraverso queste enumerazioni", e permette di "sostituire l'ingegnosità con la pazienza"[332].

Quando si passa, tuttavia, dalla metamatematica all'euristica, è opportuno evitare una simile "riduzione" e partire dal dato reale: il ragionamento matematico, sembra voler dire Turing, è costituito da due facoltà distinte, con

[330] *Ibid.*, p. 209.
[331] *Ibid.*, p. 209.
[332] *Ibid.*, p. 209.

compiti specifici ed entrambe ineliminabili (per quanto possa essere in certe circostanze un comodo artificio assumere il contrario). Proprio tale ineliminabilità, che è una delle conseguenze ineludibili dei teoremi di incompletezza, rende impraticabile il programma fondazionale descritto, pur lasciando aperta la ben più modesta possibilità di uniformare il modo arbitrario con il quale esse risultano essere 'distribuite' tra i matematici, definendone i compiti con maggior chiarezza[333].

Questo è il dato su cui si deve fondare, allora, la ricerca logica, rivolgendo l'attenzione a sistemi "non costruttivi", dei quali le logiche ordinali sono un esempio. Sistemi, cioè, nei quali alcuni passi sono intutivi, "riconoscere una formula come formula ordinale"[334], cioè un'espressione come un elemento di un sistema di notazioni, altri, "eseguire le conversioni"[335], sono meccanici.

Mutano di conseguenza anche i criteri per valutare gli obiettivi raggiunti dalle varie alternative di sistemi "non costruttivi":

> Che proprietà desideriamo che abbia una logica "non costruttiva" se se ne deve fare uso ai fini dell'espressione delle dimostrazioni matematiche? Vogliamo che essa mostri, in modo abbastanza chiaro, quando un passaggio richiede l'uso dell'intuizione, e quando esso è puramente formale. Lo sforzo basato sull'intuizione dovrebbe essere ridotto al minimo. Fatto più importante di tutti, deve essere al di là di ogni ragionevole dubbio che la logica porta a risultati corretti ogni qual volta i passi intuitivi sono corretti. È anche auspicabile che la logica sia adeguata all'espressione dei teoremi aritmetici, così che essa possa venir usata nelle discussioni metamatematiche.[336]

Da questo punto di vista, il giudizio di Turing sulle logiche ordinali Λ_H e Λ_P è la conseguenza diretta dei teoremi ad esse relativi: perché essendo, come si è visto, incompleta, con la logica Λ_H "non siamo in una posizione migliore che con un sistema logico costruttivo"; l'insoddisfazione per Λ_P "come fondamento non-costruttivo degli argomenti metamatematici", per quanto essa

[333]Il passo di Turing che si è discusso negli ultimi capoversi è quello che chiude, in [Tur39], p. 209. La necessità di una spiegazione che ne forza la struttura originaria, è motivata dal fatto che, così come è stato concepito da Turing, esso pone un problema di interpretazione aprendosi con una frase apparentemente auto-contraddittoria, la quale, testualmente, recita: "Nelle nostre discussioni, tuttavia, ci si è spinti all'estremo opposto e si è eliminato non l'intuizione, ma l'ingegnosità, e questo a fronte del fatto che il nostro scopo è stato perseguito nella medesima direzione. Si è cercato di vedere fino a che punto sia possibile eliminare l'intuizione, e lasciare solo l'ingegnosità".

[334]*Ibid.*, p. 210.

[335]Si ricordi che le logiche ordinali sono basate sul formalismo del λ-calcolo.

[336]*Ibid.*, p. 210.

sia in una certa misura una logica ordinale completa, discende invece dall'assenza di certezza riguardo al fatto che, mediante essa, "non si otterranno altro che risultati veri"[337].

Analisi di Turing dei risultati ottenuti a parte, ciò che della disamina del matematico britannico sull'essenza del ragionamento matematico ci preme sottolineare in questa sede, è però un aspetto diverso, che si comprende mettendo in relazione il lavoro di Turing sulle logiche ordinali con il suo contributo alla teoria della ricorsività e con gli scritti successivi dell'immediato dopoguerra.

Prima di rivolgere la nostra attenzione a questi ultimi, sarà bene fare qualche considerazione di 'metà percorso', sulle relazioni tra [Tur39] ed il lavoro dedicato ai "numeri computabili"[338].

3.1.3 Le macchine di Turing, le logiche ordinali e la mente umana

La letteratura critica dedicata all'opera di Turing, pare essere fortemente condizionata dall'idea che il lavoro del matematico britannico debba essere inquadrato in, e letto alla luce di, uno sfondo concettuale di tipo fortemente materialistico. Questo atteggiamento, come si è già avuto modo di discutere nella prima parte, ha condotto, innanzi tutto, ad una certa lettura del contributo di Turing del 1937, ed in particolare proprio di quell'argomento elaborato in favore della propria versone della Tesi che sancisce l'equivalenza tra il concetto informale di 'funzione calcolabile', ed il suo supposto corrispettivo rigoroso di 'funzione Turing-computabile'.

Non intendiamo, in questa sede, riportare una lista di opinioni al riguardo[339]: valga per tutte, l'interpretazione di Gödel della terza nota di [Göd90c] (della quale si è già detto e che si avrà modo di riprendere più avanti), che sembra attribuire a Turing la volontà di stabilire, con quell'argomento, un'equivalenza tra la mente umana e il modello astratto di macchina da lui definito[340]. Si è visto peraltro (cfr. §1.2), come sia possibile proporre un'interpretazione del passo cruciale, che permette di evitarne, e di considerare persino fuorviante, una lettura di tipo meccanicistico: ciò che Turing

[337] *Ibid.*, p. 210.

[338] Cioè [Tur36].

[339] Alcune posizioni sono analizzate, ad esempio, in [Col97], §2.5.

[340] Si è menzionato, affrontando la questione nella prima parte, che questa stessa interpretazione del passo di Gödel, è controversa (cfr §2.4.1).

intese mostrare analizzando direttamente il concetto intuitivo inteso, fu solo l'adeguatezza ad esso della definizione matematica da lui proposta.

Non si tratta tanto di stabilire la correttezza dell'una o dell'altra interpretazione. Questo viene fatto già, e meglio, altrove[341]. Vogliamo sforzarci piuttosto di assumere un punto di vista più neutrale nel proporre una visione d'insieme, che sia allo stesso tempo il più organica possibile, dei lavori di Turing.

Perché, e dovrebbe apparire chiaramente alla luce di quanto detto nel paragrafo precedente, la tesi di dottorato di Turing sulle logiche ordinali, pare porre dei seri problemi a quanti vogliano vedere in lui, fin dal 1937, uno strenuo sostenitore di un'ipotesi computazionale della mente. L'analisi del ragionamento matematico contenuta nel §11 di [Tur39], infatti, parla chiaramente di due componenti, l'intuizione e l'ingegnosità, rispetto alle quali l'adozione del formalismo denuncia limiti precisi che non permettono di andare oltre il tentativo di delinearne la natura e definirne i compiti.

Turing, in particolare, afferma, nel sottolineare il concorso ineludibile dell'intuizione, l'essenza "non costruttiva", e quindi non meccanica, del ragionamento matematico: siamo quindi di fronte, apparentemente, ad una sconfessione, piuttosto che all'adozione, di un'ipotesi meccanicistica.

Ciò conduce quanti leggano Turing attraverso la lente di un simile presupposto, ad adottare una qualche strategia al riguardo, ad esempio vedendo nell'indagine sulle logiche ordinali e, per coerenza, in [Tur36], come ad una prima fase nello sviluppo del pensiero di Turing, nella quale la convizione sulla natura computazionale della mente comincia a prendere forma, ma non è ancora sposata senza indugio.

Sembra questo, ad esempio, l'orientamento di Andrew Hodges laddove questi, nella biografia del matematico britannico, scrive:

> C'è comunque in questo lavoro di Turing [[Tur39]] la stessa ambiguità riscontrabile in *Computable numbers* [cioè [Tur36]]: la mente viene meccanizzata, ma al tempo stesso si accenna all'esistenza di qualcosa al di là della meccanizzazione. Qual'è il significato di tutto questo per la mente umana? A questo stadio Alan non ha ancora le idee chiare.[342]

Una simile attitudine, lascerebbe quindi supporre che l'adozione di una concezione meccanicistica della mente avvenga in modo inequivocabile negli

[341] Si sono già più volte citati al proposito, nella prima parte, i molti lavori di Wilfried Sieg.

[342] [Hod91], p. 192. *NdA*: il riferimento, anche per la traduzione, è all'edizione italiana del 1991 del libro di Hodges, come risulta dalla voce bibliografica.

scritti successivi, ed in particolare nei lavori immediatamente successivi la fine della seconda guerra mondiale.

È quanto ci proponiamo di verificare, a partire dal prossimo paragrafo, analizzando quelli che, tra i contributi tardi di Turing, ci sembrano, da questo punto di vista, essere i più rappresentativi.

3.2 Può una macchina imparare?

Della fortuna e delle imprese di Turing come criptoanalista al servizio dell'esercito britannico durante la seconda guerra mondiale, molto è noto.

Se ci si limita a guardare agli scritti dell'immediato dopoguerra senza una prospettiva storica come guida, non si trovano però che poche tracce di quell'esperienza e si riscontra invece, nonostante che non si abbia più a che fare con contributi di tipo squisitamente logico, un legame diretto con i lavori degli anni '30 del '900: avevamo lasciato Turing ad occuparsi dell'indagine intorno alle limitazioni dei sistemi formali, lo ritroviamo, in una conferenza alla London Mathematical Society del febbraio 1947, ad affrontare problemi molto simili seppur in un contesto diverso e rispetto ad obiettivi profondamente mutati. Il tentativo, è infatti quello di dare 'carne e sangue' al modello astratto di macchina da lui introdotto due decenni prima, e costruire un Motore Automatico Calcolatore, uno degli antesignani dei moderni elaboratori digitali[343].

Come risulta dal testo dell'intervento di Turing[344], gran parte degli sforzi del matematico britannico furono volti ad illustrare all'uditorio i problemi che emergono nel perseguire un simile progetto, e le possibili soluzioni ad essi. Non mancano tuttavia alcune considerazioni di carattere più generale, alle quali merita prestare attenzione.

Tra queste, quante permettono di far emergere in modo chiaro il legame tra la costruzione di un calcolatore e le proprie ricerche più teoriche in ambito logico, come nel caso del passo seguente:

> Qualche anno fa stavo indagando su ciò che può essere ora descritto come

[343]Con questo, non si intende affatto sminuire l'importanza che può aver ricoperto l'esperienza di Turing durante il conflitto mondiale, anche per tale proposito. Si vuole però evitare la ricostruzione di quella cornice, dal momento che il lavoro che si intende svolgere non è di tipo storico (per ogni informazione al riguardo, si rimanda invece al già citato [Hod91]).

[344][Tur86].

> la ricerca intorno alle possibilità e alle limitazioni teoretiche di macchine calcolatrici digitali. Presi in considerazione un tipo di macchina con un meccanismo centrale ed una memoria infinita contenuta su un nastro infinito. Questo tipo di macchina sembrava essere sufficientemente generale. Una delle mie conclusioni fu che l'idea di un processo 'secondo regole' [in originale "'*rule of thumb*' *process*"] e quella di un 'processo meccanico' fossero sinonimi. L'espressione 'processo meccanico' ovviamente significa un processo che potesse essere eseguito dal tipo di macchina che stavo considerando. Era essenziale per questi argomenti teoretici che la memoria dovesse essere infinita. Si può facilmente dimostrare altrimenti che la macchina può eseguire solo operazioni periodiche. Macchine come l'A[*utomated*]C[*omputing*]E[*ngine*] possono essere viste come versioni pratiche di questo stesso tipo di macchina.[345]

Il quesito sorge spontaneo: se sussiste una relazione così stretta tra le 'macchine di Turing' ed il calcolatore al quale egli intendeva dare vita, cosa ne è, rispetto a quest'ultimo, delle "limitazioni teoretiche" alle quali Turing era pervenuto nel corso delle indagini sulle prime? Su questo punto in particolare, si trova un accenno alla fine dell'intervento.

Nel tirare le fila della disamina, Turing si sofferma infatti su un'idea che, egli dice, accompagna spesso le ricerche nel settore della costruzione di macchine calcolatrici, e che conduce a ritenere che esse "eseguono solo quei processi che sono istruite a compiere". Ciò condiziona gli scopi della ricerca, dal momento che si ritiene conseguentemente che simili dispositivi possano essere impiegati esclusivamente come "schiavi", e siano adatti a "lavori tali che l'utente della macchina ha una conoscenza piena di ciò che sta succedendo in ogni istante"[346].

È ovvio che quest'idea soddisfi solo in parte Turing. Egli prospetta infatti la possibilità che la macchina possa essere programmata in modo tale da seguire inizialmente le istruzioni che gli sono state impartite, ma di essere in un secondo tempo in grado di agire su queste stesse istruzioni, cambiandole, e dunque conducendo autonomamente un compito che essa debba portare a termine. "Quando ciò accade", dice Turing, "ritengo che si debba essere obbligati a vedere la macchina come capace di intelligenza"[347].

Proprio sotto questo aspetto, le limitazioni teoriche del modello matematico potrebbero essere chiamate in causa, in quanto esse apparentemente

[345] [Tur86], pp. 87-88. *NdA*: la pubblicazione degli scritti di Turing in [Inc92], alla quale, per semplicità, facciamo riferimento, prevede due numerazioni di pagina per ogni contributo: quella originaria e quella, progressiva, della raccolta. Nelle citazioni ci riferiamo a quest'ultima.
[346] *Ibid.*, p. 103.
[347] *Ibid.*, p. 104.

permettono di costruire un argomento mediante il quale esprimere, “in forma più aggressiva”, la diffusa convinzione che vi sia una “contraddizione fondamentale” nell’idea di una macchina che possieda intelligenza.

L’argomento sarebbe il seguente:

> Per esempio, è stato dimostrato che rispetto a certi sistemi logici non ci può essere una macchina che sia in grado di distinguere le formule dimostrabili del sistema da quelle indimostrabili, cioè non esiste alcun test che la macchina può applicare per dividere le proposizioni in queste due classi con certezza. Quindi, se una macchina fosse costruita a tale scopo non riuscirebbe, in qualche caso, a fornire una risposta. D’altra parte, se con lo stesso problema si confrontasse un matematico, egli andrebbe alla ricerca e troverebbe nuovi metodi di prova, in modo tale da riuscire eventualmente a raggiungere una decisione relativamente ad una certa formula.[348]

Si tratta, per quanto ci è noto, della prima formulazione di un argomento che deve la sua maggiore notorietà alla pubblicazione, nel 1961, di un articolo da parte del filosofo di Oxford J. R. Lucas (al quale ne viene comunemente attribuita la paternità), e che comparirà nei contributi successivi del matematico britannico sotto la dicitura di “obiezione matematica” alla possibilità di una macchina pensante[349].

A quest’argomento, che tende ad evidenziare una differenza essenziale tra il procedere di una macchina, che segue pedissequamente le istruzioni del proprio programma, e quello dell’uomo, che invece pare muoversi, in modo del tutto indipendente da qualsiasi sistema di regole, alla ricerca di nuovi metodi di prova, Turing oppone due considerazioni.

La prima, è ispirata alle assunzioni a partire dalle quali si dimostrano i risultati matematici che l’argomento chiama in causa, e sui quali esso fonda la propria ‘morale’:

> In altre parole, se si ritiene che una macchina debba essere infallibile, allora essa non può essere anche intelligente. Ci sono molti teoremi matematici che dicono questo quasi con esattezza. Ma questi teoremi non dicono nulla riguardo a quanta intelligenza potrebbe essere esercitata se una macchina non avesse pretese di infallibilità.[350]

---

[348] *Ibid.*, p. 104.

[349] All’analisi dell’argomento di Lucas e della riformulazione di esso offerta di recente dal fisico e matematico inglese Roger Penrose, è dedicato il prossimo capitolo del lavoro. Rimane aperta la questione se quella posizione sia stata interamente elaborata da Turing, oppure se sia ispirata ad un’idea comunicata da altri al matematico britannico.

[350] *Ibid.*, p. 105.

Considerato che tutte le possibili versioni dei teoremi limitativi che possono essere indicati come il sostegno matematico dell'obiezione proposta si basano su una qualche assunzione di correttezza, occorre chiarire in quale senso, dietro la risposta di Turing, vi sia apparentemente una difesa di sistemi formali o macchine inconsistenti, ed in che modo ciò possa essere d'aiuto contro l'obiezione all'idea di macchine dotate di intelligenza.

Occorre notare, infatti, che l'idea dell'argomento in questione è più o meno la seguente: considerato che tra le proposizioni di un linguaggio formale ve ne sono alcune in particolare, le istanze delle quali non possono essere formalmente dimostrate o refutate in certi sistemi formali consistenti, una macchina alla quale venisse chiesto di applicare correttamente un test ad un enunciato opportunamente scelto, ad esempio un'istanza dell'enunciato indecidibile di Rosser (o della sua negazione), continuerebbe indefinitamente ad applicare la stessa, infruttuosa routine, senza mai giungere ad una risposta. Ed è questa reiterazione del medesimo processo, si noti bene, che costituisce l'essenza dell'obiezione mossa: un simile *modus operandi*, si dice, non ha un corrispettivo nell'atteggiamento del matematico umano, il quale, una volta compresa l'inefficacia di una certa strategia, si muoverebbe alla ricerca di metodi nuovi.

Ma supponiamo adesso che quella stessa macchina sia dotata, ad esempio, di un'istruzione sulla base della quale, quando un calcolo supera un certo lasso di tempo, entra in funzione un dispositivo che, in modo del tutto casuale, costringe la macchina a scegliere tra le due risposte possibili. In questo caso, ovviamente, la macchina in questione non applicherebbe affatto un test soddisfacente, se con esso si intende distinguere con certezza le proposizioni formalmente dimostrabili da quelle che non lo sono. Essa è anche, in un certo senso, inconsistente, dato che è perfettamente plausibile che possa dare due risposte diverse in relazione ad uno stesso enunciato. Ma in un senso proprio, la macchina non è *realmente* inconsistente, dal momento che essa esegue correttamente le istruzioni che le sono state impartite. È chiaro infatti, che non c'è alcuna contraddizione ad immaginare una macchina la cui unica risposta sia, ad esempio, l'enunciato '$0 = 1$'.

Ciò che conta, però, ai fini dell'obiezione contro la possibilità di macchine che esibiscano un comportamento intelligente, è che, rispetto ad un simile congegno, l'argomento elaborato sulla base di certi teoremi logici perderebbe parte della sua forza: fondato com'è, per come Turing lo riporta, sull'implausibilità *euristica* del fatto che alla base dell'intelligenza vi sia un procedere secondo regole, i suoi propugnatori non potrebbero applicarlo in modo diretto ad un dispositivo come quello descritto, il quale, certo, commette errori, ma si presta meno al rilievo che ogni macchina, di fronte a certi problemi, sia costretta a procedere sempre ed invariabilmente all'applicazione dello stesso ciclo di operazioni.

Una simile risposta potrebbe naturalmente venir considerata come largamente insoddisfacente.

È forse per questo che Turing aggiunge ad essa una seconda osservazione:

> Un matematico umano è stato sempre sottoposto ad una preparazione intensiva. Questa preparazione può essere vista come qualcosa di non dissimile dall'inserire tavole di istruzioni in una macchina. Non ci si deve aspettare quindi che una macchina possa ottenere grossi risultati nel costruire tavole di istruzioni per conto proprio. Nessun uomo aggiunge poi molto al corpo delle conoscenze, perché dovremmo aspettarci di più da una macchina? Sottolineando lo stesso fatto in modo diverso, la macchina deve poter avere un contatto con gli esseri umani in modo da potersi adattare ai loro standard.[351]

La possibilità di costruire un meccanismo in grado di commettere errori, dunque, non è che un primo, ma non decisivo passo per ribattere alla credenza diffusa che le macchine non possano, per la loro stessa natura, sviluppare un comportamento intelligente. Ma un'indicazione più basilare rispetto a questa posizione, la si ricava interrogandosi su come l'intelligenza umana si sviluppa: essa beneficia innanzi tutto del processo educativo al quale un individuo è sottoposto, e della possibilità di un confronto continuo con i propri simili. È solo dopo aver dimostrato che è impossibile riprodurre lo stesso processo nel caso di una macchina, allora, che si potrà essere convinti dell'esistenza di un ostacolo insormontabile nell'idea di un dispositivo meccanico con intelligenza.

Torniamo, a questo punto, per un attimo al concetto dal quale, nel presente paragrafo, si sono prese le mosse. Siamo partiti mettendo in evidenza il legame tra le indagini logiche di Turing nel corso degli anni '30, con le ricerche nel campo della neonata *computer science* appartenenti al decennio successivo. Il riflesso che le limitazioni alle quali soggiace il modello astratto di macchina, legato alle prime, hanno sulle possibilità che si aprono per il frutto delle seconde, porta Turing ad indicare la necessità di abbandonare il criterio della consistenza, da un lato, e a proporre l'idea di una sorta di 'percorso educativo' al quale sottoporre le macchine, dall'altro. Nonostante le similitudini di partenza tra le due direzioni di ricerca, si aprono in questo contributo delle prospettive che preludono all'adozione, nella pratica, di un concetto di 'macchina' che diverge da quello teorico.

[351] *Ibid.*, p. 105.

Sono queste stesse considerazioni ad essere al centro di uno scritto di Turing, ben più sistematico al proposito, dell'anno successivo.

3.2.1 Dispositivi intelligenti

Il rapporto del 1948 elaborato a conclusione dell'esperienza presso il National Phisycal Laboratory[352], non si limita a riprendere cursoriamente il tema sul quale si era chiusa la conferenza alla London Mathematical Society dell'anno precedente: ne fa piuttosto l'argomento principale del lavoro.

Turing riparte esattamente dal punto sul quale si era fermato in quell'occasione, e cioè da una critica di alcune delle ragioni addotte a sostegno dell'idea, che si rivela dietro certi "luoghi comuni" del linguaggio, secondo la quale la stessa espressione di 'macchina intelligente' sarebbe in realtà un ossimoro.

Alcune di queste ragioni ("Una mancanza di volontà ad ammettere la possibilità che il genere umano possa avere alcun rivale in campo intellettuale", oppure, "Una convinzione religiosa che ogni tentativo di costruire un simile apparecchio sia una sorta di irriverenza prometeica"[353]) sono "puramente emozionali", e perciò, secondo Turing, non possono che essere ignorate. Altre ("Il tipo estremamente limitato di macchina che è stato usato fino a tempi recenti"[354]), sono contingenti, nel senso che dipendono dallo stato della ricerca, e comunque possono essere avversate sulla base dei recentissimi esempi di calcolatori.

Infine, vi sono quelle obiezioni che hanno un fondamento matematico in certi teoremi logici:

> Recentemente, il teorema di Gödel ed i risultati ad esso legati [...] hanno mostrato che se si cerca di usare le macchine allo scopo di determinare la verità o la falsità dei teoremi matematici e non si intende tollerare risultati sbagliati occasionali, allora ogni macchina data non sarà in grado di dare in certi casi alcuna risposta. Dall'altro lato, l'intelligenza umana sembra essere in grado di scoprire metodi di potenza sempre crescente per affrontare tali problemi 'trascendendo' i metodi disponibili nella macchina.[355]

Mettendo insieme la risposta di Turing a quest'obiezione e a quella ulteriore, secondo la quale "[f]in tanto che una macchina è in grado di mo-

[352] Si veda [Hod91], pp. 476 e segg., al riguardo.
[353] [Tur48], p. 107.
[354] *Ibid.*, p. 107.
[355] *Ibid.*, p. 108.

strare intelligenza, questa deve essere vista come nient'altro che il riflesso dell'intelligenza del suo creatore", si ritrovano entrambe le componenti della posizione da lui espressa nella conferenza dell'anno precedente: fallibilità ed educazione.

Turing sottolinea infatti che:

1. "L'argomento tratto dal teorema di Gödel e da altri risultati [...] poggia in modo essenziale sulla condizione che la macchina non commetta errori. Ma questo non è un requisito per l'intelligenza"[356].

2. "Il punto di vista [...] secondo il quale l'intelligenza in un dispositivo meccanico sia solo il riflesso di quella del suo creatore è simile all'idea che il merito per le scoperte di un allievo debba essere conferito al suo insegnante"[357].

Dunque, se si abbandona l'idea che una macchina non possa commettere errori e se si garantiscono certe condizioni per poter guardare al rapporto tra la macchina e colui che l'ha materialmente creata come a quello tra insegnante e allievo, le obiezioni più razionali (e quindi, si potrebbe aggiungere, più ragionevoli), al progetto di creare un dispositivo meccanico intelligente, vengono meno.

È nella restante parte del lavoro che Turing cerca di affrontare sistematicamente le problematiche di una simile ricerca, ed è, come ci si potrebbe aspettare, soprattutto la questione dell'educazione di una macchina ad essere al centro della disamina.

Organizzare una macchina disorganizzata

Come emerge anche dalla conferenza del 1947, la valenza che Turing attribuisce al processo educativo deriva dal ruolo assolutamente essenziale che questo ricopre nel caso degli individui (umani), un fatto che egli riassume, in [Tur48], nel modo seguente:

> Sarebbe abbastanza ingiusto aspettarsi che una macchina, direttamente dalla fabbrica, sia in grado di competere alla pari con un laureato. Il laureato ha avuto contatti con gli esseri umani per venti o più anni. Questo contatto ha modificato il suo modello di comportamento per tutto quel lasso di tempo. I suoi insegnanti hanno cercato intenzionalmente di modificarlo.

[356] *Ibid.*, p. 108.
[357] *Ibid.*, p. 109.

> Alla fine di quel periodo, un gran numero di procedure standard saranno state sovraimposte all'originale modello del suo cervello. Queste procedure saranno note alla comunità nel suo insieme. Egli è quindi nella posizione di provare nuove combinazioni di queste procedure, di tentare piccole variazioni di esse, e di applicarle in modi nuovi.[358]

È dunque un lungo addestramento che sta all'origine di quel carattere, l'andare alla ricerca di nuove tecniche, che sembra distinguere in modo essenziale, come sostengono i propugnatori dell'obiezione basata sui teoremi di indecidibilità, l'uomo da una macchina.

Il problema principale che il tentativo di sottoporre una macchina ad un processo educativo simile pone, è quello di rendere possibile un'appropriata interazione con essa, in modo che si possa, innanzi tutto, dare vita a quelle modificazioni che ne sono il fulcro. È ciò che Turing, nel proprio lavoro, indica con il nome di "interferenza".

Vengono distinte due forme di interferenza[359]: quella, più estrema, di tipo "attrezzistico" (nel testo, *screwdriver interference*), mediante la quale si sostituiscono fisicamente certe parti della macchina, e quella legata all'attività di programmazione (*paper interference*) che consiste in un tipo di comunicazione con la macchina che è essenzialmente volta a modificarne le istruzioni (e quindi a mutarne il comportamento)[360].

Rispetto al processo educativo al quale un individuo viene sottoposto, occorre riconoscere allora, che "egli è soggetto ad una grande quantità di interferenza", al punto tale che quest'ultima "costituisce la regola piuttosto che l'eccezione"[361].

Ecco allora che il progetto iniziale, che è in modo vago descrivibile come il tentativo di dare vita ad una macchina che esibisca un comportamento intelligente mediante qualcosa che somigli ad un processo educativo, assume contorni più definiti:

> Se si sta cercando di costruire una macchina intelligente e si segue il modello umano nel modo più vicino possibile, si dovrebbe iniziare con una

[358] *Ibid.*, p. 118.
[359] Cfr. *ibid.*, pp. 115-116.
[360] Turing nota come la differenza tra i due tipi di interferenza sia trascurabile, fin tanto che ci si limita a considerare macchine universali come i calcolatori digitali, alludendo con ciò al fatto che, su dispositivi del genere, i quali, per loro natura, sono in grado di simulare ogni possibile processo meccanico, l'azione di macchine fisicamente diverse, ottenute cioè mediante il concorso della prima forma di interferenza, può essere simulata mediante un'appropriata opera di programmazione.
[361] *Ibid.*, p. 118.

> macchina che abbia una capacità limitata di eseguire operazioni elaborate o di reagire agli ordini (che prendono la forma di interferenza) in modo disciplinato. Applicando poi un'interferenza appropriata, in modo da imitare l'educazione, si dovrebbe sperare di modificare la macchina al punto da renderla affidabile nell'avere precise reazioni a certi comandi.[362]

In prima istanza, dunque, il programma di Turing prevede quindi la verifica del fatto che sia possibile, partendo da un dispositivo che appaia semplicemente come un'insieme di parti privo di organizzazione (ossia una macchina che non presenti una finalità precisa in funzione della quale essa sia stata costruita), giungere, al termine di un addestramento basato sull'interefenza e che sia il più simile possibile all'educazione di un individuo, ad un apparato in grado di eseguire certi calcoli in relazione ad input determinati.

Un simile programma, prevede quindi in particolare l'analisi delle strategie pedagogiche che vengono adottate nel caso degli esseri umani, ed il tentativo di implementarli, attraverso un'appropriata opera di programmazione, su un dispositivo meccanico. È questo il senso dell'esempio di "organizzazione" attraverso un sistema di 'Ricompense e Punizioni' che Turing fornisce, sistema nel quale egli individua uno dei metodi istruttivi più comunemente messi in atto nell'educazione di uno 'scolaro' umano.

La macchina descritta da Turing ed il 'procedimento educativo' scelto[363], in quello che egli riporta come un esperimento da lui effettivamente portato a termine, può essere descritto in breve ed in termini semplici, e non sarà perciò vano, per capire un pò meglio l'idea che anima la disamina del matematico britannico, ripercorrerne le linee essenziali.

Il dispositivo meccanico M, che viene sottoposto all'esperimento, non ha bisogno di essere descritto nei dettagli. È sufficiente sapere che esso è dotato di una serie di unità di memoria (o istruzioni di partenza) $M_1, \ldots, M_h$, che possiede un dispositivo C mediante il quale operare scelte casuali, e che è in grado di compiere due tipi di azioni: certe operazioni esternamente 'visibili' $A_1, \ldots, A_k$, oppure delle operazioni interne sulle istruzioni di partenza. La macchina è in comunicazione, inoltre, con un utente esterno, l'*educatore*, il quale può fornire tre tipi di impulsi ad M: uno *stimolo sensoriale* S (è un impulso 'neutro'), una *ricompensa* R ('positivo') ed una *punizione* P ('negativo')[364].

[362] *Ibid.*, p. 118.

[363] Cfr. *ibid.*, pp. 122-125.

[364] In quale senso gli impulsi debbano essere considerati neutri, positivi o negativi, diverrà chiaro a descrizione dell'esperimento completata.

Il sistema elaborato da Turing può essere descritto mediante due tabelle che indicano, rispettivamente, la *configurazione* della macchina, tabella nella quale si associa ad ogni *stato* o *situazione* possibile (ognuno di essi essendo indicato mediante un elemento della successione $1, 2, \ldots, n$), l'*istruzione esterna* che determina il comportamento della macchina, ed il *carattere* di M, dove ad ogni stato si associa la *reazione* della macchina[365].

Più precisamente, la tabella 1 di configurazione, che è stabilita in anticipo e che non può essere cambiata, avrà l'aspetto seguente:

stato	*istr. esterna*
1	c_1
2	c_2
$\vdots$	$\vdots$
n	c_n

dove, per ogni $1 \leq i \leq n$, c_i ha la forma di un'istruzione che recita "Adotta il dato contenuto nella j-esima unità di memoria" (dove, quindi, $1 \leq j \leq h$), oppure è del tipo "Reagisci allo stimolo sensoriale", oppure, infine, c_i è un'istruzione che stabilisce "Ascolta l'educatore".

La tabella 2, che descrive il carattere della macchina, ha un aspetto del tutto simile:

stato	*reazione*
1	r_1
2	r_2
$\vdots$	$\vdots$
n	r_n

In questo caso, però, il contenuto è determinato dalla tabella 1: se infatti, per un certo $1 \leq i \leq n$, c_i è dei primi due tipi descritti ("usa la j-esima unità di memoria" o "reagisci allo stimolo"), la reazione corrispondente r_i è determinata ed ha la forma di una coppia $< a_i, s_i >$, dove a_i è un'azione (una di quelle 'visibili' oppure l'operazione su un'unità di memoria), e s_i è lo stato successivo. Anche il contenuto di queste coppie viene deciso in partenza e non è soggetto a cambiamento.

[365]Per ragioni espositive, di chiarezza e concisione, facciamo ricorso ad una terminologia che solo in parte ricalca quella utilizzata da Turing.

Fin tanto che si è in casi del genere, dunque, la macchina reagisce in modo 'tradizionale', eseguendo cioè operazioni precise in corrispondenza a determinati input.

Se però c_i ha la terza forma descritta ("ascolta l'educatore"), allora la reazione r_i della macchina può essere: I ("indeterminata"), $T0$ ("tenta la soluzione '0'"), $T1$ ("tenta la soluzione '1'"), $D0$ ("adotta la soluzione '0'") oppure $D1$ ("adotta la soluzione '1'"), dove le 'soluzioni' '0' e '1' indicano, in un qualche modo ben definito ed in funzione dello stato i, lo stato successivo[366]. Le soluzioni, come si sarà notato, vengono scelte o in forma di tentativo ($T0, T1$), o in forma più definitiva ($D0, D1$).

È chiaro che ciò costituisce l'aspetto cruciale dell'educazione della macchina per ricompense e punizioni.

Infatti, nel caso in cui la voce nella tabella 2 sia I, M fa la sua scelta (in forma $T0$ o $T1$) mediante il dispositivo C di scelta casuale; l'azione dell'educatore, attraverso l'impulso di ricompensa R o quello di punizione P, fa sì che, in presenza di una reazione di tipo $T0$ o $T1$ e di un impulso R, M adotti la soluzione in forma definitiva, cambiando la riga opportuna della tabella 2 da $T0$ o $T1$ a $D0$ o $D1$ rispettivamente (il 'carattere' della macchina si 'stabilizza'); in presenza di una reazione dello stesso tipo seguita però ad una punizione P, M è indotta a sostituire invece la soluzione adottata con l'indeterminata I (M viene così 'corretta').

Quanto detto dovrebbe essere sufficiente per illustrare meglio l'idea di Turing, che prende le mosse da macchine il cui comportamento sia in parte determinato e immodificabile, e che sia invece per altri versi indeterminato e largamente passibile di mutamento grazie al concorso dell'utente esterno, per vedere se sia possibile giungere, per mezzo di un''interferenza' appropriata e modellata su una delle possibili strategie educative adottate nei confronti di un individuo, ad un'organizzazione più compiuta.

Rispetto al sistema di 'Ricompense e Punizioni', adottato in relazione ad una macchina del tutto simile a quella descritta[367], Turing dice due cose:

1. di essere riuscito, almeno sulla carta[368], "a trasformare [*organizing*] una

[366] Nel caso esaminato da Turing, ad esempio, se i è lo stato presente, '0' e '1' sono, rispettivamente, il resto delle divisioni $\frac{2i}{n}$ e $\frac{2i+1}{n}$.

[367] Il cambiamento principale (cfr. *ibid.*, p. 123.) riguarda l'inserimento di una memoria esterna, in forma di nastro, sulla quale il dispositivo esegue le operazioni 'visibili'.

[368] Il commento di Turing non si riferisce infatti ad un dispositivo meccanico *reale* (cioè un aggregato fisico di circuiti e relé), ma alla descrizione, 'su carta', di una macchina del

simile macchina in una macchina [di Turing] universale"[369];

2. che il risultato ottenuto è comunque da considerarsi "insoddisfacente", in quanto "non sufficientemente analogo al tipo di processo mediante il quale si insegnerebbe davvero ad un bambino"[370].

L'uomo è una macchina complessa?

Compatibilmente con l'insoddisfazione manifestata da Turing e derivante dalla seconda delle osservazioni riportate, il sistema modellato sull'educazione mediante ricompense e punizioni mostra che il programma del matematico britannico è, almeno in parte, praticabile: è possibile cioè educare una macchina, nel senso di portare un dispositivo da un livello di organizzazione molto basso, nel quale il comportamento può essere anche in larga parte indeterminato, ad livello decisamente più elevato di organizzazione, del tutto analogo a quello di una macchina di Turing usuale.

Questa parte del contributo di Turing si basa su un'idea precisa: utilizzare il caso del processo educativo così come esso si dà in relazione agli esseri umani, come modello per l'addestramento delle macchine. Quest'idea, almeno per come l'abbiamo vista svilupparsi fino a questo momento, non è necessariamente legata ad una concezione secondo la quale la sua praticabilità si fonda sul fatto che l'uomo *è* una macchina, e che quindi l'educazione a cui egli va incontro nel corso della propria vita è del tutto analoga a quello che si andrà ad adottare sui dispositivi di nostra costruzione.

Tuttavia, è proprio a questa concezione che sembrano alludere certe considerazioni di Turing che non abbiamo ancora riportato:

> Crediamo che ci siano grandi aree del cervello, principalmente nella corteccia, la cui funzione sia largamente indeterminata. Nel neonato queste aree non hanno grandi effetti: i loro effetti non sono coordinati. Nell'adulto esse

genere come successione di istruzioni (quello che egli indica come una *paper machine* nelle pagine iniziali del lavoro - cfr. *ibid.*, p. 113). Si tratta cioè di una descrizione del tutto analoga a quella delle usuali macchine di Turing (propriamente, una *paper machine* non è altro che un uomo che esegue il calcolo al posto della macchina, seguendone le istruzioni indicate). Si sarà intuito da ciò che si è detto, come una simile descrizione, nel caso del tipo di dispositivo considerato da Turing, sia del tutto plausibile. L' 'organizzazione', in questo caso, non è altro che il calcolo (sempre su carta) dei cambiamenti indotti dall'esecuzione delle varie istruzioni e dal sistema di ricompense e punizioni.

[369] *Ibid.*, p. 123.

[370] *Ibid.*, p. 124.

> assumono effetti notevoli e finalizzati: la forma di questi effetti dipende dall'addestramento nell'infanzia. Un'ampia reminiscenza del comportamento casuale dell'infanzia permane nell'adulto.[371]

L'osservazione di Turing ha effettivamente il sapore della novità, rispetto a quelle viste fin qui, dal momento che essa è il preludio per una connessione, più profonda di un semplice paragone di comodo, tra l'uomo e la macchina. Ciò sarebbe sufficiente per garantire all'obiettivo che egli si è prefisso tutto un altro significato.

Un significato, del quale, peraltro, Turing pare essere pienamente consapevole:

> Tutto questo suggerisce che la corteccia cerebrale del neonato sia una macchina priva di organizzazione, la quale può essere organizzata mediante un addestramento basato sull'interferenza. L'organizzazione potrebbe consistere nella trasformazione della macchina in una macchina [di Turing] universale o qualcosa del genere.[372]

Così formulata, tuttavia, l'ipotesi in questione sarebbe largamente insoddisfacente dal momento che, Turing nota:

> Ciò significherebbe che l'adulto obbedirà ad ordini dati in un linguaggio appropriato, anche se fossero molto complicati; sarebbe privo di buon senso, e obbedirebbe agli ordini più ridicoli senza battere ciglio. Quando questi ordini fossero stati eseguiti egli cadrebbe in uno stato comatoso oppure obbedirebbe a qualche ordine permanente, come quello di mangiare.[373]

Se si pretende di associare ad un'esperimento come quello di cui si è detto, l'idea che la sua ragion d'essere si fondi su una stretta analogia tra l'uomo allo stadio iniziale del processo educativo, e la macchina dalla quale si prende le mosse, ci si scontra con le differenze innegabili che sussistono tra l'esito dell'un processo, l'individuo adulto, e quello del procedimento sul primo modellato, la macchina di Turing universale. Una simile eventualità suggerisce quindi che, di per sé, il modello di macchina di Turing non sia completamente soddisfacente.

È pur vero, però, che, conseguenze implausibili come quella indicata nella citazione precedente a parte, Turing appare convinto che vi sia qualche elemento di verità nel paragone tra la corteccia cerebrale e una macchina priva

[371] *Ibid.*, p. 120.
[372] *Ibid.*, p. 120.
[373] *Ibid.*, p. 120.

di organizzazione, e tra il passaggio dall'infanzia alla maturità e quello da una macchina disorganizzata ad una macchina universale. Egli infatti ritiene che una volta compresi i passaggi implicati in quest'ultimo percorso, ci si troverebbe "in una situazione decisamente migliore per considerare come il processo organizzativo avrebbe potuto essere modificato per dare vita ad un tipo di mente più vicino alla norma"[374].

È chiaro che i paralleli in questione, sembrano soprattutto far trasparire un approccio ben più sbilanciato in favore di una concezione di tipo meccanicistico, di quanto l'esposizione in forma di programma di ricerca vorrebbe denunciare.

Tuttavia, è anche vero che, fino a questo punto, Turing si è occupato solo di una parte, seppur importante, dell'intero progetto: passare da una macchina priva di scopo, ad un dispositivo che è in grado di eseguire alla perfezione qualsiasi operazione che possa essere descritta in termini di regole meccaniche.

Su quanto resterebbe da fare, anche una volta eliminati gli strascichi dell'insoddisfazione sul modo attraverso il quale i risultati positivi nella prima direzione possono essere ottenuti, Turing torna alla fine del contributo.

3.2.2 Disciplina e Iniziativa

Sottoponendo una macchina costruita senza un intento preciso ad un opportuno addestramento, è plausibile ritenere, sulla base della riuscita dell'esperimento con il sistema di 'Ricompense e Punizioni', che si riesca a dare vita ad un dispositivo che sia un'equivalente pratico di una macchina di Turing.

Il commento da parte del matematico britannico ad una simile eventualità, nel quadro complessivo del progetto di costruire una macchina intelligente, è il seguente:

> Se la mente priva di addestramento del neonato deve mutarsi in una intelligente, essa deve acquisire sia disciplina che iniziativa. Fino a questo punto si è preso in considerazione solo la disciplina. Convertire un cervello o una macchina in una macchina universale, rappresenta la forma più estrema di disciplina. Senza qualcosa di questo tipo non è possibile mettere a punto un'appropriata comunicazione. Ma la disciplina è certamente non sufficiente, di per sé, per produrre intelligenza. Ciò che occorre in aggiunta lo chiamiamo iniziativa. Quest'affermazione avrà il valore di una definizione. Il nostro

[374] *Ibid.*, p. 120.

> scopo è quello di scoprire la natura di questo residuo così come esso si presenta nell'uomo, e di provare a copiarlo su una macchina.[375]

Per quanto rappresenti un passo essenziale, la trasformazione di una macchina priva di organizzazione in un dispositivo finalizzato a scopi determinati, anche qualora venga depurata di quegli aspetti che la rendono ancora non pienamente soddisfacente, non sarebbe che una parte dell'impresa. Ciò che resta, a dire il vero, pare ben più impegnativo: perché non solo si tratta di dare corpo ad un concetto, quello di 'iniziativa', che appare a tutti gli effetti come una scatola vuota (Turing si riferisce in buona sostanza, con quell'espressione, a ciò che resta dall'intelligenza togliendo la disciplina), ma occorre anche svelarne la natura e cercare un modo per implementare una simile caratteristica su una macchina.

Si comprende così il motivo per cui le indicazioni di Turing per una ricerca che sia volta a scoprire l'elemento residuo dell'intelligenza umana, siano decisamente meno sistematiche rispetto al resto del contributo.

Egli allude infatti a due metodologie d'indagine. L'una, che consiste nel prendere le mosse da una macchina perfettamente disciplinata, per così dire, come un calcolatore digitale, e nel cercare di "tirar fuori da essa un pò di iniziativa" attraverso una modalità di impiego che faccia sì che la macchina debba operare sempre più 'scelte' e prendere 'decisioni' con sempre maggior autonomia.

In un altro senso, si dovrebbe invece partire da una macchina disorganizzata cercando di ottenere, ad un tempo, tanto disciplina quanto iniziativa.

Quel che di maggiormente insoddisfacente c'è in questi suggerimenti di Turing, è il fatto che essi appaiono come tentativi 'alla cieca': si tratta di cercare di fare emergere una componente distintiva di un comportamento intelligente, per la quale non si è in grado di offrire una definizione ragionevole e che possiamo solo sperare, quindi, di essere in grado di riconoscere qualora essa si 'manifesti' attraverso le operazioni eseguite da una macchina.

Forse è per questo motivo che Turing decise di chiudere il proprio resoconto, abbozzando un'analisi del problema posto dalla ricerca dell'inziativa:

> Una sorta di problemi molto tipici che richiedono qualche forma di iniziativa, consiste in quelli della forma 'Trova un numero n tale che ...'. Questo tipo, copre una grande varietà di problemi. Per esempio, sono riducibili a quella forma quesiti come 'Vedi se riesci a trovare un modo di calcolare la funzione che ci metterà in condizione di ottenere i valori per argomenti ...

[375] *Ibid.*, p. 125.

> con un'accuratezza ... in un tempo ... per mezzo del [calcolatore digitale]...', dal momento che il problema è chiaramente equivalente a quello di trovare un programma da implementare sulla macchina in questione, ed è semplice mettere in corrispondenza i programmi con i numeri interi positivi in modo tale che, dato il codice numerico o il programma, l'altro sia semplice da trovare. Non dovremmo sbagliare di molto, dati i tempi, se assumessimo che tutti i problemi siano riducibili a questa forma. Si dovrà pensare nuovamente a questo, nel momento in cui verrà fuori qualcosa che non sia in modo ovvio di questa forma.[376]

L'elemento distintivo dell'intelligenza umana, l'iniziativa, si manifesta nella risoluzione di problemi. Prendiamo allora in considerazione, dice Turing, la classe più generale di essi: si tratta di quella costituita da problemi numerici di tipo esistenziale, dal momento che gran parte di quesiti apparentemente diversi, possono essere in realtà ricondotti ai primi (ad esempio, mediante un'opportuna procedura di aritmetizzazione).

Il caso a cui Turing ricorre, quello di un problema relativo alla ricerca di un programma per calcolare una funzione, acquisisce ad esempio, sotto una trasformazione di questo tipo, la forma di un problema Σ_3^0, ossia di un enunciato:

$$\bigvee z \bigwedge x \bigvee y T_1(z,x,y)$$

dove T_1 è il noto predicato di Kleene.

Ora, se sono i problemi, ed in particolare i problemi esistenziali di questo tipo, che richiedono iniziativa, in una qualche forma, occorre chiedersi: quali strategie si possono adottare per la loro soluzione? Turing ne distingue tre tipi: quella *intellettuale*, quella *genetica* e quella *culturale*.

La prima, è una strategia che appare essere sufficiente a risolvere "un certo numero di casi", ossia certe classi di problemi esistenziali:

> Iniziando con un [calcolatore digitale] si inserisce innanzi tutto in esso un programma che corrisponda alla formulazione di un sistema logico (come i *Principia Mathematica* di Russell). Esso non determinerebbe il comportamento della macchina in modo completo: ai vari istanti sarebbe possibile più di una scelta relativa al passo successivo. Si potrebbe fare in modo, tuttavia, di metter in ordine tutte le possibili combinazioni di scelte, e andare avanti fino a che la macchina non dimostra un teorema, il quale, per la sua stessa forma, possa risultare essere una soluzione al problema. Ciò potrebbe essere visto come una conversione del problema originale in un altro della stessa forma. Invece che cercare tra i valori della variabile originaria n, si ricerca tra i valori di qualcosa d'altro.[377]

---

[376] *Ibid.*, p. 126.
[377] *Ibid.*, p. 126.

La descrizione di Turing della metodologia di indagine di tipo intellettuale, sembra compatibile con ciò che è noto ai logici come 'l'estrazione di una funzione da una derivazione formale'. Si tratta, infatti, in un caso semplice, data cioè una formula ψ dimostrabile in un certo sistema formale con $\psi :\equiv \forall x \exists y \phi$ (nella quale ϕ è, ad esempio, una matrice primitiva ricorsiva), di trovare, a partire dalla prova di ψ, un termine t_x (parametrico in x) ed una dimostrazione $\mathcal{D}$ del fatto che $\phi(t_x)$ vale. Il problema, così affrontato, si bipartisce: da un lato, c'è la questione dell'estrazione di un termine da una dimostrazione di un enunciato esistenziale (componente che pone, ovviamente, il problema di trovare una procedura efficiente per tale compito), dall'altro, la generazione, o meglio la ricerca, tenendo presente l'impostazione di Turing del passo precedente, di una prova del fatto che il termine costruito soddisfa la matrice dell'enunciato esistenziale.

In relazione quindi all'esempio di Turing, quello di enunciati di tipo Σ^0_3, la questione è: come ottenere informazioni sui termini rappresentanti funzioni dimostrabilmente totali in un certo sistema d'assiomi (ad esempio quello di *Principia mathematica*) dalle dimostrazioni formali?[378]

Si capisce così in che senso, in questo modo, si ottiene una soluzione del problema originario passando attraverso una sua radicale trasformazione: si tratta, cioè, di spostare l'obiettivo della ricerca da un elemento dell'insieme degli indici di funzioni ricorsive, come nell'esempio del problema Σ^0_3, a quello dei termini e di certe dimostrazioni formali di enunciati i quali, "per la loro stessa forma", siano una soluzione del problema iniziale.

Trasformazioni di un quesito di tal genere, quindi, per Turing, offrono un primo esempio di strategie che richiedono iniziativa.

Il secondo esempio, è rappresentato dalle ricerche di tipo *genetico*:

> Esiste un'indagine genetica o evolutiva mediante la quale si va in cerca

[378]Il tentativo di dire qualcosa di più preciso ed esauriente sulla questione, ci condurrebbe inevitabilmente fuori tema. Ci limitiamo a ricordare come le strategie per la risoluzione di questo tipo di problemi sono molteplici. Tra esse, quella che si basa sul teorema di Herbrand il quale, nella versione ristretta ad enunciati Σ^0_1, dice che se $\exists x \phi$ è dimostrabile in un sistema logico e ϕ è una matrice priva di quantificatori, allora esiste una successione di termini $t_0, \ldots, t_n$ e una dimostrazione della formula $(\phi(t_0) \vee \ldots \vee \phi(t_n))$. Questo risultato viene utilizzato in combinazione con altri (ad esempio l'eliminazione delle cesure), per la risoluzione di casi complessi. Alcuni accenni alla problematica sono contenuti in [Sch77]. Due lavori recenti dedicati al tema sono, ad esempio [Sie91a] e [Sie91b], il secondo dei quali, oltre a delineare un quadro delle tematiche filosofiche coinvolte, offre la possibilità anche di ricostruire il contesto storico nel quale la problematica si è affacciata ed è venuta maturando, oltre alle connessioni principali con altre aree di ricerca.

> di un combinazione di geni, secondo il criterio della valenza ai fini della sopravvivenza. Il notevole successo di questa ricerca, conferma in qualche modo l'idea che l'attività intellettuale consiste principalmente di vari tipi di ricerche.[379]

L'idea di Turing è meno chiara in questo caso. Tuttavia, se si assume, come è plausibile fare sulla base dell'idea che non esistono problemi importanti che non abbiano una natura esistenziale, e quindi che l'oggetto d'indagine originario sia un problema del tutto analogo a quelli in relazione ai quali porta talvolta ad un successo la ricerca di tipo intellettuale descritta poc'anzi, si ottiene quest'affascinante quadro: quando il primo tipo di ricerca fallisce, si mette in moto un processo evolutivo il cui obiettivo è quello di dare vita ad una struttura genetica migliore in fatto di garanzia di sopravvivenza, e quindi più capace in termini di risoluzione di problemi. Ancora una volta, l'inziativa si cela così dietro una trasformazione dell'obiettivo originario di un quesito di partenza.

A questi due tipi di indagine, occorre, secondo Turing, aggiungerne almeno una terza, di matrice *culturale*:

> Come ho già detto, l'uomo isolato non sviluppa alcuna capacità intellettuale. È necessario per lui essere immerso in un ambiente di altri uomini, dei quali egli assorbe le tecniche durante i primi vent'anni della propria vita. Egli potrebbe forse condurre un pò di ricerca per proprio conto e fare poche scoperte che vengono passate ad altri uomini. Da questo punto di vista la ricerca di nuove tecniche [di soluzione di problemi] deve essere vista come portata avanti dalla comunità nel suo complesso, piuttosto che da individui.[380]

L'osservazione di Turing, sembra voler rafforzare lo spunto che lo ha condotto a proporre il criterio dell'educazione di una macchina come il solo ragionevole per far emergere intelligenza da dispositivi meccanici: il principio secondo il quale una macchina deve essere messa in comunicazione con una comunità di agenti, e deve, perciò, essere messa nelle condizioni di imparare, da questi, certi metodi di prova, è di per sé un valore, essendo lo stesso principio alla base dell'emergere di quella componente essenziale della ragione umana che è l'iniziativa. Una simile idea, sembra quindi essere il miglior sostegno per l'eventualità, che risulterebbe altrimenti poco fondata, che la realizzazione del progetto che Turing ha in mente possa, pur in assenza di

[379] *Ibid.*, p. 127.
[380] *Ibid.*, p. 127.

una definizione adeguata dell'intelligenza in ogni suo aspetto, avere gli esiti sperati, ossia la costruzione di una macchina pensante.

Ciò completa le (poche) indicazioni che Turing fornisce relativamente a quello che appare essere forse il problema principale del progetto di ricerca al quale è dedicato il contributo che si è preso in esame.

È importante tuttavia, per riconnettersi anche a quanto detto al termine del paragrafo precedente a proposito di una presunta 'impronta meccanicista' nelle riflessioni di Turing, soppesare accuratamente, comunque, le sue prese di posizione sulla duplice natura dell'intelligenza.

Se infatti, come si è visto, certe considerazioni di Turing sulle similitudini tra la corteccia cerebrale e una macchina disorganizzata, e tra il processo di organizzazione di quest'ultima e quello di crescita dell'individuo, possono far pensare ad una precisa motivazione cognitiva alla base del progetto per la realizzazione di una macchina intelligente, la distinzione tra 'disciplina' e 'iniziativa' sembra mescolare nuovamente le carte in tavola.

Tanto che potrebbe apparire lecito proporre un confronto con l' 'intuizione' e l' 'ingegno' che Turing aveva indicato, al termine del lavoro dedicato alle logiche ordinali, come le due facoltà alla base della conoscenza matematica.

Di certo, c'è che l'ambizione di fondo appare diversa. Con la disciplina e l'iniziativa Turing intende le due caratteristiche che riconosciamo essere all'origine di un comportamento intelligente, e che quindi possono esserne considerate le condizioni fondamentali: la capacità di eseguire certi compiti secondo regole condivise, nell'un caso (capacità della quale le macchine di Turing rappresentano la forma "più estrema"), quella di sviluppare una certa indipendenza dal contesto culturale e di andare alla ricerca di idee nuove, nell'altro. Non è chiaro quindi se, con esse, sia lecito parlare di due 'facoltà' nello stesso senso nel quale Turing individua i due aspetti peculiari all'origine della conoscenza matematica.

Ciò che il confronto dovrebbe permettere di discutere, è se si nasconda, nel passaggio da uno scritto all'altro (e quindi dall'ambito teorico a quello pratico, ambiti dei quali quei contributi sono rappresentanti), un mutamento d'opinione da una situazione nella quale Turing parla esplicitamente di un fondamento 'non costruttivo' del ragionamento matematico, considerazioni che sembrano legarsi ad una concezione non computazionale della mente umana, ad una posizione di tipo esattamente opposto.

Ebbene, non ci pare, fin qui, di aver trovato un fondamento indubitabile per una simile interpretazione.

Perché è pur vero che il progetto di Turing, di per sé, può dare adito ad una speculazione di tal genere: è possibile sostenere, ci si potrebbe chiedere, la praticabilità della costruzione di una macchina che esibisca un comportamento intelligente, senza partire dal presupposto che alla base di quest'ultimo vi sia un fondamento di tipo meccanico (anche in un senso lato del termine)?

Tuttavia si dovrebbero anche notare due specificità dell'approccio di Turing al problema:

1. che proprio la praticabilità di quel programma viene sostenuta in relazione ad un tipo di macchina che è profondamente diverso da una macchina di Turing: se quest'ultima è il prototipo dell'individuo perfettamente disciplinato, una macchina intelligente, per la natura stessa di ciò che chiamiamo intelligenza, deve essere *qualitativamente distinta* da quel modello;

2. che il successo del programma dipende esclusivamente dall'indagine intorno all'iniziativa: dal disvelamento della sua *natura* da un lato, dalla nostra capacità di riprodurla su una macchina, qualora questo sia possibile, dall'altro.

Le conseguenze di ciò, verranno riprese e discusse nell'ultimo paragrafo del presente capitolo, subito dopo l'analisi di un altro fondamentale contributo di Turing del 1950.

3.3 Le macchine alla prova del Gioco dell'Imitazione

Il lavoro di Turing del 1948, si chiude con un'osservazione che è solo apparentemente estrinseca rispetto ai problemi di tipo procedurale che occorre affrontare e risolvere nel perseguire il progetto di una macchina intelligente:

> Il punto fino al quale guardiamo a qualcosa come ad un ente che si comporta in modo intelligente, è determinato sia dal nostro stato mentale e dall'addestramento [ricevuto], sia dalle proprietà dell'oggetto che si considera. Se siamo in grado di spiegare e predire il suo comportamento, oppure se sembra esserci poca pianificazione dietro esso, si ha ben poca tentazione ad immaginare intelligenza. Quindi, è possibile che in relazione ad uno stesso oggetto un uomo lo consideri intelligente ed un altro no; il secondo uomo ne avrebbe scoperto le regole di comportamento.[381]

[381] *Ibid.*, p. 127.

La riflessione di Turing, richiama alla mente quei luoghi nei quale il matematico britannico sottolinea come il progetto di una macchina intelligente sia, nella percezione comune, minato alla base dall'atteggiamento che vede, nell' 'agire secondo regole rigide' o 'in modo meccanico', delle perifrasi per esprimere l'assenza di intelligenza. In questo senso, il passo succitato pone una questione reale, legata alla verifica, nel senso scientifico del termine, dei risultati ottenuti: una volta che si ritenga di essere riusciti nell'intento e di avere cioè dato vita ad un dispositivo che esibisce, in una qualche misura, un comportamento intelligente, occorre anche essere in grado di proporre un metodo per sottoporre a verifica l'ipotesi, in modo da essere al riparo dall'intrusione nel giudizio di qualsiasi elemento di natura psicologica o pregiudiziale.

Turing allude, come possibile soluzione in questo senso, ad un esperimento da lui effettivamente condotto, o meglio ad una 'versione idealizzata" di un esperimento da lui effettuato:

> Non è difficile concepire una macchina su carta[382] che sia in grado di giocare una partita di scacchi in modo non malvagio. Si prendano ora tre uomini A, B, C come soggetti dell'esperimento. A e C sono giocatori di scacchi piuttosto scarsi, B è l'operatore che lavora con la macchina (perché sia in grado di giocare sufficientemente veloce, è opportuno che egli sia tanto un matematico quanto uno schacchista). Si usano due stanze dotate di qualche apparecchio per comunicare le mosse, e viene giocata una partita tra C e A o la macchina. C potrebbe trovare difficile dire contro chi egli stia giocando.[383]

Siamo di fronte ad una prima, abbozzata descrizione di quello che diverrà noto, a seguito della pubblicazione nell'ottobre del 1950 sulla rivista *Mind* dell'articolo dal titolo "Macchine calcolatrici e intelligenza", come il Test di Turing.

3.3.1 La questione dell'intelligenza

Con lo stile chiaro e diretto che gli è proprio, Turing riassume in modo rimarchevole nell'attacco del lavoro i termini del problema che egli vuole porre all'attenzione del lettore:

> Propongo di considerare la questione, 'Può una macchina pensare?'. Si dovrebbe innanzi tutto prendere le mosse da una definizione del significato dei termini 'macchina' e 'pensare'. Le definizioni dovrebbero essere inquadrate in modo tale da riflettere il più possibile l'uso comune delle parole, ma

[382] *NdT*: si veda quanto detto al proposito alla nota 368.
[383] *Ibid.*, p. 127.

> un simile atteggiamento è pericoloso. Se il significato delle parole 'macchina' e 'pensare' deve essere trovato sulla base di un esame di come esse vengono comunemente utilizzate, è difficile sfuggire alla conclusione che il significato e la risposta alla questione 'Può una macchina pensare?' debbano essere indagati mediante una ricerca statistica tipo sondaggio Gallup. Ciò è assurdo. Anziché tentare una simile definizione, rimpiazzerò il quesito con un altro che è ad esso strettamente legato, e che è espresso in termini relativamente privi di ambiguità.[384]

È difficile immaginare un *incipit* più calzante per una "Rivista quadrimestrale di Psicologia e Filosofia": una questione filosofica a proposito di un legame possibile tra le macchine e l'intelligenza, viene modificata e trasformata per evitare che un tentativo di risposta ad essa, possa venir inficiato, o reso inattendibile, dalle definizioni d'uso comune dei termini che il problema chiama in causa, e quindi, in ultima analisi, dall'atteggiamento psicologico verso di essi.

La trasformazione del problema che Turing ha in mente, passa attraverso la descrizione di un gioco che si svolge tra tre individui: 'A', un uomo, 'B', una donna e 'C', che può essere invariabilmente dell'uno o dell'altro sesso, che ha il ruolo di "interrogatore". Si tratta di una versione riveduta e corretta del modello descritto nel 1948.

Al fine di garantire la correttezza della partita, i tre vengono posti in stanze separate che sono messe in collegamento mediante qualche apparecchio che, pur consentendo a 'C' la possibilità di porre domande e ad 'A' e 'B' di rispondere, impedisce che ciò avvenga attraverso un dialogo diretto, o qualsiasi altra forma di comunicazione, che permetta a 'C' di svelare i giocatori in un modo diverso dalla semplice deduzione sulla base delle risposte date. Lo scopo del gioco, infatti, consiste nell'indovinare, da parte di 'C', chi tra i due giocatori (X e Y, dal suo punto di vista) sia l'uomo (cioè 'A') e chi la donna ('B'). La difficoltà consiste essenzialmente nel fatto che il ruolo che Turing assegna ad 'A' è esclusivamente quello di trarre in inganno 'C', fingendo di essere la donna (da qui il nome del gioco, che Turing chiama "Gioco dell'Imitazione").

Già la semplice descrizione dell'esperimento, consente una prima riformulazione del problema di partenza:

> Poniamo adesso l'interrogativo, 'Cosa succederà se una macchina dovesse prendere la parte di A nel gioco?' Risponderà l'interrogatore in modo errato lo stesso numero di volte quando la partita venga giocata in questo modo

[384] [Tur50], p. 433.

> [cioè con una macchina al posto di A], rispetto al caso in cui la partita si svolga tra un uomo e una donna? Questi interrogativi sostituiscono quello di partenza 'Può una macchina pensare?'[385]

Il problema della liceità di una simile sostituzione, aleggia sullo sfondo dei paragrafi immediatamente successivi al primo, nei quali Turing è impegnato soprattutto a sottolineare i vantaggi del nuovo quesito mostrando la praticabilità dell'esperimento.

Esso, infatti, permette di tracciare "una linea netta tra le capacità fisiche e intellettuali di un uomo", così da isolare gli aspetti veramente essenziali del problema dell'intelligenza (seppure attraverso il criterio dell'imitazione), ed escludendo parimenti quelli inessenziali: nel selezionare le macchine per il gioco dell'imitazione, si potrà in particolare disinteressarsi della necessità di costruire una macchina che sia visivamente simile ad un essere umano.

Un secondo vantaggio, è invece rappresentato dal fatto che, potendo l'interrogatore spaziare su tutto lo scibile umano, si vengono così a riequilibrare i reciproci limiti dei giocatori: "[n]on vogliamo penalizzare la macchina per la sua incapacità di spiccare in una competizione di bellezza, né penalizzare un uomo per il fatto di uscire sconfitto da una gara contro un aeroplano"[386].

Ma la sostituzione del quesito di partenza con quello relativo al Gioco dell'Imitazione, consente a Turing, più di ogni altra cosa, di individuare con precisione il tipo di macchina che è maggiormente adatta a prendere parte all'esperimento ideato dal matematico britannico. Si tratta infatti di "calcolatori digitali", dei quali, nel passarne in rassegna le caratteristiche principali, Turing tiene a sottolineare soprattutto tre proprietà: (i) il fatto che "queste macchine sono concepite per eseguire ogni operazione che potrebbe essere portata a termine da un calcolatore umano"[387] (sono cioè equivalenti reali delle macchine di Turing astratte), e per tale motivo sembrano possedere un requisito essenziale per partecipare ad un gioco che consiste nell'imitare un essere vivente (per quanto non necessariamente 'calcolatore'[388]); (ii)

[385] *Ibid.*, p. 434.
[386] *Ibid.*, p. 435.
[387] *Ibid.*, p. 436.
[388] La frase che segue quella citata in precedenza, recita: "[s]i suppone che il calcolatore umano segua regole fissate; egli non ha alcuna autorità per sovvertirle in alcun caso". La proprietà indicata da Turing è dunque una conseguenza della Tesi da lui formulata fin dal 1937: in quanto tale, essa si riferisce alla completa capacità imitativa dei calcolatori, per porre la questione nei termini del lavoro che si sta prendendo in esame, rispetto ad un individuo che esegua una qualsiasi operazione di calcolo, ma non, ovviamente, qualsiasi

questo tipo di macchina è completamente deterministico nel senso che "una conoscenza ragionevolmente accurata dello stato [del calcolatore] in un certo istante, consente una conoscenza ragionevolmente accurata [di esso], qualsiasi numero di passi seguenti"; (iii) la predicibilità pressoché assoluta del comportamento di un calcolatore digitale, è all'origine dell'ulteriore vantaggio rappresentato dalla loro universalità (la proprietà di poter 'imitare' ogni altra macchina discreta), la quale consente di guardare alla decisione di confinare l'attenzione a macchina di questo tipo, come ad una restrizione irrilevante.

Sulla base delle precisazioni suddette, Turing può così formulare il problema destinato a sostituire il quesito 'Può una macchina pensare?' in termini assolutamente "privi di ambiguità" (laddove invece la precedente trasformazione conteneva ancora il riferimento generico ad una 'macchina', non ben specificata, che prendesse parte al Gioco dell'Imitazione):

> Si fissi l'attenzione su un particolare calcolatore digitale C. È vero che modificando questo calcolatore in modo che esso sia dotato di una memoria sufficiente, migliorandone opportunamente la velocità di esecuzione, e fornendolo di un programma appropriato, C può essere messo in grado di giocare in modo soddisfacente la parte di A nel gioco dell'imitazione, qualora la parte di B fosse presa da un uomo?[389]

Si noti che, tenendo presente come il compito dei due giocatori che prendono le parti di A e B non sia mutato (resta quindi quello, per A, di far finta di essere una donna, per B di aiutare l'interrogatore a fare la scelta giusta), in quest'ultima versione il gioco può essere detto 'dell'imitazione' in un duplice senso: non solo perché il giocatore che prende il posto di A (cioè la macchina) deve far finta di essere una donna, ma perché essendo il posto di B preso da un uomo, cioè un individuo di sesso maschile, anch'egli non può che fare la stessa cosa. In questo senso letterale, all'interrogatore non rimane alcuna risposta corretta: o egli indica la macchina, credendo che sia la donna, oppure, con lo stesso convincimento, sceglie l'uomo.

Tuttavia, in questo stesso senso, il gioco equivale, per le finalità preposte da Turing, al precedente: esso consente di ravvisare le capacità imitative della macchina, desumendole sulla base delle risposte dell'interrogatore, e di confrontarle direttamente con quelle di un uomo, essendo questi il suo avversario (laddove la prima trasformazione sembrava alludere alla necessità di portare a termine due serie di partite, una con un uomo al posto di A,

operazione in generale.

[389] *Ibid.*, p. 442.

l'altra con la macchina, e confrontare, nei due casi, il numero di risposte scorrette dell'interrogatore).

Appare lecita, comunque, un'interpretazione che segua meno alla lettera i passaggi operati da Turing (una lettura che sembra essere più in linea con l'odierna interpretazione del Test di Turing), e che consenta tuttavia di far emergere meglio il senso della trasformazione proposta del problema originario: il gioco si svolge, alle condizioni 'ambientali' indicate da Turing, tra un essere umano e una macchina; ad essi viene dato un compito che si suppone necessiti una qualche forma di intelligenza; l'interrogatore dà la risposta esatta solo nel caso in cui individui correttamente, sulla base delle risposte ai quesiti da lui sollevati, l'uomo e la macchina.

3.3.2 Macchine che imparano

È opportuno considerare il contributo di Turing del 1950 alla luce di un commento di Robin Gandy, una figura che fu, oltre che un allievo del matematico britannico, legato al Nostro da un rapporto di amicizia:

> L'articolo del 1950 non si proponeva tanto di dare un contributo profondo alla filosofia, quanto di fare della propaganda. Turing riteneva che i tempi fossero maturi perché filosofi, matematici e scienziati prendessero sul serio l'idea che i calcolatori non sono solo macchine per calcolare, ma entità capaci di esibire comportamenti che dobbiamo considerare intelligenti. Cercò di diffondere questa convinzione e scrisse questo articolo - a differenza dei suoi articoli matematici - rapidamente e prendendoci gusto. Lo ricordo mentre mi leggeva dei brani - sempre con un sorriso sulle labbra, a volte frenando una risata. Alcuni commenti che ho avuto modo di leggere sovraccaricano questo articolo di significati rispetto alle intenzioni iniziali.[390]

L'invito di Gandy a non dare un peso soverchio a certe riflessioni da parte di Turing, non può però essere una scusa per negare le evidenti similitudini tra le considerazioni del contributo del 1950 e quelle svolte nei lavori che si sono presi in esame in precedenza.

Né può essere un alibi per sottacere la posizione di Turing a proposito del quesito posto sulle capacità dei calcolatori digitali, opportunamente modificati, di prendere parte in modo soddisfacente al Gioco dell'Imitazione (una posizione, peraltro, decisamente moderata - sintomo certo di un'inossidabile integrità scientifica - nonostante gli scopi propagandistici del lavoro dei quali parla Gandy). Il passo succitato, anzi, sembra fornire un sostegno all'idea di

[390] [Gan96], p. 125.

misurare ciò che esattamente Turing sottoscrisse in quell'occasione, piuttosto che abbandonarsi a speculazioni che ne tradiscano lo spirito.

Tanto più, che se la 'questione dell'intelligenza' viene riformulata da Turing in termini più definiti, altrettanto precisa e priva di ambiguità appare essere la sua congettura riguardo agli esiti dell'esperimento da lui proposto:

> Credo che in un lasso di tempo di cinquant'anni, sarà possibile programmare calcolatori con una capacità di memoria di circa 10^9, che possano prendere parte al gioco dell'imitazione in modo così capace da non lasciare ad un interrogatore esperto più del 70 per cento di possibilità di fare l'identificazione corretta dopo cinque minuti di domande. La questione originaria, 'Può una macchina pensare?', la considero troppo priva di significato per essere meritoria di una discussione. Tuttavia, credo che alla fine del secolo l'uso terminologico e l'opinione comunemente diffusa si sarà mutata al punto tale che si potrà parlare di macchine pensanti senza aspettarsi di venire contraddetti.[391]

Lasciamo volentieri ad altri il compito di giudicare quanto le aspettative di Turing trovino un riscontro effettivo negli sviluppi tecnologici contemporanei.

Ci limitiamo a sottolineare come egli sia estremamente prudente nel descrivere le capacità dei calcolatori digitali alla prova del Gioco dell'Imitazione (la percentuale sarebbe ad avviso di Turing ancora largamente a sfavore delle macchine, e c'è da aspettarsi che lo sarebbe anche in misura maggiore con un interrogatorio più lungo), e come questa prudenza si accompagni, come risulta in modo chiaro dal prosieguo del passo, ad una consapevolezza del fatto che certe ipotesi, talvolta anche le meno ortodosse, possono rivelarsi inaspettatamente fruttuose per la ricerca scientifica:

> Credo inoltre che non si serve alcuna causa utile nel nascondere simili convinzioni. L'immagine popolare che gli scienziati procedano inesorabilmente da fatti ben stabiliti a fatti ben stabiliti, senza essere influenzati da nessuna congettura non dimostrata, sia abbastanza sbagliata. Data una chiarificazione di quali siano i fatti e quali le congetture, nessun danno può risultare. Le congetture sono di grande importanza dal momento che suggeriscono utili linee di ricerca.[392]

Ciò contribuisce, crediamo, a 'neutralizzare' ulteriormente le riflessioni di Turing (rispetto ad un intento platealmente propagandistico), e a renderle, per tale motivo, sufficientemente 'affidabili'.

---

[391] [Tur50], p. 442.

[392] *Ibid.*, p. 442.

Si diceva delle similitudini con i testi analizzati precedentemente. Queste, emergono soprattutto nella parte successiva del lavoro, nella quale Turing offre tanto un'analisi di alcune delle posizioni avverse alla propria, quanto alcune argomentazioni a sostegno dei convincimenti da lui espressi. Trattandosi di luoghi nei quali gli elementi 'folkloristici' del lavoro di Turing dei quali parla Gandy emergono in misura maggiore, ci limitiamo a scegliere quei passi che ci paiono essere i più rilevanti.

Primo fra tutti, quello nel quale Turing affronta, ancora una volta, l' "obiezione matematica", e cioè l'idea che certi teoremi logici indichino l'impossibilità, per le macchine, di esibire un comportamento intelligente.

Le novità sono due: una consiste nel fatto che Turing decida di proporre una versione dell'argomento che fa uso del proprio risultato relativo all'esistenza di problemi indecidibili mediante una macchina di Turing, motivando la scelta sulla base del fatto che questo risultato "è il più conveniente da considerare, dal momento che si riferisce direttamente a delle macchine, laddove gli altri possono essere usati solo in argomenti indiretti al confronto di questo"[393]; l'altro fatto nuovo, emerge invece dalle risposte all'obiezione che Turing formula, e che sembrano risentire delle condizioni (quelle del Gioco dell'Imitazione) alle quali il confronto tra uomo e macchina si svolge.

Il passo cruciale recita:

> Il risultato in questione si riferisce ad una macchina che è in sostanza un computer digitale con una memoria infinita. Stabilisce che ci sono certe cose che una macchina di questo tipo non può fare. Se questa fosse messa a punto per dare risposte a determinate domande come nel gioco dell'imitazione, ci sarebbero certi quesiti rispetto ai quali o fornirebbe la risposta sbagliata, o non fornirebbe alcuna risposta per quanto tempo gli venisse concesso per darla. [...] Le domande su cui sappiamo che la macchina deve commettere degli errori sono del seguente tipo, "Considera la macchina specificata nel modo seguente .... Risponderà mai questa macchina 'Si' ad una certa domanda?". I puntini devono essere rimpiazzati dalla descrizione standard di una qualche macchina [...]. Se la macchina descritta sta in una relazione relativamente semplice con la macchina a cui è rivolta la domanda, è possibile dimostrare o che la risposta è sbagliata o che essa non arriverà. Questo è il risultato matematico: si suppone che esso dimostri un'incapacità delle macchine alla quale non è soggetta la mente umana.[394]

Si noti, innanzi tutto, che è possibile dare un'interpretazione precisa delle parole di Turing.

[393] *Ibid.*, p. 444.
[394] *Ibid.*, pp. 444-445.

Indichiamo infatti con **M** la macchina (universale) che partecipa al Gioco e che è, quindi, sottoposta alle domande dell'interrogatore. Essa non è dunque altro che (o contiene in se) un dispositivo $\mathsf{M}(x, y)$ tale che $\mathsf{M}(m, n) = \mathtt{SI}$ ogni qual volta l'm-esima macchina (in un qualche ordinamento) risponde SI all'n-esimo quesito (in una qualche enumerazione o codifica), ossia (per ogni $m, n \in \mathbf{N}$):

$$\mathsf{M}_m(n) = \mathtt{SI} \Rightarrow \mathsf{M}(m, n) = \mathtt{SI}$$

La macchina che "sta in una relazione semplice" con **M**, è allora banalmente una macchina 'diagonale', $\mathsf{D}_{\mathbf{M}}(x)$ che, per ogni $n \in \mathbf{N}$, è definita da:

$$\mathsf{D}_{\mathbf{M}}(n) = \begin{cases} \mathtt{NO}, & \text{se } \mathsf{M}(n, n) = \mathtt{SI} \\ \mathtt{SI}, & \text{altrimenti} \end{cases}$$

Supponiamo che esista una descrizione d (in una codifica standard) per $\mathsf{D}_{\mathbf{M}}(x)$. La domanda critica è allora quella rappresentata dalla coppia (d, d), dal momento che, per definizione di $\mathsf{D}_{\mathbf{M}}$:

$$\mathsf{M}(d, d) = \mathtt{SI} \Rightarrow \mathsf{D}_{\mathbf{M}}(d) = \mathtt{NO}$$

mentre, per definizione di $\mathsf{D}_{\mathbf{M}}$ e M:

$$\mathsf{M}(d, d) = \mathtt{NO} \Rightarrow \mathsf{D}_{\mathbf{M}}(d) = \mathtt{SI} \Rightarrow \mathsf{M}(d, d) = \mathtt{SI}$$

Dal ché si deve concludere o che, nell'un caso, M può fornire risposte sbagliate, oppure, volendo mantenere l'ipotesi di correttezza, che può anche non offrire alcuna risposta per certi input.

Interpretazione del passo a parte, si è già sottolineato come la reazione di Turing all'obiezione avanzata vada letta, per essere meglio compresa, alla luce della situazione che si viene a determinare con il Gioco dell'Imitazione.

La risposta di Turing si articola infatti in tre parti. In primo luogo, egli sottolinea come, anche supponendo che la suddetta obiezione mostri una limitazione delle macchine (cioè dei calcolatori digitali), "si è soltanto affermato, senza alcuna sorta di dimostrazione, che nessuna limitazione dello stesso stesso tipo si applichi all'intelletto umano". A ciò, Turing aggiunge il fatto che alla sensazione di superiorità proveniente dal sapere come indurre in errore la macchina, non deve essere assegnata troppa importanza: "[n]oi stessi diamo troppo spesso risposte sbagliate alle domande, per essere giustificati nel sentirsi rincuorati di fronte alla manifesta fallibilità delle macchine". Egli conclude infine, appellandosi all'infinita perfettibilità, rispetto al quesito critico, delle macchine, osservando che "[...] la nostra superiorità può essere giustificata in una simile occasione in relazione alla macchina rispetto alla quale si è messo a segno il nostro insignificante trionfo. Non ci sarebbe motivo di gridare alla sconfitta di *tutte* le macchine simultaneamente. In breve, ci può essere un uomo più in gamba di una certa macchina, ma ci potrebbero essere macchine ancora più in gamba, e così via."[395].

Stando a come l'interrogazione dei due giocatori avviene nel corso del Gioco dell'Imitazione, sembra voler dire Turing, l'incapacità di una macchina a fornire una risposta (o a fornire una risposta invariabilmente corretta) in relazione a certi quesiti, potrebbe non svolgere un ruolo decisivo: perché bisogna intanto essere certi che il contendente umano sia in grado di far meglio (e non c'è alcuna dimostrazione che ciò valga in ogni caso); perché anche nel caso di una risposta scorretta da parte della macchina, questa potrebbe essere equilibrata, agli occhi dell'interrogatore, dagli errori commessi dall'altro giocatore; perché, infine, la 'vittoria', al gioco, con la macchina in questione, non comporta che non si possa costruire una macchina diversa, e più potente, con la quale 'sconfiggere' il vincitore della 'partita' precedente.

Se si escludono le novità di cui si è detto, la disamina dell'"obiezione matematica" non si discosta tuttavia in modo decisivo dai commenti relativi ad essa contenuti negli altri lavori che si sono in presi in esame. Qualche indicazione in più rispetto a questi ultimi, la si ricava dalle risposte ad altre possibili obiezioni che Turing considera.

Un caso, è quello dell'"argomento della coscienza", che sembra toccare la questione delle eventuali implicazioni filosofiche insite nella trasformazione del problema iniziale. Si sarà notato che, infatti, se vista in relazione al quesito originario sulle macchine e l'intelligenza, la scelta di Turing di considerare come equivalente il problema legato al Gioco dell'Imitazione, potrebbe far nascere il sospetto di uno sfondo di tipo 'comportamentista' dietro di essa: la sostituzione è legittima cioè, proprio perché 'essere intelligenti' non significa altro che 'esibire un comportamento intelligente'.

Dietro l'affermazione che "[n]essun meccanismo potrebbe provare (e non meramente segnalare artificialmente, un facile espediente) piacere per i propri successi, dolore quando le proprie valvole fondono, essere gratificato dall'adulazione, essere infelice per i propri errori, essere affascinato dal sesso, arrabbiarsi o essere depresso quando non riesce ad ottenere ciò che vuole", c'è dunque una vera e propria confutazione della validità, nel senso di fedeltà

[395] *Ibid.*, p. 445.

al quesito d'origine, del Gioco di Turing: "l'unico modo per mezzo del quale si potrebbe essere essere sicuri che una macchina pensa, sarebbe *essere* la macchina e sentire se stessi pensare"[396].

In risposta a quest'obiezione, Turing nota, innanzi tutto, come una simile posizione possa facilmente essere ricondotta ad una forma di solipsismo, probabilmente inaccettabile anche per i sostenitori di una simile tesi, dalla quale discende che, non solo con le macchine, ma persino di fronte ad un uomo, si dovrebbe essere portati ad accettare il fatto che anch'egli pensa solo per averlo provato in modo diretto. Si sarebbe così condotti ad assumere, insieme al fatto che le macchine non pensano, che, in questo, esse non sono affatto diverse da ogni altro essere vivente.

Ma al di là di questo facile esercizio retorico di critica filosofica, rimane ciò che Turing dice per concludere la disamina di quest'obiezione:

> Non vorrei dare l'impressione di pensare che non esiste alcun mistero riguardo alla coscienza. C'è, ad esempio, qualcosa di paradossale legato ad ogni tentativo di localizzarla. Ma non penso che questi misteri debbano necessariamente essere risolti prima che si possa rispondere alla questione che è all'attenzione di questo articolo.[397]

L'affermazione di Turing, ci permette dunque di sgombrare il campo dal sospetto che, dietro l'abbandono del quesito originario e la proposta di sostuirlo con l'indagine sulle possibilità delle macchine di partecipare proficuamente al Gioco dell'Imitazione, vi sia alcun impegno, non solo di tipo comportamentista, sulla natura della coscienza: l'idea di Turing riflette piuttosto la volontà di confinarsi ad un obiettivo praticabile (e soprattutto scientificamente misurabile), lasciando fuori tanto pregiudiziali di partenza qualsivoglia, quanto assunzioni concettuali che, a suo dire, non risulterebbero comunque decisive.

Lo stesso spirito, permea la risposta di Turing ad un'obiezione ulteriore, secondo la quale l'idea di una macchina intelligente sarebbe implausibile per il fatto che non è possibile dare un insieme finito di regole che ci permettano di descrivere le reazioni ed il comportamento che un essere umano dovrebbe avere in ogni circostanza concepibile.

La stessa osservazione, nota Turing, permette di formulare un argomento a sostegno dell'idea che l'uomo non è una macchina, in modo diretto e nella

[396] *Ibid.*, p. 446.
[397] *Ibid.*, p. 447.

forma di un'inferenza stringente[398]: "Se ogni uomo avesse un insieme di regole di condotta ben definito mediante le quali egli organizza la propria vita, non sarebbe migliore di una macchina. Ma non esistono tali norme, quindi gli esseri umani non possono essere macchine"[399].

La risposta che il Nostro elabora, si basa su una distinzione tra "regole di condotta", ossia quei precetti che sono alla base delle nostre decisioni (l'esempio di Turing è quello di una regola del tipo "Fermati quando vedi la luce rossa"), e "leggi di comportamento", cioè "le leggi di natura così come sono applicate al corpo umano". È opinione di Turing che l'inferenza di poc'anzi, possa venir vista come "non più insuperabile" se a 'regole di condotta' vi viene sostituita l'espressione 'leggi di comportamento':

> Non possiamo tuttavia convincerci altrettanto facilmente dell'assenza di leggi di comportamento così come è possibile fare con le regole di condotta. L'unico modo che conosciamo per scoprire tali leggi è l'osservazione scientifica, e non conosciamo certamente circostanze alle quali potremmo dire, 'Abbiamo cercato abbastanza. Non esistono tali leggi'.[400]

La prospettiva che si apre, è dunque quella di una ricerca scientifica potenzialmente infinita, ma che, proprio per tale motivo, non consente alcuna speculazione sui suoi esiti. Su questa base, si può quindi comprendere meglio, tanto la prudenza di Turing nel delineare le proprie previsioni (anche in relazione al modesto proposito relativo al Gioco dell'Imitazione) e le proprie convinzioni profonde, quanto la decisione nel sostenere la praticabilità del progetto, così ambizioso nelle sue conseguenze.

Ma, soprattutto, se di ricerca scientifica si tratta, quello che si può certamente fare con minori remore, è chiarirne le direttrici fondamentali. È a questo che Turing dedica il paragrafo conclusivo dell'articolo, sulla base di spunti che ci sono già noti.

Torniamo per un attimo alla previsione del matematico britannico: per la fine del novecento sarà possibile programmare un calcolatore digitale che possa partecipare al Gioco dell'Imitazione in un modo ragionevolmente soddisfacente. Quello che di rilevante c'è in questa prospettiva è il fatto che essa si riferisca a dispositivi che sono la controparte pratica di macchine di Turing: ciò vuol forse dire che, in contrasto con quanto siamo venuti sottolineando

[398] Si tratta davvero di un'inferenza stringente? L'analisi logica di essa fa piuttosto pensare ad una fallacia per falsa contrapposizione (del tipo $A \rightarrow B$, $\neg A$ dunque $\neg B$).
[399] *Ibid.*, p. 452.
[400] *Ibid.*, p. 452.

commentando gli altri scritti di Turing, è a questo modello, così come esso è, che si deve guardare?

Le osservazioni conclusive di [Tur50] fanno riemergere i due aspetti che si erano messi in evidenza precedentemente, ossia il ruolo essenziale di una qualche forma di addestramento che richiami il processo educativo da un lato, e la necessità di abbandonare il paradigma delle macchine di Turing *tout court* dall'altro.

"Anziché cercare di produrre un programma che simuli la mente adulta", si legge, "perché piuttosto non tentare di produrne uno che simuli quella del bambino?"[401].

La convinzione di Turing è che ciò possa rappresentare il punto di partenza migliore per l'indagine che si vuole intraprendere:

> Se [il programma-bambino] fosse soggetto ad un processo educativo opportuno si otterrebbe il cervello di un adulto. Presumibilmente il cervello del bambino è qualcosa di simile ad un taccuino così come lo si compra in cartoleria. Pochi meccanismi e molte pagine bianche. [...] La nostra speranza è che vi siano così pochi meccanismi nel cervello del bambino che qualcosa di simile possa essere programmato con facilità. La quantità di lavoro da svolgere nell'educazione si può assumere, in prima approssimazione, essere la stessa necessaria per il bambino umano.[402]

Con questa premessa, che, come già in [Tur48], suddivide il compito in due parti ben precise, la costruzione di una 'macchina infantile' e la messa a punto di un addestramento che ricalchi da vicino quello a cui è soggetto l'essere umano nel corso della propria vita, l'abbandono del paradigma delle macchine di Turing avviene in almeno due sensi.

In primo luogo, perché anche qualora risultasse opportuno associare al programma in preparazione qualcosa di analogo ad un vero e proprio sistema formale, l'accento dovrà essere posto soprattutto sul valore euristico dei passi inferenziali:

> Perché ad ogni stadio nell'uso di un sistema logico, c'è un grande numero di passi alternativi, ognuno dei quali è lecito fin tanto che si presta attenzione all'obbedienza alle regole del sistema. Queste scelte fanno la differenza tra un ragionatore brillante e uno pedante, non tra uno corretto e un'altro fallace.[403]

La costruzione del sistema formale, insomma, dovrà essere fatta in modo che esso, per ciò che riguarda l'ingegnosità, la capacità individuata in

[401] *Ibid.*, p. 456.
[402] *Ibid.*, p. 456.
[403] *Ibid.*, p. 458.

[Tur39] di costruire dimostrazioni matematiche ineccepibili, non necessariamente "soddisfi il più rigoroso tra i logici"[404].

In un secondo senso, il tipo di macchina al quale si deve pensare affinché il processo educativo possa risultare soddisfacente, non potrà più consentire che l'"insegnante' incaricato di istruirla abbia una completa conoscenza del suo stato in ogni momento, ed eserciti dunque la facoltà di predirne in modo assoluto il comportamento.

"[Q]uesto è chiaramente in contrasto", sottolinea Turing, "con la procedura normale nell'uso delle macchine per eseguire calcoli"[405].

Ma è pur sempre in accordo con un'immagine plausibile della ragione umana:

> Il comportamento intelligente consiste presumibilmente in un abbandono del comportamento perfettamente disciplinato implicato nell'esecuzione di un computo, ma di tipo piuttosto simile, che non dà vita ad un comportamento di tipo casuale o a *loop* ripetitivi e inutili.[406]

3.4 Menti gödeliane vs Macchine di Turing?

La terza nota sui risultati di indecidibilità di Gödel del 1972[407] che si è esaminato nel §2.4.1 della prima parte, offre, alla luce della disamina di certi lavori di Turing, la possibilità di inserire in modo organico la valutazione della posizione del matematico britannico nel quadro del presente lavoro.

Come forse si ricorderà, Gödel riferisce in quella circostanza di un "errore filosofico" da parte di Turing (contenuto nel §9 di [Tur36]) consistente nell'idea che la mente umana non possa andare oltre le procedure meccaniche di calcolo. Ad essa, Gödel contrappone la propria concezione, secondo la quale la mente umana è "costantemente in evoluzione" e, per tale motivo, essenzialmente non costruttiva.

Cosa possiamo dire, arrivati a questo punto, su tale contrapposizione?

Il confronto tra Gödel e Turing, diviene particolarmente intrigante nel momento in cui si realizza, sulla base dei rispettivi contributi, come essi offrano l'occasione di guardare alla problematica del rapporto tra menti e macchine da due punti di vista speculari, e quindi complementari. La critica di Gödel,

[404] *Ibid.*, p. 458.
[405] *Ibid.*, p. 459.
[406] *Ibid.*, p. 459.
[407] Cfr. [Göd90c].

prende le mosse da una concezione della mente come costitutivamente distinta da una macchina (di Turing), laddove il matematico britannico, ponendosi nella prospettiva di voler dare vita ad un dispositivo meccanico che sia in grado di esibire un comportamento intelligente, assume inevitabilmente, nella diatriba, la parte delle macchine.

Tuttavia, proprio in quanto riteniamo che le posizioni di entrambi siano di assoluto rilievo all'interno di un dibattito che si intende riprendere e approfondire nei prossimi capitoli, non vogliamo limitarci a simili considerazioni di carattere generale (che non sono comunque, come si vedrà al termine della disamina, prive di un certo valore), quanto piuttosto affrontare la questione in modo diretto, anche a costo di apparire fin troppo schematici.

Per chiarire se, ed in quali termini, con Gödel e Turing si sia di fronte ad una contrapposizione tra due punti di vista diametralmente opposti per ciò che concerne la natura della mente umana, occorre, ci pare, rispondere ad una serie di quesiti.

Il primo, come può forse apparire ovvio, è: Gödel era un anti-meccanicista?

Valgono in questo senso le sue stesse dichiarazioni esplicite al proposito. Egli era un sostenitore dell'idea che l'attività della mente non potesse essere ricondotta in modo esclusivo alla sola azione del cervello e delle sue parti, che per Gödel è, a tutti gli effetti, una macchina. Si è visto come questa convinzione fosse per Gödel uno dei pilastri sul quale basare la propria visione della matematica, tanto da un punto di vista ontologico quanto da quello epistemologico.

Se il primo passo si presenta come privo di complicazioni, più arduo appare invece trovare risposta alla domanda corrispondente relativa a Turing: egli era forse un meccanicista?

La tentazione sarebbe di rispondere che sì, probabilmente lo era, per quanto nei suoi lavori non sembra potersi riscontrare un'affermazione certa ed indubitabile al proposito (sicuramente non altrettanto lampante come le considerazioni di tipo opposto da parte di Gödel). Tuttavia, i limiti di una simile conclusione divengono palesi non appena si cerchi di approfondire minimamente la questione.

Dicendo che Turing era un meccanicista, intendiamo con ciò che egli fosse un sostenitore dell'idea che la mente fosse in un qualche senso equivalente ad una macchina di Turing?

C'è più di un motivo per rispondere negativamente ad un simile quesito. Tale risposta, sarebbe infatti in linea con quanto emerge unanimemente dai suoi lavori: sia da quello dedicato alle logiche ordinali, dove tra le conseguenze dei teoremi di Gödel, egli annovera la natura essenzialmente "non costruttiva" della conoscenza matematica (nel senso dell'impossibilità di ridurre ad un numero finito di regole l'apporto della facoltà intuitiva); sia dagli scritti più tardi sul progetto di una macchina pensante, nei quali Turing parla esplicitamente della necessità di abbandonare il modello delle macchine astratte da lui definite, al fine di dare vita ad un dispositivo meccanico che esibisca un comportamento intelligente.

Sia nei contributi logici che in quelli successivi, si riserva un ruolo essenziale ai teoremi limitativi della logica ed ai risultati di Gödel in particolare. Non si profila dunque una contraddizione, ci si potrebbe chiedere, dietro il ricorso ai teoremi di incompletezza per sostenere l'ineliminabilità dell'apporto (non costruttivo) dell'intuizione, da un lato, e nel non vedere nelle implicazioni di questi stessi risultati, riformulati come nell' "obiezione matematica", un ostacolo alla realizzazione del progetto di una macchina intelligente dall'altro?

Si avrebbe gioco facile ad appellarsi proprio all'analisi di Gödel delle implicazioni dei risultati di indecidibilità contenuta nella *Gibbs Lecture*, per mostrare come non vi sia, di principio, alcuna contraddizione logica nei due punti di vista: anche ritenendo che nessun sistema formale possa contenere tutti i giudizi matematici di natura intuitiva, si potrebbe di contro argomentare in favore, per riprendere la distinzione di Turing, della natura puramente algoritmica dell'ingegnosità, ossia della nostra capacità di costruire dimostrazioni ineccepibili a sostegno di quanto intuito; da ciò e dai teoremi di Gödel, discenderebbe meramente l'esistenza di giudizi intuitivi per i quali non si è in grado di offrire alcuna prova.

Tuttavia, non v'è dubbio che ci sia una differenza di contesto importante, che pure si è notato nel corso del capitolo, tra i due riferimenti ai teoremi di incompletezza, essendo il primo, quello contenuto in [Tur39], di natura fondazionale (l'ineliminabilità dell'intuizione come impraticabilità del programma di Hilbert), ed il secondo, più legato ad un ambito di ricerca che sta a metà tra la matematica, l'informatica e la robotica. Ciò lo si intuisce anche dalla risposta che Turing offre all' "obiezione matematica" negli scritti più tardi: l'idea delle macchine che sbagliano sembra pensata per rendere il progetto di una macchina intelligente euristicamente plausibile, piuttosto che fondazionalmente ineccepibile.

Proprio a tal proposito occorre fare una precisazione importante. Nel corso della nostra disamina degli scritti di Turing, si è più volte voluto sottolineare l'importanza di quei passi nei quali ci è parso di riconoscere un esplicito riferimento, da parte del matematico britannico, alla necessità di superare il paradigma delle macchine di Turing nel tentativo di dare vita ad un dispositivo meccanico intelligente, cioè in grado di 'competere' (come nel caso del Gioco dell'Imitazione) con un essere umano.

È chiaro che si potrebbe vedere in questo fatto in particolare, un elemento del tutto ininfluente rispetto al problema sul quale ci stiamo al momento interrogando, ossia il fatto se vi sia o meno, dietro il progetto di Turing, uno sfondo concettuale di tipo materialistico o meccanicistico. Anzi, proprio le osservazioni relative alle carenze del modello teorico, potrebbero indurre a pensare che la 'mossa' di Turing, l'adeguamento euristico del modello delle procedure effettive di computo, sia non solo, come si è detto, ininfluente, ma persino obbligata, rispetto ad un'interpretazione computazionale della mente.

La questione, tuttavia, è ugualmente meritevole di attenzione.

Intanto, perché quell'aspetto delle ricerche proposte da Turing è estremamente rilevante da un punto di vista storico: proprio i suggerimenti relativi alle varie linee di ricerca che possono nascere traendo spunto da una riflessione sullo sviluppo e sull'addestramento di un essere intelligente, fanno del matematico britannico un padre fondatore riconosciuto di quel settore di indagine, al quale contribuiscono le più disparate discipline, che è noto come Intelligenza Artificiale[408].

Sotto un altro aspetto, invece, le considerazioni di Turing relative ai limiti evidenti del modello rappresentato dalle funzioni calcolabili, sono importanti proprio in quanto possono costituire, come si cercherà di vedere meglio nel prosieguo, una risposta di per sé già esauriente a quel tipo di obiezioni alla tesi meccanicistica fondata su certi teoremi metamatematici. Argomenti come l'"obiezione matematica" sono stati, infatti, variamente riproposti ed elaborati (segno, certo, di una chiara lungimiranza da parte di Turing ad individuarne, per primo, la rilevanza), per contrastare, già a livello teorico, proprio il fondamento delle ricerche nel campo dell'Intelligenza Artificiale. Risulta che, al di là di fraintendimenti formali e ambiguità concettuali (aspetti ai quali è dedicato il prossimo capitolo del presente lavoro), proprio

[408]Non abbiamo qui velleità (né saremmo peraltro capaci di dare seguito ad esse, se ne avessimo) di ricostruire nel dettaglio lo stretto legame che è possibile riconoscere tra i progetti, spesso allo stato embrionale, ai quali Turing allude, e quelli realizzati in tempi successivi. Un'analisi di questo genere, che cerca di mettere in relazione il contenuto degli scritti tardi di Turing con le ricerche di H. Simon e A. Newell, è quella contenuta, ad esempio, in [Col97].

abbandonando (in un senso da doversi precisare nel capitolo 5) il paradigma logico al quale i risultati chiamati in causa si applicano, è possibile contestare la portata di simili argomentazioni. Ciò conduce, oltretutto, alla possibilità di arricchire in modo non banale la discussione sul rapporto menti/macchine.

Dunque, ci pare doveroso riconoscere che, *se* nel caso di Turing fossimo di fronte ad un sostenitore di una concezione computazionale della mente (un periodo ipotetico che, a nostro parere, è d'obbligo), non avremmo a che fare, tuttavia, né con una posizione meritevole di accantonamento, né, in ogni caso, con un punto di vista facilmente accantonabile.

Rimane comunque, per tornare alla questione della presunta contrapposizione con Gödel, la difficoltà di inquadrare con precisione la posizione di Turing, al di là dell'impressione, che si può sviluppare dalla lettura dei suoi scritti, sulla presenza o meno di un simile sfondo concettuale. E una simile difficoltà irrisolta, rende impraticabile, a nostro avviso, un confronto di merito con Gödel. All'origine della diversità di vedute dei due, sembra esserci soprattutto una differente prospettiva dalla quale essi affrontano problemi indubitabilmente legati fra loro, e che ben emerge dall'esame dei contributi più tardi di Turing se confrontati con la prospettiva gödeliana di [Göd51]: dopo l'indagine sulle logiche ordinali, quando il matematico britannico parla di macchine e di intelligenza, egli non intende riferirsi, con questi termini, ad un modello matematico astratto e a qualcosa che si predica, con consapevolezza assoluta, della 'ragione umana'.

Non ci pare neanche implausibile ritenere, a conferma di ciò, che, a certe condizioni, Turing potesse persino condividere le conclusioni di Gödel della *Gibbs Lecture*, e finire per sostenere anch'egli l'idea che la mente umana superi infinitamente le capacità matematiche di una macchina astratta. Ciò sarebbe vero, ad esempio, tenendo presente quanto Turing dice (più volte, nei suoi scritti tardi) a proposito del fatto che "[n]essun uomo isolato contribuisce poi molto al corpo delle conoscenze"[409], e supponendo, ad un tempo, di intendere con 'mente umana' il prodotto trascendentale di un'astrazione, che riassuma gli sforzi dell'intera comunità dei matematici.

Turing sembra però partire dal presupposto che nessuna limitazione teorica possa avere invece alcuna rilevanza, quando si parla di una *certa* macchina e di un *certo* individuo, e del problema di giudicare in modo obbiettivo le differenze tra i due (da questo punto di vista, consideriamo l'idea alla base del Gioco dell'Imitazione come estremamente significativa e affatto casuale).

[409] [Tur86], p. 105.

A queste condizioni, infatti, non esiste teorema matematico che possa assicurare che, in ogni situazione possibile, l'individuo che si considera di volta in volta sia sempre in grado di superare il calcolatore che si è messo a punto. Soprattutto, a queste condizioni, Turing sosterrebbe che si pone il problema dell'intelligenza in un modo che non è meno rilevante della situazione teorica e astratta che i teoremi di incompletezza, ad esempio, consentono di esaminare.

È in questo secondo senso, che egli aveva elaborato un progetto scientifico di ricerca preciso, ma con labili connessioni di partenza con uno sfondo filosofico sulla natura della mente umana.

Valga per tutte, l'analogia della "buccia di cipolla" alla quale egli ricorre per spiegare in che senso la mente potrebbe rivelare, nella realizzazione del progetto, la propria natura computazionale:

> Nel considerare le funzioni della mente o del cervello troviamo certe operazioni che siamo in grado di spiegare in termini puramente meccanici. Diciamo che ciò non corrisponde alla vera mente: è una sorta di buccia che dobbiamo tirare via se vogliamo scoprire la vera mente. Ma poi in ciò che resta scopriamo un'altra buccia da tirar via, e così via. Procedendo in questo modo arriviamo alla 'vera' mente, oppure giungiamo ad una buccia che non contiene nulla? In quest'ultimo caso la mente è meccanica nel suo complesso.[410]

Non c'è quindi alcun bisogno di postulare alcunché al riguardo: la vera natura della mente umana, sarà eventualmente una conseguenza dei risultati acquisiti.

Tornando al confronto con Gödel dal quale siamo partiti, si possono allora ravvisare le condizioni per una distizione, con quest'ultimo, sul piano squisitamente scientifico, laddove Gödel, lo si è visto nella prima parte, arrivava ad immaginare la possibilità di confutare scientificamente il meccanicismo, come riducibilità della mente al solo cervello, mentre Turing era disposto quantomeno a lasciare aperta l'alternativa.

Ma non c'é vera e propria disputa filosofica, perché manca un impegno preciso, in questo senso, da parte del matematico britannico.

Il contrasto con Turing di cui parla Gödel, dunque, c'è, se ci si mantiene ad un livello generale, perché ad un certo disimpegno dell'uno su certe tematiche, fanno da contraltare le prese di posizione esplicite dell'altro (a patto di tenere presente il materiale del *Nachlass* gödeliano dove esse compaiono con maggior

[410] [Tur50], p. 455.

regolarità). Ma questo conflitto non si riesce a renderlo più profondo, di merito cioè, oltre un certo limite: Turing, in buona sostanza, non era affatto un filosofo.

E questa, tenendo presente quanto si è concluso nella prima parte del presente lavoro, non è poi una differenza di poco conto rispetto a Gödel.

Capitolo 4

L' "obiezione matematica" all'opera

Lo scopo del presente capitolo è quello di andare a vedere più da vicino se, e come, sia possibile fare uso dei risultati limitativi della logica, ed in particolare dei teoremi di Gödel, per sostenere l'idea che la mente umana non possa essere una macchina.

Prima di esaminare come quella possibilità, che emerge già negli scritti di Turing ma che abbiamo visto profilarsi anche in alcuni lavori dello stesso Gödel, sia stata ripresa ed elaborata da autori come Lucas e Penrose, vogliamo brevemente soffermarci sul problema posto dalla tesi meccanicista.

Che vi sia un problema di tipo terminologico dietro la posizione secondo la quale è possibile fornire, per usare una formula altrettanto generica ma diversa da quella richiamata nel capoverso iniziale, una spiegazione computazionale dei processi razionali, lo si poteva intuire già dalla disamina della posizione di Turing nell'ultimo paragrafo del capitolo precedente.

Con l'espressione 'tesi meccanicista' si intende infatti comunemente indicare un punto di vista secondo il quale, in un qualche senso, la mente umana (o anche la ragione o l'intelletto), è interamente riducibile ad un'attività di tipo meccanico. Il problema sta proprio nel precisare in quale senso si sostenga una posizione del genere (o qualcosa di equivalente ad essa).

In prima approssimazione, infatti, si possono distinguere due significati possibili:

1. in un primo senso, la mente è riducibile ad una macchina nel senso che il risultato della propria attività non può andare oltre quello di un

dispositivo meccanico, anche se la mente *non è*, materialmente, una macchina (indichiamo questa posizione come un meccanicismo di tipo *estensionale*);

2. altrimenti, si potrebbe ritenere che la mente umana sia riducibile ad una macchina in quanto essa *è*, a tutti gli effetti, un dispositivo meccanico, seppur con le sue specificità: le sue parti sono costituite da cellule, neuroni, sinapsi, anziché valvole, circuiti o quant'altro (questa è la versione *intensionale* del meccanicismo).

Per quanto a prima vista diverse (propriamente, secondo la prima posizione, la mente umana è *meccanizzabile*, in linea teorica, laddove, sulla base della seconda, essa è *meccanica* nella pratica), le due forme di meccanicismo non portano ad esiti opposti: se si sostiene la versione intensionale, si vorrà forse concedere che sia possibile in effetti trovare un equivalente estensionale della mente, se si ritiene invece che la formulazione estensionale sia quella più plausibile, si potrà riconoscere che le differenze costitutive possano essere messe in secondo piano perché inessenziali da un punto di vista pratico.

Quando il problema del meccanicismo si misura su di un piano logico, ossia in relazione a certi risultati dell'indagine metamatematica, occorre tenere presente inoltre che si ha a che fare principalmente con una forma *teorica* dell'ipotesi meccanicista, piuttosto che con una sua versione *pratica*: dicendo che la mente umana è riducibile ad una macchina, non stiamo pensando, con quest'ultimo termine, ad un oggetto sensibile (ad esempio ad un calcolatore digitale), bensì a quel modello astratto rappresentato dalle macchine di Turing.

Ad un primo sguardo, questa formulazione della tesi meccanicista presenta il vantaggio di semplificare notevolmente la situazione. In un senso, infatti, essa rende superflua la distinzione tra meccanicismo intensionale ed estensionale: nessuno vorrà sostenere che la mente umana è *letteralmente* una macchina di Turing, bensì che per ogni operazione della mente (che non sia necessariamente il calcolo di una funzione numerica), possiamo scrivere una lista finita di istruzioni per arrivare allo stesso risultato (si avrebbe così una sorta di 'correlato cognitivo' della Tesi di Church-Turing: tutto ciò che la mente *fa*, è Turing-computabile).

Appellandosi all'universalità del modello matematico, poi, si avrebbe con questa formulazione un modo semplice per passare rapidamente dal piano pratico a quello teorico: se si riesce a dimostrare che per una data operazione della mente esiste una macchina in grado di eseguirla, possiamo concludere che, almeno in linea di principio, esiste una macchina di Turing in grado di eseguire esattamente lo stesso compito. Per contrapposizione, possiamo, per lo stesso motivo, ritenere implausibile di poter arrivare in pratica a costruire un dispositivo meccanico per simulare le operazioni della mente, se si è mostrato, facendo uso dei risultati limitativi, che non sussiste la possibilità logica che una macchina di Turing possa essere equivalente alla ragione umana.

Ciò non toglie che, come rivelano già le precedenti considerazioni ad un simile livello 'ingenuo', certe distinzioni e la necessità di precisare i termini della questione siano tutt'altro che secondarie. Questo fatto ben lo si apprezza considerando proprio le posizioni al riguardo di Gödel e Turing.

Per poter dire che Gödel era anti-meccanicista, in quanto sostenitore dell'idea che la mente umana sorpassi le capacità di una macchina di Turing, occorre specificare che con 'mente umana' non si intende 'cervello', rispetto al quale Gödel era invece un meccanicista intensionale. Turing, dal canto suo, era probabilmente anch'egli un anti-meccanicista teorico (nessuna macchina di Turing può contenere in sé il frutto dell'attività intuitiva della mente umana), pur essendo, c'è da presumerlo almeno, un meccanicista sul piano pratico: egli riteneva possibile produrre un dispositivo in grado di essere scambiato per un ente dotato di intelligenza (dove il problema se esso *sia*, in effetti, tale, verrebbe considerato come qualcosa di indeterminato, e di indeterminabile su base scientifica).

Con questa premessa, intendiamo solo indicare alcuni temi che ci pare importante vengano tenuti presenti in relazione all'analisi delle supposte confutazioni della tesi meccanicista ispirata dai teoremi di indecidibilità. Alcuni di essi, verranno ripresi e discussi nuovamente nel paragrafo conclusivo del presente capitolo.

4.1 L'argomento gödeliano di J. R. Lucas

Nel 1961 esce sulla rivista "Philosophy", a firma del filosofo inglese J. R. Lucas, un articolo dal titolo "Menti, macchine e Gödel"[411]. Il contributo, si ripromette di confutare il meccanicismo mostrando, attraverso il riferimento decisivo ai teoremi di incompletezza, come "le menti non possano esse spiegate come macchine"[412].

[411] [Luc61].

[412] [Luc61], p. 43. *NdA*: le citazioni tratte dall'articolo di Lucas si riferiscono alla ristampa del lavoro nella raccolta curata da A. R. Anderson, come recita la voce bibliografica.

Sarà opportuno, prima di esaminare più a fondo gli snodi essenziali dell'analisi di Lucas, riprodurre innanzi tutto nella forma schematica di un'inferenza, i passaggi che costituiscono ciò che è divenuto noto in letteratura come l' "argomento di Lucas", reso celebre dal vivace dibattito al quale esso ha dato adito[413]:

(A) "Il [primo] teorema di Gödel, afferma che in ogni sistema formale consistente [e ricorsivamente assiomatizzabile] che sia sufficientemente potente per riprodurre l'aritmetica elementare, ci sono formule che non possono essere dimostrate-nel-sistema, ma che noi possiamo riconoscere come vere."[414]

(B) "Il teorema di Gödel deve valere per macchine cibernetiche, poiché è nell'essenza dell'essere una macchina che essa sia una istanza concreta di un sistema formale."[415]

(C) "Segue [da **(A)** e **(B)**] che data una macchina consistente e capace di eseguire l'aritmetica elementare, c'è una formula che essa è incapace di produrre come vera [...] ma che noi possiamo vedere come tale."[416]

(D) "Ne consegue che nessuna macchina può essere un modello completo e adeguato della mente, che la mente è essenzialmente diversa da una macchina."[417]

Tra i punti cruciali dell'argomentazione, vi è la giustificazione della premessa **(A)**, che Lucas elabora nel modo seguente:

> Essenzialmente, si considera la formula che dice, in effetti, "Questa formula è indimostrabile-nel-sistema". Se questa formula fosse dimostrabile-nel-sistema, si avrebbe una contraddizione: infatti, se fosse dimostrabile-nel-sistema non sarebbe indimostrabile-nel-sistema, così che "Questa for-

[413] Tutto ciò nonostante che Lucas non sia stato affatto il primo ad elaborare l'argomento in questione. A parte il caso di Turing, del quale si è già avuto modo di dire, Smart, nel suo [Sma61], propone una critica dell'argomento così come esso viene formulato da P. Rosenbloom, in "Elements of Mathematical Logic" (Dover, 1950), a p. 208, da J. G. Kemeny in "A Philosopher Look at Science" (van Nostrand, 1959), a p. 224, e da E. Nagel e J. R. Newman nella loro monografia dedicata ai teoremi di Gödel dal titolo "Gödel's Proof" (Routledge e Kegan Paul, 1959), in particolare alle pp. 100-101.
[414] *Ibid.*, pp. 43-44.
[415] *Ibid.*, p. 44.
[416] *Ibid.*, p. 44.
[417] *Ibid.*, p. 44.

mula è indimostrabile-nel-sistema" sarebbe falsa; allo stesso tempo, se fosse dimostrabile-nel-sistema non sarebbe falsa, bensì vera dal momento che in un sistema consistente solo formule vere e nessuna falsa possono essere dimostrabile-nel-sistema. Dunque, la formula "Questa formula non è dimostrabile-nel-sistema" non è dimostrabile-nel-sistema ma indimostrabile-nel-sistema. Inoltre, se la formula "Questa formula non è dimostrabile-nel-sistema" è indimostrabile-nel-sistema, è vero che la formula è indimostrabile-nel-sistema, quindi "Questa formula non è dimostrabile-nel-sistema" è vera.[418]

Al di là di una indubbia osticità di lettura del passo, non dovrebbe risultare difficile ricostruire ciò che Lucas ha in mente con l'argomento **(A)**-**(D)**. Dato un sistema formale S che soddisfa le ipotesi dei teoremi di Gödel (oppure, equivalentemente, una "macchina cibernetica" adatta alla dimostrazione di teoremi aritmetici), possiamo costruire un enunciato che *noi* riusciamo a riconoscere come vero, laddove S non è in grado né di dimostrarlo, né di refutarlo. Poichè facendo ciò, 'trascendiamo' le capacità del sistema S (facciamo qualcosa che S non è in grado di fare), non possiamo essere (equivalenti a) S[419].

Il punto essenziale che si cela dietro il riferimento alla nostra capacità di 'trascendere' ogni sistema formale che soddisfi le ipotesi dei teoremi di Gödel, lo si può spiegare in vari modi.

Quello che Lucas pare voler indicare, riferendosi alla verità di ciò che 'dice' l'enunciato indecidibile di Gödel, sembra essere il seguente: dimostrando il primo teorema di Gödel si ricava in particolare che, per un dato S, c'è un enunciato G_{S} indimostrabile; si ha quindi che l'enunciato $\forall x \neg Dim_{\mathsf{S}}(x, \ulcorner G_{\mathsf{S}} \urcorner)$,

[418] *Ibid.*, p. 44.

[419] Si sarà notato che la giustificazione da parte di Lucas del passaggio cruciale del proprio argomento, contenuta nel passo succitato, rivela apparentemente una confusione tra la consistenza di un sistema formale (che equivale a dire che 'nessuna formula falsa è formalmente dimostrabile in esso') e la correttezza del medesimo ('tutto cio che si dimostra in esso è vero'). Curiosamente, una tendenza a confondere queste due proprietà risulta abbastanza frequente in questo genere di discussioni sui teoremi di Gödel (alcuni dei passaggi dubbi dello stesso tenore che si possono trovare nella discussione da parte di Penrose dell'argomento analogo da lui elaborato - cfr. [Pen96], p. 104 e segg - sono stati stigmatizzati da Feferman in [Fef95]). La sorpresa nasce dal fatto che proprio i teoremi di incompletezza offrono notoriamente l'occasione di chiarire bene la differenza tra le due nozioni: se G_S è l'enunciato di Gödel indecidibile in un sistema S, allora $\mathsf{S} \cup \neg G_S$ è consistente, ma 1-inconsistente e dunque, essendo la 1-consistenza una forma debole di correttezza (come è noto, si dimostra che S è 1-consistente sse è corretto rispetto agli enunciati di tipo Σ^0_1), non è corretto.

che esprime questo fatto, è vero[420]; ma questo stesso enunciato è, per costruzione, logicamente equivalente a G_{S}; dunque, quest'ultimo è vero e indimostrabile in S[421].

L'idea, è quindi quella secondo la quale, per mezzo di un unico argomento (la prova di Gödel), si ricava tanto la verità ai nostri occhi dell'enunciato G_{S}, quanto la sua indimostrabilità in S. Dal ché segue, che rispetto ad ogni sistema formale entro il quale si sia in grado di eseguire un *minimum* di teoria dei numeri elementare, un requisito irrinunciabile se si intende affermare in relazione ad esso l'equivalenza con le capacità matematiche della ragione umana, possiamo esibire una verità aritmetica che non fa parte dei teoremi del sistema. Sulla base di una corrispondenza tra sistemi formali e macchine (come recita la premessa **(B)**), si può concludere che nessuna macchina è equivalente alla mente umana.

Quanto detto, dovrebbe bastare per comprendere quali potrebbero essere i punti di forza di una simile argomentazione, qualora essa venisse riconosciuta come ineccepibile: si sarebbe infatti in grado di presentare un argomento che ha i requisiti del *rigore*, in quanto basato su un teorema logico (fatto, questo, particolarmente rilevante proprio in quanto si intende confutare una tesi filosofica), e dell'*uniformità*, nel senso che quel teorema garantisce la possibilità di mostrare l'inadeguatezza di *ogni* sistema formale (che soddisfi certi requisiti minimali).

Infatti, se si è 'trasceso' S facendo riferimento (come nella nostra spiegazione di poc'anzi del passo di Lucas) al concetto di 'verità' per il linguaggio di S, che non è formalizzabile in S stesso (almeno, non fino al punto tale da poter dimostrare la verità dell'enunciato G_{S}), e se è pur vero che la verità dell'enunciato indecidibile può essere desunta in un opportuno sistema S^* che estende S (e nel quale sia possibile definire tarskianamente un predicato di verità per il linguaggio del sistema di partenza), è anche vero che, per

[420] Ovviamente, indichiamo con $Dim_{\mathsf{S}}(x, y)$ il predicato primitivo ricorsivo di dimostrabilità per S, che sussiste, cioè, se e solo se x è il codice di una dimostrazione la cui ultima formula è codificata da y.

[421] In questi termini, il riferimento alla sola ipotesi di consistenza di un sistema formale può essere giustificato: l'argomento di Lucas sfrutta esclusivamente il fatto che l'enunciato di Gödel è indimostrabile (e non il fatto che esso risulta essere anche irrefutabile), che discende, appunto, dalla sola consistenza della teoria assiomatica. Come è noto, la consistenza di una teoria assiomatica equivale alla correttezza per la collezione di enunciati di tipo Π^0_1, ossia proprio della forma logica dell'enunciato indecidibile. In *questo* senso, è vero che un sistema consistente è anche corretto.

mezzo dello stesso argomento, si può dimostrare che c'è un enunciato, G_{S^*}, indecidibile in S^*, che noi riconosciamo essere vero.

Il punto essenziale, come dimostrerebbe, iterata tale procedura, una simile 'versione infinitaria' dell'argomento di Lucas (perché non esiste estensione di un sistema logico, o di un dispositivo meccanico, rispetto alla quale non si abbia la possibilità di applicare la medesima costruzione), è che, a detta del filosofo di Oxford, con i teoremi di Gödel si dispone di uno strumento per evidenziare, nella 'partita filosofica' con il meccanicista, i difetti della tesi avversaria:

> È come un gioco. Il meccanicista ha la prima mossa. Egli fornisce *un - qualsiasi*, ma *definito* - modello meccanico della mente. Io punto l'indice su qualcosa che esso non è in grado di fare, ma che la mente può fare. Il meccanicista è libero di modificare il suo esempio, ma ogni volta che lo fa, ho il diritto di cercare i difetti del modello rivisto. Se il meccanicista riesce ad esibire un modello rispetto al quale non sono in grado di scoprire alcun difetto, la sua tesi è stabilita: se non può non è dimostrata; e dal momento che - come risulta - egli non è in grado [di fare ciò] in modo necessario, [la sua tesi] è refutata.[422]

Un'ampia parte del lavoro di Lucas, è dedicata proprio a fugare ogni dubbio che l'argomento da lui proposto non sia sufficiente a questo scopo.

A tal fine, egli si dedica all'analisi di due possibili obiezioni: (i) l'idea che il modello al quale si applica l'argomento elaborato (le macchine di Turing, sostanzialmente) sia euristicamente, prim'ancora che logicamente, inadeguato, e che, dunque, la forza dell'inferenza proposta dovrebbe essere misurata rispetto a sistemi che permettano, in particolare, conclusioni per via non deduttiva; (ii) il fatto che i teoremi di incompletezza consentano in realtà una conclusione ipotetica, *se* un sistema formale è consistente *allora* l'enunciato di Gödel è indecidibile: ciò, si potrebbe ritenere, sembra inficiare l'uso categorico che, al contrario, Lucas fa, in **(A)**-**(D)**, di quel risultato.

Riguardo alla prima possibile obiezione, Lucas è, a dire il vero, estremamente sbrigativo. Egli si limita infatti a sottolineare come la costruzione di una macchina euristicamente plausibile "presenti delle difficoltà". Essa dovrebbe, in particolare, essere dotata di un meccanismo di selezione delle congetture, ossia delle inferenze non deduttive, dal momento che non tutte le proposizioni indecidibili sono sullo stesso piano: nel caso ad esempio dell'enunciato di Gödel, la suddetta macchina dovrebbe essere in grado di assumere, seppur in forma ipotetica, la verità dell'enunciato indecidibile, ma

[422] *Ibid.*, p. 50.

non della sua negazione, la quale è, sulla base dell'argomento proposto, falsa seppur parimenti consistente con gli altri principi di base. Lucas ritiene quindi che non sia affatto chiaro, come si possa ovviare alle difficoltà senza adottare criteri di scelta puramente casuali.

La conclusione, è quindi che o non si raggiunge la plausibilità euristica voluta, oppure non si riesce affatto ad evitare gli effetti dell'argomento gödeliano:

> In breve, comunque sia progettata una macchina, o essa procede in modo casuale, o in accordo a regole definite. Fin tanto che essa procede casualmente, non riusciamo a smascherarla: ma la sua azione non sarà certo una convincente parodia di un comportamento intelligente; fin tanto che il suo modo di procedere segue regole definite, il metodo di Gödel può essere usato per produrre una formula che la macchina non è in grado di asserire essere vera per mezzo di quelle regole, mentre noi, essendo all'esterno del sistema, possiamo riconoscerla come tale.[423]

L'obiezione fondata sull'ipotesi di consistenza, dalla quale dipende la possibilità di applicare i teoremi di incompletezza, viene da Lucas interpretata in un modo del tutto peculiare. Egli non si limita, infatti, come ci si potrebbe aspettare, a considerare l'obiezione che la consistenza sia un requisito essenziale per l'uso dei teoremi di Gödel che egli fa[424]. Piuttosto, egli sembra aver in mente una posizione più complessa che utilizza il secondo teorema di incompletezza[425]: poiché l'enunciato di Gödel è indimostrabile e vero solo se il sistema in relazione al quale è stato costruito è consistente, occorre innanzi tutto accertare quest'ultimo fatto; ma sulla base del secondo teorema di Gödel, una prova di consistenza non può essere assoluta: si ha dunque che, in ultima istanza, "la macchina è consistente, supposto che lo siamo noi"[426]; prima ancora che utilizzare il primo teorema di incompletezza per mostrare che l'uomo non è una macchina, occorrerebbe dunque far vedere che l'uomo è in grado di fare ciò che ad un sistema formale non è consentito per il secondo teorema di Gödel: asserire la propria consistenza rimanendo consistente.

Ciò spiega perché Lucas prenda in esame tutta una serie di motivi per i quali la mente umana non può essere inconsistente.

[423] *Ibid.*, pp. 51-52.

[424] Banalmente, se si adotta il criterio di Lucas secondo il quale l'enunciato di Gödel, che 'dice di se stesso' di essere indimostrabile, è vero in quanto si verifica ciò che esso dice, si ricava che se il sistema è inconsistente (e non si fanno restrizioni logiche) esso è dimostrabile e dunque falso.

[425] Cfr. *Ibid.*, p. 52.

[426] *Ibid.*, p. 52.

Non lo è, in primo luogo, perché per quanto anche localmente contraddittori, gli esseri umani non lo sono al punto da poter sostenere che ciò caratterizzi il modo di procedere della mente: "[g]li esseri umani, per quanto non perfettamente consistenti, non sono così inconsistenti da essere fallaci"[427]. Una dimostrazione di questo fatto, è ad esempio il modo con il quale si tende ad evitare ogni forma di contraddizione.

Ma, per Lucas, non appare verosimile neanche l'idea, che taluni potrebbero sostenere a fronte del fatto che rifuggiamo dalle contraddizioni, che la mente umana possa essere inconsistente, seppure provvista di una sorta di 'regola d'arresto' la quale, entrando in funzione ogni qual volta una contraddizione si profila all'orizzonte di un ragionamento deduttivo, impedisce che si giunga in effetti ad una tale conclusione:

> Nessun credito vanta l'uomo che, abbastanza intelligente da vedere qualche mossa argomentativa avanti, evita di essere condotto a riconoscere la propria inconsistenza, chiudendo le porte [ad essa] nel momento in cui egli realizza dove andrebbe a finire il proprio argomento. Piuttosto, lo riteniamo inconsistente ugualmente, non, in questo caso, per aver affermato e negato la stessa proposizione, ma perché egli utilizza [certe regole di inferenza per parte dell'argomento] e si rifiuta di usare [fermandosi] la stessa regola di inferenza.[428]

Colui che voglia sostenere, in qualsiasi forma, l'idea che la conoscenza umana potrebbe essere inconsistente, deve spiegare perché, a detta di Lucas, una simile ipotesi appaia del tutto in contrasto con la totale assenza di tolleranza che mostriamo nei confronti di una qualsiasi contraddizione che dovesse presentarsi in una data area del sapere. Questa stessa ragione, viene opposta a quanti ritengano possibile che, per quanto inconsistente, la mente umana lo sia in un modo così complesso da non aver ancora mostrato in modo palese questa sua natura: "[s]iamo determinati a non essere inconsistenti, e siamo risoluti ad estirpare l'inconsistenza, dovesse mai presentarsi"[429].

Ciò è sufficiente, secondo Lucas, per rigettare ogni tentativo di sfuggire all'argomento gödeliano per questa via:

> Non solo possiamo in modo corretto dire semplicemente che *sappiamo* di essere consistenti, a prescindere dai nostri errori, ma dobbiamo in ogni caso *assumere* che lo siamo, se si vuole che il pensiero sia possibile; in più,

[427] *Ibid.*, p. 53.
[428] *Ibid.*, p. 55.
[429] *Ibid.*, p. 55.

siamo selettivi, non diremmo qualsiasi cosa e di qualunque genere, come farebbero invece delle macchine inconsistenti: ed infine possiamo, in un certo senso, *decidere* di essere consistenti, nel senso che possiamo risolverci a non tollerare contraddizioni nel nostro pensiero e nei nostri discorsi, e ad eliminarle, se esse dovessero manifestarsi, ritirando e cancellando un corno della contraddizione.[430]

La critica dell'obiezione basata sul ruolo della consistenza nei teoremi di incompletezza, è il passaggio preparatorio per la conclusione di Lucas che è volto a completare la trasformazione del risultato di Gödel in una vera e propria tesi filosofica.

Dal fatto che l'enunciato indecidibile è auto-referenziale, il filosofo inglese trae l'indicazione che la differenza tra le macchine, che sono "oggetti inanimati", e gli esseri umani, in quanto "esseri conscienti", stia proprio nell'incapacità delle prime, e nella possibilità dei secondi, di riflettere su se stessi e sulla propria attività:

> Una macchina può essere costruita per così dire per "considerare" la sua stessa azione, ma non è in grado di prendere questa "in considerazione" senza diventare una macchina differente, cioè la vecchia macchina con una "nuova parte " aggiunta. Ma è qualcosa di inerente alla nostra idea di una mente cosciente che essa possa riflettere su se stessa e criticare le sue stesse azioni, e nessuna parte extra è necessaria a fare ciò: essa è già completa, e non possiede alcun tallone d'Achille.[431]

L'idea alla base di questo passo è il punto cruciale della conclusione dell'articolo di Lucas: nessuna macchina, la quale, come si può mostrare facendo uso dei teoremi di incompletezza, non è in grado di comprendere il proprio funzionamento in modo soddisfacente, può essere equiparata alla mente umana, che ha, dopotutto, nell'autocoscienza (come risulta dalla capacità di sfuggire, per contro, all'argomento gödeliano) un suo punto distintivo.

Ciò è vero al punto tale per Lucas, che una macchina che non cadesse sotto gli effetti dell'argomento da lui elaborato non sarebbe essa stessa una macchina: chiunque potesse vantare di aver dato vita ad un simile dispositivo, condurrebbe non tanto a negare la valenza dell'argomentazione basata sui teoremi di Gödel, quanto piuttosto a ritenere "che egli abbia creato una mente, nello stesso senso in cui noi procreiamo individui allo stato delle cose"[432].

[430] *Ibid.*, p. 56.
[431] *Ibid.*, p. 57.
[432] *Ibid.*, p. 58.

I teoremi di incompletezza, svolgono la funzione essenziale di permettere l'individuazione di un confine preciso in questo senso. Non solo perché, in quanto risultati che si applicano ad ogni sistema formale (o macchina di Turing) 'sufficientemente potente', rivelano un limite essenziale di questi 'oggetti inanimati'; ma anche perché, applicandosi *solo* ad essi, denotano l'elemento discriminante di quanti, in quanto 'esseri coscienti', o meglio, 'auto-coscienti', a quei risultati sfuggono.

4.2 Alcune osservazioni

Rispetto alla discussione introduttiva del capitolo, si sarà forse notato come la disamina della tesi meccanicista in relazione a certi risultati limitativi della logica, presenti due indubbi vantaggi. Il primo, consiste nella possibilità di dare una formulazione precisa di quella tesi: essa si riduce all'idea che vi sia un sistema formale (o una macchina di Turing) la cui estensione, ossia l'insieme dei teoremi (o degli *output*), coincide con quella della mente umana. Sotto questo punto di vista, anche la tesi anti-meccanicista di Lucas può essere indubbiamente formulata in modo chiaro: anche confinandosi ad un sottoinsieme proprio degli enunciati di un linguaggio formale (la collezione delle formule chiuse di tipo Π^0_1), dato un qualsiasi sistema formale, consistente, ricorsivamente assiomatizzabile e 'contenente' una certa quantità di aritmetica primitiva ricorsiva, c'è un enunciato che risulta vero alla mente umana e che è indimostrabile (addirittura indecidibile) nel sistema.

Un secondo vantaggio lo si è già ricordato in apertura di capitolo, e, sebbene non sia menzionato in modo esplicito, sembra condizionare la discussione di Lucas. Se si assume che una macchina di Turing è in grado, in linea di principio, di riprodurre, attraverso un'opportuna 'programmazione', ogni operazione meccanica, non è restrittivo limitarsi a considerare questo modello astratto e matematicamente preciso: piuttosto, se esso dovesse rivelarsi inadeguato (nel senso estensionale suddetto) rispetto alla mente, si dovrebbe coerentemente presumere che non può esserci dispositivo fisico in grado di far meglio.

Detto questo, occorre certo vedere se si può ritenere l'argomentazione di Lucas, e le sue considerazioni a latere, indiscutibilmente soddisfacenti.

Un primo aspetto di cui si deve tenere conto, è il contesto nel quale la posizione del filosofo inglese matura e che è ben descritta, dallo stesso Lucas, in un recente articolo retrospettivo sul proprio lavoro del 1961:

> Il mio articolo originario, “Minds, Machines and Gödel”, fu scritto sull'onda del lavoro di Turing uscito su *Mind* nel 1950, ed era stato concepito per mostrare che la mente umana non è una macchina di Turing. Perché dunque, non ho formulato l'argomento in termini del teorema di Turing, che è più facile da dimostrare e si applica direttamente alle macchine di Turing, anziché del teorema di Gödel, che è così terribilmente difficile da dimostrare e non si applica così direttamente o in maniera ovvia alle macchine? La ragione è che il teorema di Gödel mi dava di più: esso solleva questioni relative alla verità che riposa in modo evidente sulla natura della mente umana, laddove il teorema di Turing non fa altrettanto; non solo si dimostra infatti che l'enunciato Gödeliano è indimostrabile-nel-sistema, ma anche che esso è vero. [...] Il teorema di Turing, sebbene concluda con lo stesso risultato negativo del teorema di Gödel, che certi compiti non possono essere portati a termini da un computer idealizzato, non raggiunge l'elemento positivo ulteriore che noi, in quanto agenti razionali, possiamo portare a termine proprio il compito che al computer non riesce.[433]

Dunque, lo spunto per la critica al meccanicismo fu rappresentata dalla lettura dell'articolo di Turing del 1950. Lucas, ritenne cioè di potersi opporre a Turing (supponendo che egli abbia voluto affermare una tesi di quel tipo), riformulando l' “obiezione matematica”, considerata dal matematico britannico nel proprio contributo, utilizzando i teoremi di incompletezza piuttosto che il teorema dell'arresto. Questo, per riuscire a mostrare ad un tempo, non solo che c'è qualcosa che una macchina di Turing non è in grado di fare (dimostrare un certo enunciato Π^0_1), ma anche che, su questo, la mente umana esibisce il proprio carattere distintivo portando a termine il compito.

Non ci sarebbe quindi bisogno di alludere, come si è visto fare Turing discutendo la propria versione dell' “obiezione matematica”, alla possibilità del matematico umano di scoprire *nuove* strategie dimostrative che lo potrebbero metter in grado di risolvere il problema della veridicità dell'enunciato di Gödel, laddove invece una macchina è costretta ad applicare sempre lo stesso ciclo di operazioni: già la stessa dimostrazione dei teoremi di incompletezza permette, secondo quanto sostiene Lucas, di concludere ciò. E rispetto a questo fatto, a nulla varrebbero ovviamente le risposte elaborate al proposito dallo stesso Turing, in particolare, quella relativa alla possibilità di poter dare vita a sistemi sempre più potenti.

Su questa conclusione, che ci pare essere implicitamente chiamata in causa nel passo di Lucas succitato, occorre fare, a nostro avviso, due osservazioni.

La prima, consiste nel fatto che la posizione di Turing si comprende se

[433] Cfr. [Luc96], p. 103

contestualizzata, come si è cercato di chiarire nel capitolo precedente, allo scopo che egli si era realmente ripromesso: mostrare la praticabilità della costruzione di una macchina che esibisse un comportamento intelligente (al punto da essere scambiata, a certe condizioni, per un essere umano), senza fare alcuna assunzione di tipo filosofico sull'intelligenza, o, tanto meno, sulla natura della mente umana.

In secondo luogo, prima ancora di entrare nel merito del giudizio sulla correttezza logica dell'argomento di Lucas, si dovrebbe notare che è possibile dare alle risposte di Turing una veste che permette loro di acquisire senso, anche in relazione alla posizione del filosofo inglese[434]. Risulta infatti, che quello stesso 'punto di forza' sul quale sembra far leva Lucas, ossia l' 'uniformità' dell'argomento (dal quale discende la possibilità di mostrare l'inadeguatezza di ogni macchina di Turing, per mezzo della stessa, unica costruzione reiterata), può essere usato per segnalarne un elemento di debolezza.

Infatti, il metodo di Gödel mediante il quale, dato un insieme ricorsivamente enumerabile di assiomi e regole di inferenza S, si costruisce l'enunciato G_{S} che 'dice di se stesso di essere indimostrabile in S', non è solo uniforme ma anche effettivo. Esiste dunque una macchina di Turing che associa ad un dato sistema formale S, il sistema $\mathsf{S}' := \mathsf{S} \cup \{G_{\mathsf{S}}\}$ (o, meglio, all'insieme dei teoremi del primo, l'insieme dei teoremi del secondo). Questo fatto, che è ovviamente legato all'idea costitutiva delle logiche ordinali, potrebbe essere visto come il sostegno rigoroso al punto di vista Turing, secondo il quale non dovremmo "mostrare soddisfazione" per il fatto di essere riusciti ad indurre in errore una macchina, dal momento che potremmo dare vita ad un dispositivo più potente, e poi ad un altro ancora più potente, e così via[435].

Come si è visto, anche Lucas, come Turing, espone la propria posizione utilizzando la metafora del gioco (che racchiude, come si è cercato di mostrare nel paragrafo precedente, l'essenza dell'uniformità della propria argomentazione). Nel gioco da lui immaginato, nel quale il meccanicista propone un modello computazionale della mente e l'anti-meccanicista risponde mostrandone i difetti, non ci sarebbe affatto bisogno per il primo, fin tanto che il suo

[434]L'obiezione seguente viene ricondotta da J. J. C. Smart, nel suo [Sma61], ad un suggerimento di Quine, e proposta come una sorta di controparte matematica del proprio rilievo (mosso tuttavia non nei confronti dell'argomento di Lucas, ma di suoi equivalenti - si veda quanto detto alla nota 413), legato alla possibilità di dare vita ad una macchina che ad ogni stadio, per via induttiva, sia in grado di aggiungere alle proprie conoscenze l'enunciato indecidibile di Gödel.

[435]Cfr. [Tur50], p. 445, e la nostra discussione del passo nel §3.3.2.

avversario ha come strategia quella dell'argomento gödeliano, di fornire ad ogni 'partita' un modello diverso: sarebbe sufficiente piuttosto proprio quella procedura effettiva che, partendo da un certo S_0, in corrispondenza ad ogni numero naturale (o se vogliamo adottare una prospettiva ancor più idealizzata, in corrispondenza ad ogni ordinale costruttivo), genera la successione $\mathsf{S}_0', \mathsf{S}_0'', \ldots$.

Ma allora, è certamente conveniente, per Lucas, vedere nella possibilità dell'anti-meccanicista di evidenziare il difetto di ogni proposta e di giocare la partita all'infinito, non semplicemente una prova della propria superiorità contingente, bensì addirittura la dimostrazione della sua *diversità*, ma appare anche arbitrario non considerare affatto la capacità analoga (di 'rispondere' indefinitamente) del modello elaborato, sulla scorta della costruzione di Gödel, dal suo avversario.

Forse, Turing avrebbe pensato che, oltre a riferirsi ad un tipo di macchina di per sé insoddisfacente, sono le condizioni del gioco (che, in sostanza, lascia al meccanicista l'ultima parola), a creare l'impressione di una differenza essenziale tra i giocatori, e che, in circostanze 'più corrette' per la macchina (presumibilmente sulla falsariga del suo Gioco dell'Imitazione), questa stessa impressione potrebbe venire meno[436].

Questo, è quanto ci pare doveroso osservare, innanzi tutto, rispetto all'idea che il contributo di Lucas possa essere letto come una risposta al lavoro di Turing del 1950.

Oltre a ciò, ovviamente, c'è la questione di merito se l'argomento elaborato dal filosofo inglese possa essere considerato a tutti gli effetti come logicamente ineccepibile. Sotto questo aspetto, la questione cruciale, che non appare essere trattata da Lucas in modo esauriente, è quella della consistenza. Essa è tanto più importante, in quanto è alla base della possibilità di rispondere in modo affermativo alla domanda: Ma è proprio vero che nel gioco descritto in precedenza, la mente umana è in grado di continuare la partita all'infinito?

Dietro l'idea che l'argomento di Lucas permetta, di fronte ad ogni 'proposta' del meccanicista, di esibire un enunciato vero ma indimostrabilmente tale in termini meccanici, si cela infatti, come si è già detto, un'assunzione

[436] Questa osservazione potrebbe apparire estremamente vaga. Si rimanda, per una spiegazione più dettagliata, al §5.3, nel quale si riprende la discussione del gioco dell'imitazione, in relazione ad un modello logico euristicamente più plausibile dei sistemi formali, e si misurano, sui risultati, le aspettative di Turing.

sulla possibilità di trasformare il primo teorema di Gödel da ipotetico (se S è consistente, allora è sintatticamente incompleto), a categorico.

Il modo con il quale Lucas risponde all'obiezione basata sul fatto che la consistenza di un sistema formale è un requisito essenziale per l'argomento gödeliano da lui elaborato, può essere fuorviante.

Non si tratta infatti semplicemente di confutare l'idea, contro la quale Lucas si accanisce, che la mente umana potrebbe essere equivalente ad una macchina di Turing inconsistente. Probabilmente, egli è portato fuori strada dai riferimenti di Turing al requisito di fallibilità per le macchine da lui pensate come controparte della ragione umana. Tuttavia, da questo punto di vista, si è già cercato di mostrare, nel capitolo precedente, come quelle allusioni vadano intese alla luce di un proposito che è (come lo era rispetto all'analisi di Gödel della *Gibbs Lecture*), sotto certi aspetti inconfrontabile con lo scopo che Lucas si prefigge: laddove quest'ultimo cerca di mostrare l'impossibilita teorica di una certa concezione filosofica, Turing si propone di sostenere la plausibilità pratica di una ricerca scientifica.

Il fatto è, piuttosto, che ciò che dai teoremi di incompletezza si ricava è che, dato un certo sistema formale S, *se* esso è consistente, allora (i) G_{S} è indimostrabile in S, e, in quanto tale, (ii) G_{S} è un enunciato vero. Nel dire, quindi, come fa Lucas, che la mente umana è *sempre* in grado di riconoscere come vero G_{S} (per *ogni* sistema formale consistente S), si sta in realtà dicendo che la mente umana è sempre in grado di accertare in modo veritiero il fatto che ne è alla base, ossia la consistenza di S.

Una simile ipotesi, appare dubbia per due motivi: da un punto di vista logico, perché si fonda l'esistenza di un enunciato Π^0_1 vero intuitivamente e indimostrabile formalmente, sulla capacità di riconoscere la verità di un enunciato, quello che esprime la consistenza del sistema formale in questione, anch'esso di tipo Π^0_1 (non si riduce, quindi la verità di G_{S} a quella di un enunciato più semplice); da un punto di vista matematico, perché con ciò si potrebbe voler dire che si è sempre in grado di fornire una prova della consistenza di un sistema formale, cosa che, rispetto a teorie assiomatiche particolarmente complesse, non è affatto pacifica.

Occorre notare inoltre, che niente nell'affermare che la mente umana sia in grado di "asserire", "conoscere" o "decidere" circa la propria consistenza, come fa Lucas in un passo citato, è logicamente incompatibile con la tesi meccanicista, fin tanto che ciò non voglia dire "dimostrare". Solo in quest'ultimo caso, infatti, il matematico umano esibirebbe un carattere che, sulla scorta del secondo teorema di Gödel, non è riproducibile a livello dei sistemi formali.

Lo sforzo di Lucas per mostrare come l'evenienza di un meccanismo inconsistente alla base della mente sia altamente implausibile, insomma, non riesce ad eliminare completamente il dubbio che il suo argomento non abbia, in relazione alle ipotesi dei teoremi di incompletezza, un punto debole decisivo[437].

In conclusione, quindi, ci pare che (oltre a non fugare i dubbi circa la correttezza dell'argomento ivi presentato) non solo il lavoro di Lucas non offra una critica corretta alla posizione di Turing, perché non pare coglierne il vero spirito, ma che esso rischi persino di confermare l'idea del matematico britannico che, tra i problemi del progetto di una 'macchina pensante', vi potessero essere quegli atteggiamenti radicati nei luoghi comuni secondo i quali una macchina è *per definizione* priva di intelligenza.

Un approccio al problema di quel tipo, sembra essere sottinteso nella parte finale dell'articolo di Lucas, laddove egli afferma che una macchina che non dovesse cadere sotto l'argomento gödeliano, "non sarebbe più una macchina nel senso della nostra discussione"[438]. Una simile affermazione è infatti banalmente vera, pur non aggiungendo niente alla forza dell'argomentazione, se con essa Lucas intende che l'inferenza da lui costruita, in quanto basata sui teoremi di Gödel, si applica solo alle 'macchine' in un'accezione precisa dell'espressione; ma, se letta come asserzione del fatto che una macchina è tale solo fin tanto che ad essa è applicabile l'argomento gödeliano, comporterebbe sì, se assunta, che non si possa ovviamente mostrare che una certa macchina non cade sotto l'inferenza, tuttavia non in un modo più stringente di una banale *petitio principii*.

Ci pare soprattutto, che la scelta espositiva di Lucas della propria lettura dei teoremi di incompletezza, la quale presenta, accanto ad un argomento che vorrebbe avere i requisiti del rigore, considerazioni non prive di ambiguità, non permetta di esercitare fino in fondo l'indubbio vantaggio offerto dalla possibilità, di cui si è detto in apertura di paragrafo, di formulare in termini rigorosi una tesi filosofica e la sua critica.

Diverso, è il caso della riformulazione dell'argomento gödeliano ad opera di Roger Penrose.

[437]Questo aspetto, ovviamente, è stato ampiamente evidenziato dalla vastissima letteratura critica relativa alla tesi di Lucas. Per una discussione di alcune delle obiezioni avanzate, si può vedere le parti del volume di Douglas Hofstadter [Hof79] dedicate alla questione, oltre ché considerare la bibliografia al proposito riprodotta alla pagina web dello stesso Lucas (`http://users.ox.ac.uk/~ jrlucas`).

[438]*Ibid.*, p. 58.

4.3 Il punto di vista di Penrose

Fin dal suo libro del 1989 dal titolo "La mente nuova dell'imperatore"[439], il fisico e matematico inglese Roger Penrose ha ripreso, cercato di rielaborare ed inserito, l'argomento di Lucas nel quadro di un complesso tentativo di dare vita a quello che si è rivelato essere, soprattutto nelle sue opere successive[440], il progetto di una "Scienza della Coscienza" volta ad indagare i misteri della mente umana.

All'interno di una ricerca che non ha, com'è ovvio, solo implicazioni di tipo logico, ma che è anche strettamente connessa a questioni di fisica (la meccanica quantistica, in particolare, viene utilizzata per proporre, in positivo, una possibile spiegazione della natura non computazionale della mente) e di filosofia (lo sfondo concettuale è sostanzialmente quello di una posizione di tipo platonista), l'argomento 'alla Lucas' di Penrose ha, come nel caso del filosofo inglese, il compito di offrire una confutazione del meccanicismo.

A differenza di Lucas, però, l'inferenza di Penrose (così come essa viene formulata, ad esempio, in [Pen96][441]) non è costruita a partire dai teoremi di incompletezza di Gödel, bensì sfrutta il Teorema dell'Arresto di Turing ed i risultati dell'analisi di Kleene della nozione di funzione ricorsiva parziale[442].

Il frutto principale (ai fini della trattazione di Penrose) di questa miscela, è la seguente generalizzazione del risultato di Gödel ad opera di Kleene[443]: non esiste alcun sistema formale corretto e completo per il predicato $\bigwedge y \overline{T}_1(x, x, y)$, non esiste cioè alcun sistema formale S che possieda una formula $\phi(x)$ tale che, per ogni $n \in \mathbf{N}$, $\mathsf{S} \vdash \phi(\overline{n})$ se e solo se $\bigwedge y \overline{T}_1(n, n, y)$.

Il Teorema di Forma Normale su cui la dimostrazione di questo fatto è basata in ultima analisi (e l'equivalenza tra funzioni μ-ricorsive e funzioni Turing-computabili), consente a Penrose di rielaborare il risultato in termini

[439] [Pen89].

[440] In particolare in [Pen96].

[441] Si veda [Pen96], §2.5.

[442] Poiché i risultati rispetto ai quali l'argomento di Penrose è debitore fanno ormai parte di un bagaglio di conoscenze logico-matematiche acquisito, ci limiteremo ad indicarli, qualora possa risultare utile, senza offrirne una trattazione esauriente. Un'analisi puntuale del *background* logico dell'argomento di Penrose, può essere trovato in [Fef95]. Anche per ragioni di spazio, si è omesso di dare conto del dibattito suscitato dai lavori di Penrose che è altrettanto ampio di quello seguito al contributo di Lucas. Per un resoconto, si rimanda, ad esempio, alla trattazione offerta in [Ant96] e [Ant00], ed alla bibliografia ivi citata, di parte delle obiezioni avanzate.

[443] Cfr. [Kle43].

di algoritmi (ossia macchine di Turing), e di problemi di terminazione delle computazioni.

Si ha, infatti, che esiste una macchina di Turing binaria $\mathbf{M}^*(x, y)$ che enumera tutte le macchine di Turing che agiscono su un solo input. Sia quindi:

$$\mathsf{M}_0(x),\ \mathsf{M}_1(x), \mathsf{M}_2(x), \ldots$$

l'ordinamento generato da M^*, rispetto al quale, cioè, vale, per ogni $i, n \in \mathbf{N}$:

$$\mathsf{M}^*(i, n) \simeq \mathsf{M}_i(n)$$

Supponiamo adesso, dice Penrose[444], che esista una procedura effettiva, ossia una macchina di Turing $\mathsf{M}_A(x, y)$ la quale, fornendo solo 'risposte' corrette, permette di stabilire, per ogni data coppia (m, n) (con $m, n \in \mathbf{N}$), che l'm-esima procedura dell'ordinamento, non termina sull'input n.

Adottando un simbolismo standard, si può esprimere questo fatto dicendo che, per ogni $m, n \in \mathbf{N}$, vale:

$$\mathsf{M}_m(n) \uparrow \Rightarrow \mathsf{M}_A(m, n) \downarrow \tag{4.1}$$

e, per l'ipotesi di correttezza:

$$\mathsf{M}_A(n, m) \downarrow \Rightarrow \mathsf{M}_n(m) \uparrow \tag{4.2}$$

In particolare, allora, per ogni $n \in \mathbf{N}$:

$$\mathsf{M}_A(n, n) \downarrow \Rightarrow \mathsf{M}_n(n) \uparrow \tag{4.3}$$

Essendo $\mathsf{M}_A(n, n)$ una procedura effettiva che si applica al solo input n, c'è un $k \in \mathbf{N}$ tale che:

$$\mathsf{M}_A(n, n) \simeq \mathsf{M}_k(n) \tag{4.4}$$

il che implica in particolare che $\mathsf{M}_A(n, n) \downarrow \Leftrightarrow \mathsf{M}_k(n) \downarrow$.

Ma da (4.3), per $n = k$, segue:

$$\mathsf{M}_A(k, k) \downarrow \Rightarrow \mathsf{M}_k(k) \uparrow \tag{4.5}$$

e da (4.4) e (4.5):

$$\mathsf{M}_k(k) \downarrow \Rightarrow \mathsf{M}_k(k) \uparrow \tag{4.6}$$

[444] [Pen96], p. 101. *NdA*: il riferimento (anche per la traduzione dei passi citati nel seguito) è all'edizione italiana, presso Rizzoli, del libro di Penrose, come risulta dalla voce bibliografica.

Dunque, $\mathsf{M}_k(k)\uparrow$.

Segue quindi da (4.4), che $\mathsf{M}_A(k,k)\uparrow$. Ciò significa che (sotto l'ipotesi (4.2)), esiste un $k \in \mathbf{N}$ tale che:

$$\mathbf{M}_k(k)\uparrow \ \& \ \mathbf{M}_A(k,k)\uparrow$$

che contraddice (4.1).

Si ha quindi una prova del fatto che la terminazione delle procedure di computo ad un solo input, non è un fatto accertabile mediante un algoritmo corretto.

Fin qui il teorema. Da qui, come nel caso di Lucas, si diparte l'argomento di Penrose che, seguendo il modello del primo, possiamo schematizzare come segue:

(A′) "nessun insieme di regole computazionali (come [M_A]) conoscibilmente valido, può essere sufficiente ad accertare che i computi non si arrestano";

(B′) "poiché a partire dalla conoscenza di [M_A] e dalla sua validità è possibile costruire un computo [$\mathsf{M}_k(k)$] che possiamo *vedere* che non si arresta mai";

(C′) "[da **(A′)** e **(B′)**] possiamo dedurre che [M_A] *non può* essere una formalizzazione dei procedimenti disponibili ai matematici per accertare che i computi non si arrestano".

Quindi, Penrose conclude:

(D′) "I matematici umani non stanno usando un algoritmo conoscibilmente valido per accertare la verità matematica."[445]

Osservando che, per come la procedura è stata definita, il risultato dell'applicazione di $\mathbf{M}_A(x,y)$ è una 'dimostrazione' di enunciati della forma 'La computazione dell'x-esima macchina di Turing dell'enumerazione non termina sull'input y', e che questi enunciati sono di tipo Π^0_1 (essendo, per mezzo dell'analisi di Kleene, equivalenti proprio ai predicati $\bigwedge y \overline{T}_1(x,x,y)$, con T_1 primitivo ricorsivo), Penrose ricava poi un raffinamento di (**D′**), sostituendolo con la seguente formulazione:

(D″) "I matematici umani non stanno usando un algoritmo conoscibilmente valido per accertare la verità dei Π^0_1-enunciati."[446]

[445] *Ibid.*, p. 103.
[446] *Ibid.*, p. 134.

4.3.1 La tesi meccanicista da Lucas a Penrose

L'elemento che distingue nel modo più evidente l'articolo di Lucas dalla trattazione di Penrose, parrebbe essere il fatto che, laddove il primo fonda la propria argomentazione anti-meccanicista sui teoremi di Gödel, Penrose sfrutta invece una versione del teorema dell'arresto di Turing. La ragione della scelta di Lucas si fondava, come si è visto nel passo retrospettivo citato nel §4.2, sul fatto che, a suo dire, i teoremi di incompletezza consentono di mostrare, al contrario di quanto si possa fare appellandosi al risultato di Turing, l'esistenza di un enunciato vero *e* indimostrabile, rendendo così più diretta la critica filosofica.

In realtà, fin tanto che, come si è cercato di far vedere, il risultato utilizzato da Penrose si base su una forma generalizzata dei teoremi di Gödel, questa differenza appare essere più superficiale che di sostanza, come dimostra anche il fatto che, dettaglio tecnico a parte, l'idea alla base dell'argomento di Penrose (come riportato in **(A′)**-**(D′)**/**(D″)**) è esattamente la stessa alla quale ricorre Lucas: c'è un enunciato che la mente umana è in grado di riconoscere come vero (e che esprime il fatto che una certa macchina di Turing non si arresta su un dato input) ma che, per mezzo dello stesso argomento, si dimostra non poter essere accertato come tale mediante una procedura effettiva di computo.

Il vero elemento discriminante, allora, è piuttosto un altro, e risiede in particolare nella conclusione che Lucas e Penrose traggono dalle rispettive analisi.

Come si diceva nel §4.2, il vantaggio che la trattazione di Lucas offre, consiste principalmente nel fatto di poter dare una veste più definita alla tesi meccanicista, la quale, in generale, recita:

(M) La mente umana è (o è equivalente a) una macchina.

Ciò che il filosofo inglese ritiene di poter confutare, mostrando, sulla base dell'argomento gödeliano, l'esistenza di un enunciato di tipo Π^0_1 che è vero e formalmente indimostrabile, è la seguente affermazione:

($\mathbf{M}_L$) Esiste un sistema formale S nel quale sono dimostrabili tutti gli enunciati (matematici) riconoscibili come veri dalla mente umana.[447]

[447] La dicitura **($\mathbf{M}_L$)** sta ovviamente ad indicare la variante di **(M)** che è al centro della critica di Lucas. Questa notazione viene estesa, in modo ovvio, al caso di Penrose.

La mente è meccanica, in questo senso, perché tutti i metodi di cui essa fa uso per fondare un giudizio sulla verità di una proposizione matematica, sono di natura formale (o sono formalizzabili).

La critica di Lucas, come si è detto, assume in modo indebito che la mente umana sia sempre in grado, dato un qualunque sistema d'assiomi, di discriminare, con certezza assoluta, se esso sia consistente o meno. Poiché l'assunzione è cruciale per quanto Lucas intende sostenere (dal momento che se un sistema formale è inconsistente, l'enunciato indecidibile di Gödel è falso e dimostrabile), quest'assunzione rende l'argomentazione implausibile, o comunque discutibile.

Penrose si muove nella stessa direzione, ma è sufficientemente accorto da rendere la propria critica immune, per quanto si è visto fin qui, dall'obiezione che può essere mossa all'inferenza di Lucas. Ciò che la conclusione $(\mathbf{D}')$ confuta è infatti la seguente affermazione:

$(\mathbf{M}'_P)$ C'è un sistema formale S, che è *conoscibilmente corretto* e nel quale sono dimostrabili tutti gli enunciati (matematici) riconoscibili come veri dalla mente umana.

Sulla base di $(\mathbf{D}'')$, considerando cioè la forma logica dell'enunciato vero e indimostrabile, Penrose ritiene di poter considerare come contraddittoria, in realtà, già la seguente istanza della precedente:

$(\mathbf{M}''_P)$ C'è un sistema formale S conoscibilmente corretto e nel quale sono dimostrabili tutti i Π^0_1-enunciati riconoscibili come veri dalla mente umana.

Ciò ci consente un ulteriore passo avanti nel precisare i termini della tesi: se si restringe l'attenzione agli enunciati di tipo Π^0_1, l'assunzione di correttezza generale può essere allora sostituita con quella di Π^0_1-validità la quale equivale, come è noto, alla consistenza del sistema d'assiomi[448].

Si ricaverebbe, quindi, la confutazione della seguente affermazione:

$(\mathbf{M}^*)$ Esiste un sistema formale S che sappiamo essere consistente ed è tale che l'insieme dei Π^0_1-enunciati dimostrabili in esso coincide con la

[448] L'argomento è il seguente: se S è consistente e $\mathsf{S} \vdash \forall x R(x)$ con R ricorsiva, allora $\mathsf{S} \vdash R(\bar{n})$, per ogni $n \in \mathbf{N}$. Se per qualche $m \in \mathbf{N}$, valesse $\models \neg R(m)$, essendo S Σ^0_1-completo (come risulta dimostrabile dal momento che S è consistente e sufficientemente potente per ipotesi), si avrebbe $\mathsf{S} \vdash \exists x \neg R(x)$, contro la consistenza di S. Il viceversa (se S è Π^0_1-valido allora S consistente), è, naturalmente, banale.

collezione degli enunciati, della stessa forma, riconoscibili come veri dalla mente umana (quest'ultima classe, quindi, è ricorsivamente enumerabile).

Il percorso di analisi, prima ancora che di critica, della tesi meccanicista potrebbe essere considerato concluso, avendo noi apparentemente sostituito i termini più problematici contenuti nella tesi **(M)**, con i concetti chiari, di natura logica, di **(M***)[449].

La differenza principale tra la disamina di Lucas e quella di Penrose, consiste allora nella diversa considerazione riservata alle ipotesi sulle quali poggia il teorema da essi chiamato in causa. In particolare, il costo, per Penrose, di un maggior grado di correttezza nell'argomento proposto, consiste nelle varianti della tesi **(M***) che, come si evince dalla sua formulazione, rimangono a rigor di logica praticabili, ossia:

1. un'*ipotesi m-scettica*[450]: i metodi matematici di prova sono tutti racchiusi in un algoritmo non valido;

2. un'*ipotesi m-agnostica*: i metodi matematici di prova sono tutti racchiusi in un algoritmo corretto ma non conoscibilimente tale.

Già a questo livello, dunque, il ricorso per finalità 'filosofiche' dei risultati di indecidibilità denuncia un limite preciso.

Ciò spiega, in particolare, l'ampia parte (in [Pen96], dal §3.2 in poi e per tutto il cap. 3), nella quale Penrose cerca di argomentare in sfavore tanto della possibilità che alla base della conoscenza matematica vi possa essere un algoritmo non valido (§3.4), quanto che vi sia un algoritmo inconoscibile (§3.5).

È tra le pieghe di questa discussione informale, che, quindi, in quanto tale, indebolisce l'argomento proposto (per quanto implausibili si vogliano considerare le alternative rimaste in piedi), che si fa strada una seconda versione dell'infererenza di Penrose la quale, aspirando a recuperare su un piano tutto rigoroso, ossia logico-deduttivo, la (presunta) generalità dell'argomento di Lucas, vorrebbe ovviare ai limiti della formulazione più 'classica' (se correttamente intesa) dell'argomento gödeliano.

---

[449]Come chiariremo nel §4.4.1, anche una simile formulazione (Gödel *docet*), mantiene notevoli margini di ambiguità, non irrilevanti ai fini delle conclusioni che Penrose intende raggiungere.

[450]La dicitura adottata intende segnalare il fatto che si tratta della combinazione di una posizione di tipo meccanicistico con un'assunzione scettica (rispettivamente, agnostica).

4.4 Il “Nuovo Argomento Centrale”

L’impressione di alcuni commentatori che quelle parti del terzo capitolo di [Pen96], dedicate alla discussione delle conclusioni della versione dell’argomento basato sul teorema dell’arresto (o meglio su una versione generalizzata del teorema di Gödel), contenessero in realtà un vero e proprio tentativo di rafforzare l’inferenza originaria, è stata confermata dallo stesso Penrose nella risposta, in [P$^+$95], alle obiezioni mosse al proprio lavoro.

Nella nuova veste, l’argomento, che ‘abbandona’ Turing e torna a fare un uso diretto, diciamo così, dei teoremi di incompletezza (si parla quindi di sistemi formali, e non di algoritmi, che “incorporano tutti i metodi matematici di prova umanamente accessibili”), intende spiegare come potrebbe ragionare un matematico umano di fronte ad un qualsiasi sistema formale F, al quale si possa applicare la costruzione gödeliana. La formulazione più chiara di quello che può essere indicato, tenendo presente i riferimenti contenuti nella letteratura, come il “Nuovo Argomento Centrale” di Penrose, è contenuta in [P$^+$95]:

> Sebbene non so di essere F, necessariamente concludo che se lo fossi, il sistema F sarebbe corretto e, in più, sarebbe corretto anche F', dove F' è ottenuto da F per l’aggiunta dell’asserzione “Io sono F”. Quindi, percepisco che dall’assunzione che sono F, segue la verità dell’enunciato di Gödel $G(\mathsf{F}')$ e, inoltre, che esso non sarebbe un teorema di F'. Ma allora ho appena concluso che “se fossi F, $G(\mathsf{F}')$ sarebbe vera”, e conclusioni di questo genere sono proprio quelle che si suppone debba raggiungere F'. Poichè sono in grado di percepire qualcosa che va al di là delle capacità di F', deduco che, dopo tutto, non posso essere F. In più, questo vale di ogni sistema (Gödelizzabile) che prenda il posto di F.[451]

Dal momento che lo stesso Penrose riconosce come sia in questa versione rafforzata che l’argomento presentato in [Pen96] permette di concludere, a suo avviso, che “non possiamo essere identici ad alcun sistema F (Gödelizzabile) verosimile, che si sia in grado di giungere ad una credenza empirica in esso oppure no!”, occorre soffermarci brevemente sulle novità che il passo contiene rispetto alla formulazione precedente della posizione dello scienziato inglese.

L’idea di Penrose, è che per mezzo di un simile ragionamento, un matematico umano, confrontando le proprie “convinzioni matematiche incontrovertibili” (che, in [Pen96], sono anche indicate da Penrose come le $\star$-affermazioni),

[451] [P$^+$95], §3.2. *NdA*: i riferimenti a [P$^+$95] sono privi del numero di pagina, riferendosi, come risulta dalla voce bibliografica, alla versione *on line* del contributo.

con le conclusioni (i teoremi) di F, che ad esse corrisponderebbero se egli 'fosse' F (un *qualsiasi* sistema al quale si possano applicare i teoremi di incompletezza), giunga alla conclusione che egli "non può essere identico", in fin dei conti, ad un tale F.

Il ragionamento, prende le mosse, come in ogni argomento per assurdo che si rispetti, dall'ipotesi "Io sono F", sulla quale Penrose non dice altro, se non che essa "è meramente un'abbreviazione per 'F incapsula tutti i metodi matematici di dimostrazione accessibili all'uomo'"[452]. Intesa come definizione, la spiegazione di Penrose non è di grande aiuto, sostituendo un'espressione ambigua con un'altra che, per altri versi, lo è altrettanto. Tenendo presente anche il tipo di contraddizione alla quale i teoremi di Gödel consentono di giungere (l'esistenza di una "convinzione matematica incontrovertibile" che è anche formalmente indimostrabile), si può plausibilmente assumere che dall'ipotesi "Io sono F" discenda il fatto che la collezione delle "convinzioni incontrovertibili" del matematico umano coincida con l'insieme dei teoremi del sistema in questione.

Indicando allora con $\mathbf{M}^\star$ la prima classe e con $\mathbf{T}_\mathsf{F}$ il secondo insieme (ed assumendo di indicare con $SONO(\mathsf{F})$ l'ipotesi di partenza), si ha cioè che:

$$SONO(\mathsf{F}) \Rightarrow \mathbf{M}^\star = \mathbf{T}_\mathsf{F} \tag{4.7}$$

vale[453].

Come si evince chiaramente dall'*incipit* del passo precedente, Penrose ritiene che dall'ipotesi che $SONO(\mathsf{F})$ valga, segue anche la correttezza del sistema F e, in più, del sistema F' ottenuto da questo per aggiunta dell'enunciato che esprime proprio quell'ipotesi (ovverosia, $\mathsf{F}' := \mathsf{F} \cup \{Sono_\mathsf{F}\}$, dove $Sono_\mathsf{F}$ è un qualche enunciato aritmetico che è vero, sotto l'interpretazione intesa dei termini, se e solo se $SONO(\mathsf{F})$ vale).

Ciò significa quindi che, assumendo di indicare con $VAL(\mathsf{F})$ la validità del sistema F, si ha anche:

$$SONO(\mathsf{F}) \Rightarrow VAL(\mathsf{F})\ \&\ VAL(\mathsf{F}') \tag{4.8}$$

[452] *Ibid.*, §3.2.

[453] Scrivendo $\mathbf{M}^\star = \mathbf{T}_\mathsf{F}$, stiamo ovviamente facendo astrazione dal fatto che gli enunciati che sono verità matematiche incontrovertibili, e che quindi appartengono a $\mathbf{M}^\star$, sono scritti in un linguaggio che è diverso da quello al quale appartengono gli enunciati di $\mathbf{T}_\mathsf{F}$. Ciò significa che quell'uguaglianza presuppone una qualche forma di traduzione dell'un linguaggio nell'altro, e viceversa.

Se si tiene conto di quanto si è detto nel paragrafo precedente a proposito dell'ipotesi di correttezza in relazione alla forma logica dell'enunciato indecidibile, e assumendo di indicare con $CON(\mathsf{F})$ la consistenza del sistema F (a cui la restrizione della validità alla classe dei Π^0_1-enunciati è, come si è detto, equivalente), si può sostituire la (4.8) con la più debole:

$$SONO(\mathsf{F}) \Rightarrow CON(\mathsf{F}) \;\&\; CON(\mathsf{F}') \tag{4.9}$$

Su (4.9) in particolare, Penrose dice in [Pen96]:

> [S]e accetto ["Io sono F"], allora accetto anche tutti gli enunciati Π^0_1 generati dal sistema [$\mathsf{F} \cup \{$"Io sono F"$\}$]. Quindi devo accettare che [$\mathsf{F} \cup \{$"Io sono F"$\}$] sia valido[454] sulla base di quest'assunzione.[455]

Accettare un certo formalismo, sembra voler dire Penrose, ed accettare in particolare che esso 'contenga' tutti i metodi matematici di prova accessibili ad un matematico umano, significa avere 'fiducia' in esso, ossia credere nella sua validità. Questa fiducia, può essere vista quindi come il riflesso, indotto dall'ipotesi che "Io sono F" valga, della convinzione del matematico negli strumenti inferenziali di cui egli fa uso.

L'inferenza (4.9) sarebbe perciò il prodotto del seguente ragionamento: se sono F, F è consistente in quanto io lo sono (i miei strumenti conoscitivi sono corretti per gli enunciati Π^0_1); dunque, se $SONO(\mathsf{F})$ è vero, nessuna formula falsa (di tipo Π^0_1) può essere dimostrata in F a partire da questo fatto; segue che $\mathsf{F}' := \mathsf{F} \cup \{Sono_{\mathsf{F}}\}$ è consistente.

Di capitale importanza, come forse risulterà in modo inequivocabile tra un momento, è anche quel passaggio della citazione di Penrose nel quale egli dice che, avendo concluso che se le ipotesi fatte valessero allora l'enunciato indecidibile in F' ($G_{\mathsf{F}'}$) sarebbe vero, si ottiene la contraddizione fondamentale dall'osservazione ulteriore che "conclusioni di questo genere sono proprio quelle che si suppone debba raggiungere F'". La spiegazione di quanto Penrose afferma, non è difficile se si pensa alle definizioni date: se "Io sono F" vale, potendo concludere sulla base dei teoremi di incompletezza che, se F' è corretto rispetto ai Π^0_1-enunciati, allora $G_{\mathsf{F}'}$ è indimostrabile in F' ed è, per giunta, un enunciato vero (cioè è una convinzione matematica incontrovertibile), concludo da ciò, proprio perché "Io sono F" vale, e quindi F contiene tutti i metodi matematici umanamente accessibili, che quel fatto (la verità

[454] *NdT*: nel senso ristretto di Π^0_1-valido.
[455] [Pen96], p. 212.

di G_{F} sotto l'ipotesi "Io sono F"), dovrebbe anche essere dimostrabile in F, e, in quanto per definizione ne è un'estensione, in F' di conseguenza.

Se adesso pensiamo a ciò sulla base della notazione adottata e delle considerazioni fatte, il passaggio di Penrose può essere reso nel modo seguente: dal primo teorema di Gödel segue che c'è un enunciato indimostrabile in F', il quale è anche incontrovertibilmente vero se F' è consistente (in simboli: $CON(\mathsf{F}') \Rightarrow (G_{\mathsf{F}'} \notin \mathbf{T}_{\mathsf{F}'})\ \&\ (G_{\mathsf{F}'} \in \mathbf{M}^\star)$); per (4.9) si ha dunque in particolare che $SONO(\mathsf{F}) \Rightarrow (G_{\mathsf{F}'} \in \mathbf{M}^\star)$ vale; da ciò, concludo però anche che l'enunciato che esprime questo stesso fatto (ossia $Sono_{\mathsf{F}} \rightarrow G_{\mathsf{F}}$) sarebbe una conseguenza matematica incontrovertibile dell'ipotesi da cui sono partito, che $SONO(\mathsf{F})$ valga (passaggio che può essere reso scrivendo, sotto l'ipotesi $SONO(\mathsf{F})$, $(Sono_{\mathsf{F}} \rightarrow G_{\mathsf{F}'}) \in \mathbf{M}^\star$); dal ché, sulla base di (4.7), ossia del fatto che, assunto $SONO(\mathsf{F})$, verità incontrovertibile e dimostrabilità in F si equivalgono, si avrebbe che da $SONO(\mathsf{F})$ segue $(Sono_{\mathsf{F}} \rightarrow G_{\mathsf{F}'}) \in \mathbf{T}_{\mathsf{F}}$[456].

Mettendo insieme quanto detto finora, ci si accorge che è possibile restituire il ragionamento di Penrose nella forma della derivazione di una contraddizione dalla tesi meccanicista, rappresentata dall'ipotesi che $SONO(\mathsf{F})$ valga (assieme alle relative conseguenze). Quest'inferenza, avrebbe la seguente struttura:

1. $SONO(\mathsf{F}) \Rightarrow (CON(\mathsf{F})\ \&\ CON(\mathsf{F}'))$ (4.9)
2. $CON(\mathsf{F}') \Rightarrow (G_{\mathsf{F}'} \notin \mathbf{T}_{\mathsf{F}'}\ \&\ G_{\mathsf{F}'} \in \mathbf{M}^\star)$ (I teor. di Gödel)
3. $SONO(\mathsf{F}) \Rightarrow (G_{\mathsf{F}'} \in \mathbf{M}^\star)$ (segue da 1 e 2)
4. $SONO(\mathsf{F}) \Rightarrow ((Sono_{\mathsf{F}} \rightarrow G_{\mathsf{F}'}) \in \mathbf{T}_F)$ (3 è incontrovertibile, (4.7) vale)
5. $SONO(\mathsf{F}) \Rightarrow (G_{\mathsf{F}'} \in \mathbf{T}_{\mathsf{F}'})$ (4 e $\mathsf{F}' := \mathsf{F} \cup \{Sono_{\mathsf{F}}\}$)
6. $SONO(\mathsf{F}) \Rightarrow (G_{\mathsf{F}'} \notin \mathbf{T}_{\mathsf{F}'})$ (1 e 2)
7. $\overline{SONO}(\mathsf{F})$ (5 e 6)

Una tale resa possiede, a nostro avviso, due vantaggi. Consente, da un lato, per mezzo della contraddizione, di sostituire la formula conclusiva usuale di questo tipo di argomenti (poiché si 'intuisce' qualcosa che 'va al di là' delle

[456] La spiegazione si riferisce, nella formalizzazione dell'argomento presentata sotto, al passaggio dal punto 3 al punto 4.

capacità del sistema formale, si conclude che non siamo 'identici' ad esso), che può essere considerata non interamente soddisfacente.

È possibile poi su questa base, dall'altro lato, individuare in modo immediato il passo che si intende compiere, con l'argomento in questa versione rispetto alla precedente, verso una maggiore generalità: subordinando la consistenza del sistema formale F preso in esame (e della sua estensione F') all'ipotesi che esso 'contenga' tutti i metodi matematici di prova umanamente accessibili, Penrose può evitare di considerare un F che sia conoscibilmente consistente (come nel caso della conclusione dell'argomento nella prima versione), sostituendo a tale assunzione (via $SONO(\mathsf{F})$) l'*ipotesi* che esso sia tale. In altre parole, tutto ciò che occorre concedere a Penrose perché il suo argomento si applichi, è il fatto che un matematico umano in quelle circostanze (assumendo cioè $SONO(\mathsf{F})$), sia per ciò stesso condotto a concludere, sulla base di una sorta di convinzione innata nei propri strumenti conoscitivi, la correttezza formale del sistema. Una concessione (quella della validità dell'*inferenza* (4.9)), che anche il 'meccanicista scettico' o quello 'agnostico' potrebbero essere disposti a concedere, pur sostenendo che i metodi matematici di prova accessibili all'uomo non sono consistenti o, rispettivamente, che non sono conoscibilmente tali.

Ci pare perciò opportuno ritenere, almeno in prima istanza, l'argomento proposto come il correlato rigoroso del Nuovo Argomento Centrale di Penrose. Si ha così, in fondo, che il percorso che si è fin qui condotto ha portato non solo ad una precisazione di certi termini ambigui dell'istanza meccanicista, ma anche ad una formulazione apparentemente rigorosa, e, in quanto tale, passibile di un'analisi interamente formale, della critica ad essa.

Quest'ultimo compito, in particolare, è quanto ci ripromettiamo di portare a termine nei prossimi due paragrafi.

4.4.1 Le convinzioni matematiche incontrovertibili: ancora sulla *Gibbs Lecture* di Gödel

Prima ancora che affrontare la questione della correttezza logica e della plausibilità metamatematica del nuovo argomento di Penrose, vogliamo fare qualche breve considerazione sul punto raggiunto, ispirata alle riflessioni di Gödel sulle implicazioni 'filosofiche' dei teoremi di incompletezza.

Si sarà certo notata una somiglianza tra l'uso che Penrose e Lucas fanno di questi risultati, e le osservazioni di Gödel relative al rapporto tra menti e macchine nella *Gibbs Lecture* prima, e nel volume di Wang del 1974 poi[457].

Ci si può convincere di ciò, riconsiderando quel passo della conferenza del 1951 che recita:

> È *questo* teorema che rende l'incompletabilità della matematica particolarmente evidente. Perché, *rende impossibile che chiunque mettendo insieme un sistema d'assiomi e regole ben definito, faccia la seguente affermazione su esso in modo consistente: Percepisco* (*con certezza matematica*) *gli assiomi e le regole in questione come corretti e credo inoltre che essi contengano l'intera matematica.*[458]

Se si tralascia per il momento il fatto che Gödel concentri la propria attenzione sul secondo teorema di incompletezza piuttosto che sul primo, non si potrà negare che ci sono similitudini sostanziali tra la posizione di Penrose e l'analisi di Gödel. Non fosse altro che per il fatto che le osservazioni di Gödel si riferiscono alla "matematica propriamente detta", ossia la classe delle proposizioni "percepite con certezza matematica", e che tale espressione richiama da vicino la collezione delle "verità matematiche incontrovertibili" dalla quale prende le mosse Penrose.

Ad ulteriore conferma di quanto detto, si può riprendere la notazione adottata in precedenza e ricostruire, su questa base, l'argomento che Gödel fornisce per mostrare come ricavare la contraddizione del passo succitato dal secondo teorema di incompletezza.

Come si ricorderà, vale infatti quanto segue:

1. $VAL(\mathsf{F}) \in \mathbf{M}^\star$ ("Percepisco che F è corretto")
2. $CON(\mathsf{F}) \in \mathbf{M}^\star$ (1 e $(VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F})) \in \mathbf{M}^\star$)
3. $Con_\mathsf{F} \notin \mathbf{T}_\mathsf{F}$ (II teor. di Gödel)
4. $\mathbf{M}^\star \neq \mathbf{T}_\mathsf{F}$

Se si assume, come si è fatto nel paragrafo precedente, che, in assenza di una definizione precisa di cosa significhi l'espressione "Io sono F", valga quantomento $SONO(\mathsf{F}) \Rightarrow \mathbf{M}^\star = \mathbf{T}_\mathsf{F}$, la conclusione ottenuta, sarebbe sufficiente allo scopo di ricavare, da essa, come nel caso del nuovo argomento di Penrose, $\overline{SONO}(\mathsf{F})$.

[457] Si vedano, della prima parte del lavoro, i §§2.4 e 2.4.1.

[458] [Göd51], p. 309.

Non sarà sfuggita però, a questo punto, una differenza essenziale tra Gödel e Penrose: il primo, infatti, presenta il proprio argomento come *auto*-confutazione, che conduce quindi ad un'incompatibilità logica tra l'*affermazione* dell'incontrovertibilità della correttezza del sistema F e la *convinzione* che F contenga tutta la matematica propriamente detta. In sostanza, ciò a cui conduce l'argomento presentato da Gödel, è il fatto che le due modalità presenti nella tesi del passo riportato in precedenza (il *percepire con certezza matematica* la correttezza di F, e il *credere* che esso contenga tutta la matematica), devono essere tenute necessariamente distinte.

Ciò potrebbe spiegare l'insistenza da parte di Gödel sul secondo teorema di incompletezza piuttosto che sul primo. Per quanto, infatti, non sarebbe difficile ricostruire lo stesso argomento sfruttando l'enunciato che "dice di se stesso' di essere indecidibile in F, si dovrebbe comunque partire dall'assunzione sulla validità del sistema per essere in grado di ottenere la contraddizione. Ben più diretto, quindi, appare l'argomento così come esso viene concepito da Gödel.

Sotto questo aspetto, la conclusione di Gödel non è molto diversa da quella a cui giunge Penrose lungo le linee più tradizionali dell'argomento gödeliano. Tuttavia, rispetto a ciò a cui quest'ultimo aspira con la nuova formulazione, si tratta di un passo indietro, dal momento che l'intera argomentazione dipende dall'ipotesi sulla correttezza del sistema, la quale, dopotutto, potrebbe essere contestata secondo le linee che abbiamo già indicato. Questo spiega, ovviamente, la semplicità dell'argomento di Gödel in contrapposizione alla formulazione più complessa che abbiamo ricavato dalle riflessioni dello scienziato inglese.

Ciò non toglie che quanto Gödel dice a proposito delle proprie conclusioni, non possa essere di un qualche rilievo anche in merito alle circostanze presenti.

Si ricorderà infatti che Gödel ritiene che si debba procedere con cautela nel ricavare, dall'argomento proposto poc'anzi, che "nessun ben definito sistema di assiomi corretti può contenere tutta la matematica propriamente detta", o in termini più vicini a quelli del presente capitolo, che la collezione $\mathbf{M}^{\star}$ non può essere ricorsivamente enumerabile. A tale proposito, Gödel propone la distinzione tra una "matematica in senso oggettivo" (la classe delle proposizioni matematiche vere) e la "matematica in senso soggettivo" (la classe delle proposizioni dimostrabili).

Non si tratta di una mossa casuale: dopotutto, infatti, si potrebbe obiettare che l'argomentazione precedente si basa su un concetto, quello di "percezione con certezza matematica" (o, equivalentemente, di "verità matematica incontrovertibile") che non è affatto autoevidente, ma che, al contrario, risulta essere intrinsecamente vago. Occorre quindi, in primo luogo, dichiarare che cosa si intenda con quell'espressione, ovverosia rispondere al quesito: *Che cos'è* $\mathbf{M}^\star$?

Gödel distingue due possibili interpretazioni abbiamo detto: (i) $\mathbf{M}^\star$ è la collezione degli enunciati matematici veri, (ii) $\mathbf{M}^\star$ contiene tutti e soli gli enunciati dimostrabili. Rispetto alle alternative, gli esisti dell'analisi sono diversi: nel primo caso, ci si trova certamente, con l'argomento surriportato, di fronte ad una prova del fatto che la classe $\mathbf{M}^\star$ non può essere generata mediante un procedimento effettivo; ma ciò non è affatto vero nel secondo caso, nel quale, mostrando che la consistenza (e quindi, *a fortiori*, la validità) di un sistema formale non potrebbe far parte delle verità matematiche incontrovertibili qualora queste costituissero un insieme ricorsivamente enumerabile di proposizioni, si sarebbe fornito soltanto un argomento dal quale discende esclusivamente il fatto che quell'enunciato non potrebbe essere esso stesso percepito con certezza matematica. Si sarebbe quindi mostrato l'esistenza di enunciati matematici indecidibili in senso assoluto.

L'analisi logica della tesi meccanicista non è compiuta, si potrebbe perciò dire, fin tanto che il risultato di essa continua a chiamare in causa un concetto, quello di verità matematica, che si presta a molteplici interpretazioni, dalle quali, per di più, dipende l'esito della disamina.

La conclusione disgiuntiva alla quale Gödel perviene[459], e che è il frutto di quest'analisi, è, però, solo in un senso l'indice di un esito incerto: lo è sicuramente se si considera che essa è tutto quanto può essere detto sulla questione (sul rapporto cioè tra menti e macchine) da un punto di vista puramente matematico; non lo è, tuttavia, se, accanto ad essa, si considerano quelle ragioni (delle quali si è dato conto nel §2.4.2) che, a detta di Gödel, conducono inevitabilmente ad una posizione opposta alla "filosofia materialista". In ultima analisi, quella stessa conclusione porta Gödel ad argomentare, ovviamente su un piano squisitamente filosofico, a favore del platonismo come unica filosofia della matematica realmente praticabile.

Curiosamente, anche Penrose viene spinto dalle proprie conclusioni verso una forma di realismo nei confronti degli enti matematici[460]. Ma, ancor prima di ciò, in contrasto con quanto Gödel riteneva fosse lecito fare sulla base dei teoremi di incompletezza, egli è anche convinto di aver fornito, con la nuo-

[459] Si veda §2.4.

[460] Cfr. [Pen96], §8.7.

va versione dell'argomento gödeliano, una confutazione rigorosa dell'istanza meccanicista.

Se questa convinzione sia legittima da un punto di vista logico, è quanto vogliamo misurare con il prossimo paragrafo, riservandoci di tornare sugli aspetti concettuali del problema nel paragrafo immediatamente successivo, che chiude il presente capitolo.

4.4.2 Aspetti logici e metamatematici: le critiche a Penrose di Pudlák e Lindström

Nel passare all'analisi dei fondamenti logici del nuovo argomento di Penrose, procederemo distinguendo, apparentemente, la questione della correttezza dell'inferenza ideata dallo scienziato inglese, da alcune considerazioni di carattere più specificamente metamatematico. Diciamo apparentemente, perché alla base di entrambi i filoni di riflessione vi è la discussione di quella che appare essere, già dall'analisi da noi condotta nel §4.4, l'assunzione cruciale di Penrose che abbiamo indicato come (4.9), ossia l'inferenza:

$$SONO(\mathsf{F}) \Rightarrow CON(\mathsf{F})\ \&\ CON(\mathsf{F}')$$

Considerando anche la giustificazione che Penrose sembra dare per essa, con questa assunzione si vuole dare conto di una presunta capacità del matematico umano di 'andare oltre' un formalismo dato, 'intuendo', sotto l'ipotesi che esso contenga tutti i metodi matematici accessibili, una 'verità' (la consistenza del medesimo), che trascende le capacità deduttive del sistema formale. È dunque già qui che si pone il fondamento perché, attraverso l'argomento gödeliano, si possano raggiungere le finalità che ci si propongono.

Ed è dunque prevedibilmente in relazione a quest'inferenza che occorre misurare la fondatezza di queste stesse finalità.

Una confutazione dell'argomento di Penrose

La formulazione dell'assunzione (4.9) e, conseguentemente, dell'argomento di Penrose, può essere ampiamente criticata. Come si è già fatto notare, si può considerare innanzi tutto insoddisfacente il fatto che con essa ci si riferisca ad un'ipotesi, "Io sono F", della quale, (i) non si dice alcunché di esauriente, a parte darne una perifrasi, "F incapsula tutti i metodi di prova matematici accessibili all'uomo", che è quantomeno altrettanto vaga ed ambigua, (ii) si sostiene per contro, come fa Penrose, che essa esprime un enunciato che non solo ha una natura prettamente matematica, ma che è, oltretutto, esprimibile nel linguaggio formale $\mathcal{L}_{\mathsf{F}}$ sul quale F è basato (un'affermazione verso la quale ci 'obbliga' la stessa assunzione (4.9) nel riferirsi al sistema F' ottenuto da F proprio mediante l'enunciato corrispondente a $SONO(\mathsf{F})$).

Si consideri allora, in sostituzione di (4.9), la seguente inferenza:

$$VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F}^+) \tag{4.10}$$

dove adesso $\mathsf{F}^+ := \mathsf{F} \cup \{Val_{\mathsf{F}}\}$.

Come (4.9), e forse anche meglio, (4.10) possiede una chiara giustificazione intuitiva: se si assume che un certo sistema d'assiomi F sia valido, si ha, in particolare, che l'enunciato che esprime questo fatto, Val_{F}, è vero e che, dunque, il sistema che si ottiene da F per aggiunta di quell'enunciato è in particolare consistente.

Per di più, è possibile riformulare l'argomento di Penrose in modo che esso dipenda da questa assunzione, piuttosto che da (4.9).

A tal fine è sufficiente assumere, modificando così le ipotesi (4.7) e (4.9) del §4.4 in ragione di (4.10), che valga anche:

$$SONO(\mathsf{F}) \Rightarrow \mathbf{M}^\star = \mathbf{T}_{\mathsf{F}} \;\&\; VAL(\mathsf{F}) \tag{4.11}$$

Si ha così la seguente versione modificata del Nuovo Argomento Centrale[461]:

1. $VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F}^+)$ (4.10)
2. $CON(\mathsf{F}^+) \Rightarrow (G_{\mathsf{F}^+} \in \mathbf{M}^\star \;\&\; G_{\mathsf{F}^+} \notin \mathbf{T}_{\mathsf{F}^+})$ (I Teor. di Gödel)
3. $VAL(\mathsf{F}) \Rightarrow (G_{\mathsf{F}^+} \in \mathbf{M}^\star)$
4. $VAL(\mathsf{F}) \Rightarrow (G_{\mathsf{F}^+} \notin \mathbf{T}_{\mathsf{F}^+})$ (3 e 4 seguono da 1 e 2)
5. $\mathbf{M}^\star = \mathbf{T}_{\mathsf{F}} \Rightarrow (Val_{\mathsf{F}} \rightarrow G_{\mathsf{F}^+} \in \mathbf{T}_{\mathsf{F}})$ (1-3)[462]

[461] Questa versione, adattamento alla notazione cambiamenti di secondaria importanza a parte, è dovuta a Per Lindström (cfr. [Lin01], p. 244).

[462] La giustificazione di questo passo corrisponde, nella versione del nuovo argomento di Penrose del §4.4, a quanto detto in relazione al punto 4: poiché sulla base di 1-3 siamo disposti a vedere in 3 una conclusione incontrovertibile (e quindi un elemento di $\mathbf{M}^\star$), concludiamo, dall'ipotesi $\mathbf{M}^\star = \mathbf{T}_{\mathsf{F}}$, che questa stessa conclusione, sottoforma di enunciato, deve essere un teorema di F.

6. $\mathbf{M}^\star = \mathbf{T}_F \Rightarrow (G_{\mathsf{F}^+} \in \mathbf{T}_{\mathsf{F}^+})$ (5 e def. di F^+)

7. $\overline{(VAL(\mathsf{F})\ \&\ \mathbf{M}^\star = \mathbf{T}_\mathsf{F})}$ (4 e 6)

8. $\overline{SONO}(\mathsf{F})$ (4.11)

Stando così le cose, si potrebbero sì evitare le obiezioni derivanti dal fatto di far dipendere l'intero argomento da un'assunzione, (4.9), la quale appare implausibile, prim'ancora che per ragioni di merito, perché presume la possibilità di formalizzare $SONO(\mathsf{F})$ anche in assenza di una definizione adeguata di ciò che tale proposizione significa esattamente, ma occorre pur sempre definire $VAL(\mathsf{F})$ in modo tale che si abbia che:

(i) $VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F}^+)$, è un'inferenza valida;

(ii) c'è una formula Val_F del linguaggio di F, che la esprime[463].

Come si è avuto modo di apprezzare nel §4.4, Penrose sembra indirizzato ad identificare, con $VAL(\mathsf{F})$ la validità per enunciati di tipo Π^0_1, in virtù della forma logica dell'enunciato indecidibile di Gödel.

Tuttavia, si avrebbe in questo caso[464]:

$$VAL(\mathsf{F}) := (\Pi^0_1 - VAL(\mathsf{F})) \Leftrightarrow CON(\mathsf{F})$$

Ed è anche noto, perché deriva proprio dai teoremi di Gödel, che l'ipotesi (4.10) non è valida in generale. Infatti, da una metà del secondo teorema di incompletezza, si ha che[465]:

$$CON(\mathsf{F}) \Rightarrow CON(\mathsf{F} + \neg Con_F)$$

Come conseguenza della formalizzazione della sintassi, è però possibile dimostrare che il sistema $\mathsf{F}^* := \mathsf{F} + \neg Con_\mathsf{F} + Con_{\mathsf{F}+\neg Con_\mathsf{F}}$ non è consistente[466].

---

[463] Ciò significa che Val_F è vera, sotto l'interpretazione intesa, se e solo se $VAL(\mathsf{F})$ vale.

[464] Si veda quanto detto alla nota 448.

[465] Per semplificare la notazione, qui e nel seguito scriviamo $\mathsf{F} + \phi$ in luogo di $\mathsf{F} \cup \{\phi\}$.

[466] Qui, e per tutto il paragrafo, assumiamo di considerare solo sistemi formali 'sufficientemente potenti', come di solito si esprime una delle ipotesi cruciali dei teoremi di incompletezza, ossia in grado di formalizzare adeguatamente la sintassi di un linguaggio formale. Tipicamente, è sufficiente considerare sistemi che siano estensioni della teoria $\mathsf{\Sigma}^0_1 - \mathsf{IA}$, la quale è ottenuta dall'aritmetica di Peano PA restringendo lo schema d'induzione alle sole formule di tipo Σ^0_1. Per una questione di brevità, si presupporranno spesso le proprietà fondamentali derivanti dall'aritmetizzazione della sintassi e dalla formalizzazione della medesima, per quanto si cercherà di indicare con precisione i riferimenti ai risultati chiamati in causa.

Infatti, dalla chiusura del predicato di derivabilità $Teor$ sotto le leggi logiche e dal teorema di deduzione formalizzato[467], segue che, per ogni formula ϕ, vale[468]:

$$\mathsf{F}^* \vdash \neg Con_{\mathsf{F}} \rightarrow \neg Con_{\mathsf{F}+\phi}$$

Dunque, in particolare:

$$\mathsf{F}^* \vdash \neg Con_{\mathsf{F}} \rightarrow \neg Con_{\mathsf{F}+\neg Con_{\mathsf{F}}}$$

da cui, immediatamente, $\mathsf{F}^* \vdash \perp$.

Si è così fornito l'esempio di un sistema $(\mathsf{F} + \neg Con_{\mathsf{F}})$ valido, nel senso assunto dall'espressione, rispetto al quale non vale, tuttavia, che, aggiungendo ad esso la formula che ne esprime la validità, si ottiene un sistema consistente (come invece vorrebbe, sotto questa definizione di validità, l'inferenza (4.10)).

Ancora più allarmante per l'argomento di Penrose nella versione suddetta, è però il seguente fatto: l'inferenza (4.10) non è valida neppure prendendo, per $VAL(\mathsf{F})$, una qualche forma di validità ristretta ad una classe Γ di formule. Quest'affermazione la si può giustificare come segue.

Notoriamente, lo strumento al quale si ricorre per esprimere in un linguaggio formale la validità ristretta a determinate classi di formule, è rappresentato dai cosiddetti *principi di riflessione*, termine con il quale si intende indicare uno dei due schemi seguenti (che esprimono, rispettivamente, una forma non uniforme, 'locale', di validità, ed una forma uniforme della medesima):

$$Teor_{\mathsf{F}}\left(\ulcorner\phi\urcorner\right) \rightarrow \phi \qquad (\Gamma - Rfn_{\mathsf{F}})$$
$$\forall x[Teor_{\mathsf{F}}\left(\ulcorner\phi(\dot{\bar{x}})\urcorner\right) \rightarrow \phi(x)] \qquad (\Gamma - RFN_{\mathsf{F}})$$

dove $(\Gamma - Rfn_{\mathsf{F}})$ vale per ogni enunciato $\phi \in \Gamma$, e $(\Gamma - RFN_{\mathsf{F}})$ per ogni formula $\phi \in \Gamma$ contenente la sola x come variabile libera[469].

---

[467]Cioè il risultato dal quale discende:

$$\mathsf{F}^* \vdash Teor_{\mathsf{F}}(\ulcorner\phi \rightarrow \psi\urcorner) \leftrightarrow Teor_{\mathsf{F}+\phi}(\ulcorner\psi\urcorner)$$

per ogni formula ϕ, ψ.

[468]La conclusione del testo si ricava applicando le cosiddette "condizioni di derivabilità', le proprietà fondamentali del predicato $Teor$ (cfr. ad esempio [Smo77], p. 827), all'istanza dell'*a-fortiori* $(\perp \rightarrow (\phi \rightarrow \perp))$ e applicando il teorema di deduzione formalizzato.

[469]In luogo di $(\Gamma - RFN_{\mathsf{F}})$ si sarebbe equivalentemente potuto prendere lo schema:

$$\forall x Teor_{\mathsf{F}}\left(\ulcorner\phi(\dot{\bar{x}})\urcorner\right) \rightarrow \forall x\phi(x) \qquad (\Gamma - RFN'_{\mathsf{F}})$$

Definendo la proprietà $VAL(\mathsf{F})$ dell'argomento come la validità ristretta ad una classe Γ di formule, si può dimostrare, come nel caso precedente, che vale:

$$VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F} + \neg Val_{\mathsf{F}}) \tag{4.12}$$

sia che l'enunciato Val_{F} sia definito a partire da $(\Gamma - Rfn_{\mathsf{F}})$, sia che sia definito rispetto a $(\Gamma - RFN_{\mathsf{F}})$.

Supponiamo infatti che Γ sia una classe di enunciati per i quali è definibile, in F, un predicato di verità, ossia un predicato $Tr_{\Gamma}(x)$ tale che, per ogni $\phi \in \Gamma$:

$$\mathsf{F} \vdash \gamma \leftrightarrow Tr_{\Gamma}(\ulcorner \gamma \urcorner) \tag{4.13}$$

In tal modo potremmo indicare con $Val_{\mathsf{F}}^{\Gamma}$ l'enunciato che si ottiene, in ragione di (4.13), a partire da $(\Gamma - Rfn_{\mathsf{F}})$ per chiusura universale rispetto ai gödeliani degli elementi di Γ.

Tuttavia, è semplice dimostrare che la teoria $\mathsf{F}' := \mathsf{F} + \neg Val_{\mathsf{F}}^{\Gamma}$ è Γ-conservativa su F (ossia che vale che, se $\gamma \in \Gamma$ e $\mathsf{F}' \vdash \gamma$, allora $\mathsf{F} \vdash \gamma$): se infatti $\gamma \in \Gamma$ e $\mathsf{F}' \vdash \gamma$, si ricava immediatamente dal teorema di Löb[470] che, valendo $\mathsf{F} + Val_{\mathsf{F}}^{\Gamma} \vdash Teor_{\mathsf{F}}(\ulcorner \gamma \urcorner) \rightarrow \gamma$, si può ottenere $\mathsf{F} \vdash Val_{\mathsf{F}}^{\Gamma} \rightarrow \gamma$, da cui, unitamente all'ipotesi $\mathsf{F}' \vdash \gamma$, segue, per logica, $\mathsf{F} \vdash \gamma$[471].

Questo risultato ci consente di concludere che vale:

$$\Gamma - VAL(\mathsf{F}) \Rightarrow \Gamma - VAL(\mathsf{F} + \neg Val_{\mathsf{F}}^{\Gamma})$$

da cui la precedente (4.12), con $VAL(\mathsf{F})$ sostituito da $\Gamma - VAL(\mathsf{F})$, segue immediatamente.

---

per ogni formula $\phi \in \Gamma$ contenente la sola variabile libera x. Si dimostra infatti (lo ha fatto Feferman per primo - cfr. [Fef62], p. 276) che $(\Gamma - RFN_{\mathsf{F}})$ e $(\Gamma - RFN'_{\mathsf{F}})$ sono logicamente equivalenti in $\Sigma_1^0 - \mathsf{IA}$.

[470] Si tratta del teorema secondo il quale:

$$\mathsf{F} \vdash Teor_{\mathsf{F}}(\ulcorner \phi \urcorner) \rightarrow \phi \Leftrightarrow \mathsf{F} \vdash \phi$$

per ogni enunciato ϕ (si veda, ad esempio, [Smo77], p. 845).

[471] L'argomento nel caso in cui $Val_{\mathsf{F}}^{\Gamma}$ sia costruita, in modo analogo al caso considerato nel testo, a partire dallo schema $(\Gamma - RFN_{\mathsf{F}})$, è esattamente lo stesso, tranne che per il fatto che si deve fare uso del teorema di Löb in versione uniforme, sulla base del quale, per ogni formula ϕ contenente un'unica variabile libera:

$$\mathsf{F} \vdash \forall x[Teor_{\mathsf{F}}(\ulcorner \phi(\dot{\bar{\phi}})) \rightarrow \phi(x)] \Leftrightarrow \mathsf{F} \vdash \forall x \phi(x)$$

(si veda, ad esempio, *ibid.*, p. 848).

Ma ciò significa allora che, per mezzo di una costruzione esattamente analoga al caso della validità come consistenza, ossia considerando il sistema $\mathsf{F}^{**} := \mathsf{F} + \neg Val^{\Gamma}_{\mathsf{F}} + Val^{\Gamma}_{\mathsf{F}+\neg Val^{\Gamma}_{\mathsf{F}}}$, si ottiene il controesempio voluto[472].

Il sospetto, che a questo punto si fa pressante, che vi sia qualcosa di illegittimo nell'assunzione (4.10), diviene una certezza sulla base della seguente:

PROPOSIZIONE 4.4.1 (Lindström 2001) *Sotto le ipotesi:*

(i) *che $VERO(\mathsf{F}) \Rightarrow VAL(\mathsf{F})$, dove $VERO(\mathsf{F})$ indica la validità senza restrizioni di F;*

(ii) *che $VAL(\mathsf{F})$ sia esprimibile mediante una formula Val_F di $\mathcal{L}_F$,*

l'inferenza $VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F}^+)$ non è generalmente valida.

Dimostrazione: Sia ϕ la formula tale che, per il Lemma di Diagonalizzazione:

$$\mathsf{F} \vdash \phi \leftrightarrow \neg Val_{F+\phi} \tag{4.14}$$

Supponiamo che $VERO(\mathsf{F})$ valga. E supponiamo inoltre che valga $\overline{VAL}(\mathsf{F}+\phi)$ (cioè, che il sistema formale $\mathsf{F}+\phi$ non sia valido). Allora si avrebbe anche (per (ii)) che $\neg Val_{\mathsf{F}+\phi}$ è vera. Da questo e (4.14), si ottiene in particolare che ϕ è vera, che, insieme all'ipotesi di correttezza di F, dà $VERO(F+\phi)$, da cui (per (i)) $VAL(\mathsf{F}+\phi)$. Contraddizione.

Dunque:

$$VERO(\mathsf{F}) \Rightarrow VAL(\mathsf{F}+\phi)$$

D'altra parte, segue ovviamente da (4.14) che $\overline{CON}(\mathsf{F}+\phi+Val_{\mathsf{F}+\phi})$.∎

Siamo così portati a concludere che l'argomento di Penrose nella sua nuova e ambiziosa versione non sia valido in generale, dal momento che non appare

[472] L'argomento considerato nel testo sfrutta la possibilità di esprimere la validità mediante un singolo enunciato. Una simile possibilità è garantita, in un sistema come $\mathsf{\Sigma^0_1} - \mathsf{IA}$, per Γ coincidente con la collezione degli enunciati Σ^0_n (equivalentemente, Π^0_n), per ogni $n \in \mathbf{N}$. L'argomento dato continua a valere, considerando la teoria $\mathsf{F}' := \mathsf{F} + \neg(\Gamma - Rfn_{\mathsf{F}})$ (o $\mathsf{F}'' := \mathsf{F} + \neg(\Gamma - RFN_{\mathsf{F}})$), volendo indicare con ciò il sistema d'assiomi che si ottiene aggiungendo la negazione di ogni istanza del principio di riflessione corrispondente. In questo caso, tuttavia, Val_{F} potrebbe corrispondere ad uno schema d'assioma (ad esempio, anche in PA, ciò è vero per $\Gamma = \bigcup_{n\in\mathbf{N}} \Pi^0_n$).

affatto chiaro come si possa far sì che la sua premessa principale divenga soddisfacibile sotto le condizioni volute.

È pur sempre vero, si potrebbe obiettare, che si è fornita una confutazione di un argomento che non è quello di Penrose, ma che è ottenuto da esso sostituendo con (4.10) la premessa:

$$SONO(\mathsf{F}) \Rightarrow CON(\mathsf{F} + Sono_{\mathsf{F}})$$

la quale non è stata affatto confutata.

A ciò si potrebbe tuttavia rispondere che, come si è già notato, questa assunzione appare arbitraria, a meno che non si dia una definizione precisa di $SONO(\mathsf{F})$ che ne mostri, allo stesso tempo, l'esprimibilità come enunciato di un linguaggio formale. Fino a quel momento, la generalità dei risultati presentati sembra sufficiente a concludere, con Lindström, quanto meno che l'argomento così come esso è stato concepito originariamente da Penrose, "per quanto possibilmente vero in generale, non sia incontrovertibile in modo ovvio"[473].

La struttura della conoscenza matematica

La confutazione dell'argomento di Penrose che si è appena riportato, pone un quesito sul quale vale la pena soffermarsi per rendere più proficua l'analisi dei fondamenti logici dell'inferenza.

Se si guarda al ragionamento di Penrose con l'intenzione di eliminarne certi elementi di ambiguità si è condotti ad isolare l'inferenza:

$$VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F}^{+})$$

Essa, per vari motivi, può essere considerata il vero fulcro del problema che la posizione di Penrose pone. Innanzi tutto, questa regola ci consente di fornire una versione ragionevole dell'argomento che si suppone mostri un'intrinseca superiorità della mente umana (o una sua differenza essenziale) rispetto alle macchine (passando attraverso un correlato astratto e matematicamente trattabile di esse).

Per di più, l'inferenza in questione sembra possedere un'intrinseca validità intuitiva (se F è considerato corretto, la proposizione che esprime questo fatto è vera, e dunque produce, aggiunta agli assiomi di F, una teoria consistente).

[473][Lin01] p. 247.

Prim'ancora che un argomento generale mediante la proposizione 4.4.1, si sono tuttavia forniti esempi di possibili definizioni di $VAL(\mathsf{F})$ che mostrano come la convinzione nella correttezza di una simile regola sia infondata a livello metamatematico[474].

Quali indicazioni dovremmo dunque trarre dal fatto che essa non è generalmente vera?

Non è difficile vedere, che al problema è possibile dare una connotazione che prescinde del tutto dalla questione della correttezza o meno di argomenti come quello di Penrose, o, meglio, che ripropone il quesito se la mente sia o meno equivalente ad una macchina di Turing, sotto una luce diversa.

L'inferenza che si dimostra cruciale, asserisce infatti che a fronte dell'assunzione di correttezza di un certo sistema d'assiomi, un corpo di conoscenze acquisito, siamo sempre legittimati ad estendere tale sistema (in modo consistente) per mezzo di un enunciato che, sulla base del secondo teorema di incompletezza, è da esso indipendente, ma che è anche, per l'assunzione, vero. Quell'inferenza, come già si era visto nella giustificazione da parte di Penrose del corrispettivo di essa da lui assunto, esprime la nostra convinzione, o, per meglio dire (considerato il fatto che essa non è generalmente valida), l'illusione, che, nel processo di costruzione delle nostre convinzioni matematiche, si

[474] In un caso si è infatti fornito una confutazione dell'istanza della regola suddetta secondo la quale:

$$\frac{Con_F}{Con_{F+Con_F}}\ (R)$$

L'argomento sfrutta, come si è detto, quella parte del secondo teorema di Gödel che stabilisce $CON(\mathsf{F}) \Rightarrow CON(\mathsf{F}+\neg Con_F)$. Utilizzando la versione formalizzata del risultato di Gödel, si può allora dimostrare, mediante l'argomento riportato in precedenza nel testo, che la regola (R) è (formalmente) inconsistente con ogni sottosistema E di F tale che $\mathsf{E} \vdash Con_{\mathsf{F}}$. Un fatto analogo si verifica (per mezzo di un ragionamento del tutto simile) per una versione più generale e più forte di (R), ossia rispetto alla regola:

$$\frac{\Sigma^0_n - RFN_F}{\Sigma^0_n - RFN_{F+\Sigma^0_n - RFN_F}}\ (R^*)$$

Si dimostra infatti (cfr. [Pud99], p. 339) che vale:

$$\mathsf{\Sigma^0_1} - \mathsf{IA} \vdash (\Sigma^0_n - RFN_F) \rightarrow (\Sigma^0_n - RFN_{F+\neg\Sigma^0_n - RFN_F})$$

che è il corrispettivo della parte formalizzata del risultato di Gödel che si usa per confutare la validità di (R) (si ricordi che, in virtù del Σ^0_n predicato di verità per Σ^0_n-formule - cfr. [Smo77], p. 843 - è sempre possibile guardare a $(\Sigma^0_n - RFN_F)$ come ad un enunciato).

è sempre in grado di 'trascendere' il sistema delle conoscenze accettate, per mezzo di enunciati veri ed indipendenti da esso.

Per chiarire ulteriormente questo fatto, si può in un certo senso raffinare, sul piano logico, quanto si è detto finora.

In un dato formalismo, vi sono molteplici vie (un aspetto, questo, al quale si è già avuto modo di accennare), per esprimere forme ristrette della validità di un sistema assiomatico F. Tra esse, le principali sono rappresentate dai seguenti schemi:

$$\begin{array}{rcl} R_{\mathsf{F}} & := & \forall x[\neg Dim_{\mathsf{F}}(x, \ulcorner R(\mathsf{F})\urcorner) \to \exists y \leq x Dim_{\mathsf{F}}(y, \ulcorner \neg R(\mathsf{F})\urcorner)] \\ Con_{\mathsf{F}} & := & \neg Teor_{\mathsf{F}}(\ulcorner 0 = \overline{1}\urcorner) \\ \omega - Con_{\mathsf{F}} & := & Teor_{\mathsf{F}}(\ulcorner \exists x \phi\urcorner) \to \exists x \neg Teor_{\mathsf{F}}(\ulcorner \neg\phi(\dot{\overline{x}})\urcorner) \\ \Sigma^0_n - RFN_{\mathsf{F}} & := & \forall x[Teor_{\mathsf{F}}(\ulcorner \phi(\dot{\overline{x}})\urcorner) \to \phi(x)] \end{array}$$

Di queste formule, R_{F} corrisponde all'enunciato di Rosser, il quale, come è noto, risulta essere indecidibile in ogni sistema formale consistente, e che, per ricorrere ad un modo di dire più volte usato in relazione all'enunciato di Gödel, 'asserisce di se stesso' di essere indimostrabile (in un sistema siffatto) in modo 'locale' (in quanto refutabile mediante una prova che viene prima, secondo l'ordine stabilito dall'aritmetizzazione, della propria dimostrazione). Dato lo schema $(\omega - Con_{\mathsf{F}})$, che per ogni formula ϕ del linguaggio di F con la sola variabile libera x, esprime l'ω-consistenza del sistema, indichiamo poi, com'è d'uso, con $(1 - Con_{\mathsf{F}})$ la sua restrizione a formule primitive ricorsive (che contengano, cioè, al massimo quantificatori limitati). Per ciò che concerne infine $(\Sigma^0_n - RFN_{\mathsf{F}})$, si tratta, come risulta da quanto detto in precedenza, del principio uniforme di riflessione per F ristretto alle sole Σ^0_n-formule (per ogni $n \in \mathbf{N}$).

Rispetto a questi principi, alcuni risultati standard da tenere presenti (e che, in quanto standard, ci limiteremo ad enunciare rimandando ad altri testi per una trattazione esauriente[475]), sono i seguenti:

PROPOSIZIONE 4.4.2 (1) $\mathsf{\Sigma^0_1 - IA} \vdash \phi \to Teor_{\mathsf{F}}(\ulcorner \phi\urcorner)$, *per ogni* Σ^0_1-*enunciato* ϕ *(*Σ^0_1*-completezza formalizzata di* F*);*
(2) $\mathsf{\Sigma^0_1 - IA} \vdash Con_{\mathsf{F}} \leftrightarrow (\Pi^0_1 - Rfn_{\mathsf{F}}) \leftrightarrow (\Pi^0_1 - RFN_{\mathsf{F}})$*;*
(3) $\mathsf{\Sigma^0_1 - IA} \vdash (1 - Con_{\mathsf{F}}) \leftrightarrow (\Sigma^0_1 - Rfn_{\mathsf{F}})$*;*
(4) $\mathsf{\Sigma^0_1 - IA} \vdash (\Sigma^0_n - RFN_{\mathsf{F}}) \leftrightarrow (\Pi^0_{n+1} - RFN_{\mathsf{F}})$*;*

[475]Tra gli altri, si può ad esempio indicare il già citato [Smo77].

(5) $\Sigma^0_1 - \mathsf{IA} + \phi \vdash (\Sigma^0_n - RFN_{\mathsf{F}}) \rightarrow (\Sigma^0_n - RFN_{\mathsf{F}+\phi})$ *(per ogni* Π^0_k*-enunciato* ϕ *ed ogni* $n \geq k$*).*

Un fatto che, alla luce di quanto indicato, merita allora di essere rimarcato, è che, i principi indicati in precedenza, formano una gerarchia (in $\Sigma^0_1 - \mathsf{IA}$), considerando che vale:

$$\Sigma^0_1 - \mathsf{IA} \vdash \ldots \rightarrow (\Sigma^0_2 - RFN_{\mathsf{F}}) \rightarrow (\Sigma^0_1 - RFN_{\mathsf{F}}) \rightarrow (1 - Con_{\mathsf{F}}) \rightarrow Con_{\mathsf{F}} \rightarrow R_{\mathsf{F}}$$

È possibile dimostrare che, ed arriviamo così allo spunto di riflessione che, seguendo Pudlák[476], vogliamo proporre, a vari livelli di questa gerarchia di principi che esprimono la nostra 'fiducia' in un dato formalismo, valgono (in $\Sigma^0_1 - \mathsf{IA}$) relazioni della forma:

$$A_{\mathsf{F}} \rightarrow A_{\mathsf{F}+B_{\mathsf{F}}}$$

dove A_{F} e B_{F} sono principi tra quelli da noi indicati, con A_{F} dimostrabilmente più forte di B_{F}.

Scelto F tale che $\Sigma^0_1 - \mathsf{IA} \subseteq \mathsf{F}$, vale infatti la seguente:

PROPOSIZIONE 4.4.3 (Pudlák 1999) (1) $\mathsf{F} \vdash Con_{\mathsf{F}} \rightarrow Con_{\mathsf{F}+R_{\mathsf{F}}}$
(2) $\mathsf{F} \vdash (1 - Con_{\mathsf{F}}) \rightarrow (1 - Con_{\mathsf{F}+Con_{\mathsf{F}}})$
(3) $\mathsf{F} \vdash (\Sigma_{n+1} - RFN_{\mathsf{F}}) \rightarrow (\Sigma_{n+1} - RFN_{\mathsf{F}+\Sigma_n - RFN_{\mathsf{F}}})$
(4) $\mathsf{F} \vdash (\omega - Con_{\mathsf{F}}) \rightarrow (\omega - Con_{\mathsf{F}+Con_{\mathsf{F}}})$

Dim.: (1) è una conseguenza immediata di una parte del teorema di Rosser.

(2) Si assuma $1 - Con_{\mathsf{F}}$. Supponiamo inoltre $Teor_{\mathsf{F}+Con(\mathsf{F})}(\ulcorner\phi\urcorner)$, per ogni Σ^0_1-formula ϕ.

Il teorema di deduzione formalizzato stabilisce che:

$$\mathsf{F} \vdash Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\phi\urcorner) \leftrightarrow Teor_{\mathsf{F}}(\ulcorner Con_{\mathsf{F}} \rightarrow \phi\urcorner)$$

dove $(Con_{\mathsf{F}} \rightarrow \phi)$ è una Σ^0_1-formula sotto l'ipotesi ϕ lo sia.

Da 4.4.2, (3) si ricava allora:

$$\mathsf{F} + (1 - Con_{\mathsf{F}}) + Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\phi\urcorner) \vdash Con_{\mathsf{F}} \rightarrow \phi$$

[476] Cfr. [Pud99].

da cui, infine, poiché chiaramente $\mathsf{F} + (1 - Con_{\mathsf{F}}) \vdash Con_{\mathsf{F}}$, si ricava:

$$\mathsf{F} + (1 - Con_{\mathsf{F}}) \vdash Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\phi\urcorner) \to \phi$$

Avendo assunto $\phi \in \Sigma^0_1$, la tesi segue quindi nuovamente da (3) di 4.4.2.

(3) Utilizzando il Π^0_{n+1}-predicato di verità per Π^0_{n+1}-formule[477], si ha che, dimostrabilmente in $\mathsf{\Sigma^0_1} - \mathsf{IA}$, $(\Pi^0_{n+1} - RFN_{\mathsf{F}})$ è logicamente equivalente ad un Π^0_{n+1}-enunciato. Quello schema, può essere dunque considerato come un Π^0_{n+1}-enunciato.

Segue allora da 4.4.2, (5) che vale:

$$\mathsf{F} + (\Pi^0_{n+1} - RFN_{\mathsf{F}}) + (\Sigma^0_{n+1} - RFN_{\mathsf{F}}) \vdash (\Sigma^0_{n+1} - RFN_{\mathsf{F}+(\Pi^0_{n+1}-RFN_{\mathsf{F}})})$$

La tesi segue allora da una doppia applicazione del risultato (4) di 4.4.2.

(4) Assumiamo $(\omega - Con_{\mathsf{F}})$ e $Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\exists x\ \phi(x)\urcorner)$.

Dal momento che si ha:

$$\vdash (Con_{\mathsf{F}} \to \exists x\ \phi(x)) \leftrightarrow (Con_{\mathsf{F}} \to (Con_{\mathsf{F}} \wedge \exists x\ \phi(x)))$$

Si ottiene, combinando le proprietà basilari di $Teor$ con il teorema di deduzione formalizzato[478]:

$$\begin{array}{lll} \mathsf{F} & \vdash & Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\exists x\ \phi(v_0)\urcorner) \leftrightarrow \\ & \leftrightarrow & Teor_{\mathsf{F}}(\ulcorner Con_{\mathsf{F}} \to \exists x\ \phi(x)\urcorner) \leftrightarrow \\ & \leftrightarrow & Teor_{\mathsf{F}}(\ulcorner Con_{\mathsf{F}} \to (Con_{\mathsf{F}} \wedge \exists x\ \phi(x))\urcorner) \leftrightarrow \\ & \leftrightarrow & Teor_{\mathsf{F}}(\ulcorner\exists x \exists y[Dim_{\mathsf{F}}(y, \ulcorner\bot\urcorner) \vee (Con_{\mathsf{F}} \wedge \phi(x))]\urcorner) \end{array}$$

Da ciò si ricava, per contrazione dei quantificatori:

$$\mathsf{F} \vdash Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\exists x\ \phi(x)\urcorner) \leftrightarrow Teor_{\mathsf{F}}(\ulcorner\exists x[Dim_{\mathsf{F}}(x, \ulcorner\bot\urcorner) \vee (Con_{\mathsf{F}} \wedge \phi(x))]\urcorner)$$

Posto $\mathsf{F}^* := \mathsf{F} + (\omega - Con_{\mathsf{F}}) + Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\exists x\ \phi(x)\urcorner)$, segue allora, applicando $(\omega - Con_{\mathsf{F}})$ al secondo membro dell'equivalenza:

$$\mathsf{F}^* \vdash \exists x\, [\neg Teor_{\mathsf{F}}(\ulcorner\neg Dim_{\mathsf{F}}(\dot{\bar{x}}, \ulcorner\bot\urcorner) \wedge \neg(Con_{\mathsf{F}} \wedge \phi(\dot{\bar{x}}))\urcorner)]$$

[477] Cfr. [Smo77], p. 843.

[478] Usiamo $\bot$ come abbreviazione per una formula falsa.

Fissiamo un tale v. Poiché $\mathsf{F}^* \vdash Con_{\mathsf{F}}$ e $\mathsf{F}^* \vdash Con_{\mathsf{F}} \rightarrow \neg Dim_{\mathsf{F}}(v, \ulcorner\perp\urcorner)$, segue per 4.4.2, (1) che $\mathsf{F}^* \vdash Teor_{\mathsf{F}}(\ulcorner\neg Dim_{\mathsf{F}}(v, \ulcorner\perp\urcorner)\urcorner)$. Quindi, per chiusura di $Teor_{\mathsf{F}}$ sotto le leggi logiche:

$$\mathsf{F}^* \vdash \neg Teor_{\mathsf{F}}(\ulcorner\neg Dim_{\mathsf{F}}(v, \ulcorner\perp\urcorner) \wedge \neg(Con_{\mathsf{F}} \wedge \phi(v))\urcorner) \rightarrow \neg Teor_{\mathsf{F}}(\ulcorner\neg(Con_{\mathsf{F}} \wedge \phi(v))\urcorner)$$

da cui:

$$\mathsf{F}^* \vdash \exists x \neg Teor_{\mathsf{F}}(\ulcorner\neg(Con_{\mathsf{F}} \wedge \phi(\dot{x}))\urcorner)$$

E allora, per logica e teorema di deduzione formalizzato:

$$\mathsf{F}^* \vdash \exists x \neg Teor_{\mathsf{F}+Con_{\mathsf{F}}}(\ulcorner\neg\phi(\dot{x})\urcorner)$$

da cui discende la tesi. ■

Questo risultato ci consente di completare la discussione che abbiamo iniziato con la confutazione di una versione formalizzata del nuovo argomento di Penrose, dal momento che, la proposizione 4.4.3 ci mette nella condizione di 'giustificare' l'inferenza cruciale sulla quale quell'argomento, per come lo si è analizzato, pare fondarsi. Ciò può essere facilmente chiarito mediante l'esempio seguente.

Sulla base di (2) di 4.4.3, si può concludere che per ogni sistema formale F al quale si applichino i teoremi di incompletezza, vale:

$$\mathsf{F} \vdash (1 - Con_{\mathsf{F}}) \rightarrow (1 - Con_{\mathsf{F}+Con_{\mathsf{F}}})$$

Dalla formalizzazione dei teoremi di Gödel, segue invece che:

$$\mathsf{F} \vdash (1 - Con_{\mathsf{F}}) \rightarrow Con_{\mathsf{F}+Con_{\mathsf{F}}}$$

Mettendo i due risultati assieme si ricava così:

$$\begin{array}{ccl}
1 - Con_{\mathsf{F}} & \Rightarrow \mathsf{F} \vdash & Con_{\mathsf{F}^+} \\
\Downarrow & & \\
\mathsf{F} \vdash 1 - Con_{\mathsf{F}^+} & \Rightarrow \mathsf{F} \vdash & Con_{(\mathsf{F}^+)^+} \\
\Downarrow & & \\
\mathsf{F} \vdash 1 - Con_{(\mathsf{F}^+)^+} & \Rightarrow \mathsf{F} \vdash & Con_{((\mathsf{F}^+)^+)^+} \\
\vdots & & \vdots
\end{array}$$

(dove $\mathsf{F}^+ := \mathsf{F} + Con_{\mathsf{F}}$).

Capita dunque che l'inferenza $VAL(\mathsf{F}) \Rightarrow CON(\mathsf{F} + Con_{\mathsf{F}})$, che si era rivelata essere scorretta per $VAL(\mathsf{F}) := CON(\mathsf{F})$ per quanto apparentemente risultasse il contrario da un punto di vista intuitivo, possa essere verificata in una forma leggermente modificata (facendo cioè un'assunzione più forte, o meglio, assumendo il più piccolo principio che estende l'assunzione di partenza, Con_{F}, nella gerarchia succitata degli schemi legati alla validità di F). Tuttavia, non sarà sfuggito che, in questo caso, l'assunzione in questione risulta essere (dimostrabilmente in $\mathsf{\Sigma}_1^0 - \mathsf{IA}$), più forte, non solo della conclusione immediata, ma anche di tutta la successione di enunciati che si ottengono in F mediante la reiterazione di quell'inferenza.

Rispetto all'argomento di Penrose, ciò mette in luce in modo inequivocabile come quanto si suppone che una simile inferenza possa mostrare, la 'superiorità' del matematico umano rispetto ad ogni sistema formale, il suo 'correlato meccanico', sia già contenuta, in realtà, nell'assunzione su cui l'argomento fornito si fonda. Se si mettono assieme le conclusioni alle quali siamo giunti seguendo varie linee direttrici, l'esito della disamina dell'argomento di Penrose non parrebbe lasciare in piedi alcuna alternativa gradita allo scienziato inglese: o si decide di trincerarsi dietro una formulazione che è solo in apparenza più rigorosa della tesi filosofica contro cui si argomenta, oppure si riconosce la necessità di un approfondimento ulteriore e allora, o, attraverso un'analisi concettuale sulla scorta di quella di Gödel della *Gibbs Lecture* si è comunque in grado di svelare un significato dei termini in gioco che è logicamente compatibile con una tesi di tipo meccanicistico, oppure si giunge a far vedere, facendo uso di strumenti logici e metamatematici di indagine, come l'argomento poggi su un'assunzione che presume la tesi, e che, oltretutto, si presta ad essere contestata mediante una prova (quella di Lindström) sufficientemente generale.

Ma che cosa significa, invece, una tale conclusione rispetto a quell'inferenza come espressione di una convinzione teoretica, quella nella possibilità di poter sempre estendere un corpo di conoscenze acquisito mediante enunciati nuovi, cioè veri ed indipendenti dai principi fino ad un certo punto assunti?

Pudlák trae dal risultato l'indicazione che le assunzioni che guidano il progresso della conoscenza in matematica non siano sempre ed inevitabilmente vere in modo incontrovertibile, ma che possa capitare che, né più, né meno che in altri settori del sapere scientifico, "[s]i faccia uso di conoscenze meno sicure per congetturare la verità laddove non si sia in grado di dedurla da conoscenze e dati più sicuri"[479].

Piuttosto, dunque, che un indizio circa la superiorità dell'uomo su (ogni) macchina, sarebbe possibile intravedere, dietro certi risultati dell'indagine metamatematica, la necessità di rivedere, spogliandolo, in una certa misura, del dogmatismo che talvolta lo caratterizza, il paradigma classico della matematica come di un ambito nel quale si procede inequivocabilmente per mezzo del massimo rigore deduttivo: "[l]a sensazione che si possa sempre progredire e rendere le nostre assunzioni sempre più forti non riflette una nostra specifica abilità, ma è causata dalla progressiva, graduale diminuzione della nostra convinzione nella loro verità. Dopo tutto allora, la vaghezza è presente nel pensiero matematico, sebbene non nel processo deduttivo, ma nella decisione riguardo a quali assiomi si dovrebbe accettare."[480]

4.5 Alcune valutazioni di chiusura

Rispetto alla questione della quale ci siamo occupati nel presente capitolo, ci preme distinguere due aspetti.

Da un lato, c'è quanto potremmo ascrivere all'analisi e la precisazione di una tesi filosofica e che ci pare essere, non solo assolutamente legittimo, ma anche altamente auspicabile. Nel passare attraverso la critica di Lucas prima, e di Penrose poi, infatti, siamo stati in grado di dare una veste sempre più definita all'idea che, per dirla nei vaghi termini d'origine, la mente umana abbia una natura computazionale. Siamo cioè stati in grado di sostituire, o rendere superflue, le espressioni ambigue che compaiono nella tesi "La mente umana è una macchina", considerandone dei corrispettivi che non solo sono privi dell'ambiguità di quelle di partenza, ma che rendono quell'istanza passibile di un'indagine rigorosa a suo carico.

Quest'operazione, di andare a vedere cosa c'è dietro una posizione filosofica anche mediante gli strumenti della logica e della metamatematica, ci pare, come si diceva, del tutto degna.

Certo, non si tratta di un'operazione così rigorosa da non poter essere sottoposta a critica nei vari passaggi che essa comporta[481]. Questo fatto

[479] [Pud99], p. 341.
[480] *Ibid.*, p. 342.
[481] Anche se, si potrebbe sostenere, l'intera discussione è stata condotta nello spirito di quel "rigore informale" di cui parla Kreisel nel suo [Kre87] e che caratterizza a suo dire, in determinate circostanze, le problematiche di filosofia della matematica.

lo si è evidenziato, e cercato di discutere, nel caso della confutazione del nuovo argomento di Penrose. Ma ciò non significa che si possa ritenere la conclusione raggiunta come il frutto di un'improvvida analisi, e prendere spunto da ciò per rifugiarsi dietro una formulazione che mantiene, in una certa misura, quell'ambiguità della tesi filosofica che si intende confutare: il propugnatore della critica del meccanicismo, è invece costretto a fornire una propria, rigorosa interpretazione delle nozioni sulle quali egli fonda il convincimento che intende sostenere, e, in questo senso, deve mostrarne la praticabilità contribuendo con ciò al processo di chiarificazione al quale si è dato inizio.

C'è poi l'aspetto di merito, la valutazione di quello a cui conduce un simile processo di analisi e critica di una concezione filosofica.

Nella disamina della tesi meccanicistica, così come essa si è venuta delineando sulla base delle posizioni che abbiamo avuto modo di analizzare fino a questo punto, occorre distinguere tre momenti: quello logico, con l'elaborazione di un argomento fondato sui risultati dell'indagine metamatematica; quello scientifico, che considera la possibilità di un progetto di ricerca volto a svelare la natura della mente umana; quello filosofico, infine, che porta ad una serie di riflessioni sulla conoscenza, e sulla conoscenza matematica in particolare.

Queste tre componenti compaiono in diversa misura in coloro che sono intervenuti nel dibattito suscitato dall'"obiezione matematica". Esse, ad esempio, sono tutte presenti nel caso di Penrose, il quale, non solo ha l'ambizione, come si è visto, di riformulare, ed in modo decisivo, l'argomento gödeliano di Lucas, ma offre anche, in positivo, una spiegazione della coscienza facendo uso della fisica quantistica, ed associa, a tutto questo, uno sfondo concettuale che vede il problema della conoscenza passare dalla chiarificazione dei rapporti che si instaurano tra i tre mondi, quello mentale della coscienza, quello sensibile della fisica, quello astratto della matematica, che sono inter-indipendenti ed ontologicamente equivalenti (cioè dotati di un medesimo statuto d'esistenza).

Ma queste componenti sono anche presenti nelle due figure, Gödel e Turing, alle quali ci siamo richiamati continuamente e per un semplice motivo: perché a nostro avviso le rispettive posizioni in merito al rapporto tra menti e macchine, non devono essere considerate né come una mera curiosità storica, né come un orpello che impreziosisce e conferisce legittimità al dibattito sulla valenza filosofica dei teoremi di incompletezza. Ci pare piuttosto che le posizioni da loro assunte, dimostrino una maggiore sensibilità per la delicatezza delle nozioni in gioco, ed aggiungano sostanziali spunti di riflessione.

Gödel ad esempio, come si ricorderà per averlo più volte sottolineato nella prima parte del lavoro, era un convinto sostenitore della possibilità di indagare scientificamente il problema della coscienza o della natura della mente umana, ed era convinto che una simile ricerca avrebbe condotto ad una confutazione dell'interpretazione materialista di riduzione della mente alla sola attività cerebrale[482]. Egli era anche un propugnatore del platonismo in matematica, una posizione da lui sviluppata ad un certo livello di 'sofisticazione' filosofica, in relazione ad una riflessione affatto banale sulla natura del rapporto conoscitivo che si instaura con i concetti astratti. Per quanto legate ai risultati di incompletezza fossero (un fatto che si è cercato di evidenziare nel corso del nostro esame degli scritti gödeliani), entrambe queste posizioni erano indipendenti dalle conclusioni sul rapporto tra menti e macchine che, pure, Gödel aveva tratto dall'analisi dei propri celebri teoremi.

In particolare, nel fare ciò egli aveva già indicato, come si è ribadito al §4.4.1, la dipendenza degli esisti di una simile analisi dalla chiarificazione di una nozione, il corrispettivo della "conoscenza matematica incontrovertibile" di Penrose, che di per sé non risulta affatto essere un'idea chiara e distinta.

Prim'ancora che attraverso l'individuazione di certi limiti logici delle inferenze proposte, è già questo un elemento che pare mettere a serio rischio la conclusività del dibattito: perché dietro la necessità di chiarire cosa si intenda con la collezione $\mathbf{M}^\star$ delle verità matematiche incontrovertibili, o con la "matematica propriamente detta" che contiene gli enunciati percepiti con certezza matematica, sembra profilarsi la priorità di risolvere la stessa questione, il disvelamento della 'reale natura' della disciplina, che è stata all'origine dell'annosa diatriba sui fondamenti della matematica (e che sembra, quindi, riproporsi qui ad un 'livello cognitivo', legata cioè al problema dell'essenza della ragione umana).

Ciò non significa, però, che quanto è possibile concludere dall'indagine sul fondamento metamatematico e sulla correttezza logica dell'inferenza di Penrose, non offra spunti di un certo interesse.

[482]In [Wan96], Wang cita (p. 207, 6.5.3) una riflessione di Gödel relativa al fatto che "[l]a meccanica quantistica è solamente finitaria", la quale potrebbe far pensare che, in contrasto con quanto sostiene Penrose, egli non vedesse nella fisica dei quanti uno strumento risolutivo per quell'indagine scientifica sulla mente umana alla quale alludeva (ovviamento questo accenno è poco più che una mera curiosità considerando il carattere cursorio dell'osservazione gödeliana).

Le considerazioni di Pudlák a latere di certe implicazioni metamatematiche della questione, ad esempio, meritano di essere ribadite.

Secondo il logico ceco, l'analisi della convinzione, implicitamente assunta, o, come nel caso di Penrose, resa esplicita attraverso certe inferenze che la incarnano, che si possa sempre andare oltre un insieme di principi dato, fa trasparire una peculiarità euristica alla base del processo di costruzione della conoscenza in campo matematico. Questa, emergerebbe dal fatto che, in corrispondenza al risalire lungo quella gerarchia di enunciati che sono espressione, anche in forma parziale, della nostra fiducia in un certo formalismo (un procedimento che, come si è visto, può costituire il fondamento per giustificare inferenze del tutto simili a quelle da cui Penrose prende le mosse), possono corrispondere, a livello concettuale, dei salti niente affatto pacifici.

Ciò lo si può vedere bene se si considera che, volendo proseguire quella stessa gerarchia andando oltre lo schema $(\Sigma_n^0 - RFN_{\mathsf{F}})$, e quindi considerare un principio (ad esempio, $(\Sigma_0^1 - RFN_{\mathsf{F}})$) che esprima la validità di un sistema formale rispetto alla collezione di tutti gli enunciati aritmetici, si sarebbe costretti, tenendo presente che il predicato di verità aritmetica non è aritmeticamente definibile per il ben noto teorema di Tarski, a cambiare cornice formale, passando dall'aritmetica ad un qualche sottosistema dell'analisi. Ma ciò significa fondare la nostra fiducia nella teoria dei numeri, un ambito rispetto al quale riteniamo di esercitare l'intuizione nella sua forma più indiscutibile, su un sistema di nozioni che non godono affatto di una fiducia altrettanto incondizionata.

Un simile esempio, forse più di ogni altro, chiarisce perché l'inferenza che ne è alla base, anche qualora venisse esercitata in relazione a principi che esprimono forme di validità ristrette a certe classi di formule, e che quindi, in quanto tali, non comportano dei 'salti concettuali' come quello al quale si è appena alluso, può dare origine ad un'immagine (l'idea che si ricorre ad assunzioni incerte), la quale, pur apparendo inconsueta per la disciplina alla quale si riferisce, non è tuttavia del tutto sorprendente. Di fronte a questioni rispetto alle quali non siamo nella posizione di poter prendere una decisione sulla base di principi noti (tipicamente: nel caso di nuovi assiomi), si può dare vita a più linee di sviluppo di una certa area della matematica, alle quali guardiamo come ad estensioni veritiere in misura proporzionale alla nostra fiducia nelle assunzioni che le hanno generate.

Se ci astraiamo, generalizzando, dall'ambito metamatematico, si può trarre da riflessioni di questo genere uno spunto che non pare essere ininfluente rispetto alla problematica che siamo venuti analizzando nel presente capitolo, e che, se si eccettua qualche riferimento contenuto negli scritti di Turing, non si è, fino ad ora, approfondito come forse avrebbe meritato.

Il fatto è, che se si intende dare, a livello formale, un modello plausibile dei procedimenti adottati dal matematico umano, non si può considerare di secondaria importanza il ruolo ricoperto da certe strategie euristiche come quella che sembra emergere da quest'analisi. Come in parte Turing pareva voler sottolineare, prima ancora che su un piano squisitamente logico, i sistemi formali e le procedure effettive di computo, denotano i loro limiti, anche quelli sui quali vertono certi teoremi logici, proprio in relazione a simili aspetti del problema. Si tratta quindi di vedere se, ed in quale misura, un simile proposito, riformulare in modo euristicamente più soddisfacente le nozioni in gioco, risulti essere praticabile.

È ciò di cui ci vogliamo occupare con il prossimo capitolo.

Capitolo 5

La questione euristica

Nel passare al vaglio della metamatematica l'argomento di Penrose, si sarà assistito ad uno slittamento di valenza dell'analisi. Si è passati cioè da considerare la plausibilità o meno di una posizione avversa ad una tesi filosofica di tipo meccanicista (opportunamente riformulata, come si è visto, in termini logici), ad alcune osservazioni relative piuttosto a certi aspetti del processo di 'costruzione' di un corpo di conoscenze, e dunque non necessariamente legate al problema del rapporto fra menti e macchine (anche se, come si è detto, in quelle considerazioni si possono ravvisare limiti ulteriori per una posizione come quella di Lucas e Penrose).

Si è così visto, ad esempio, che dietro certe assunzioni che riflettono la nostra convinzione che un simile processo sia sempre diretto ad un'estensione progressiva delle conoscenze raggiunte ad un certo stadio, si potrebbero nascondere particolari assunzioni relative alla validità, in un qualche senso, dei principi assunti. È vero che simili assunzioni, mettendoci nelle condizioni di trascendere un sistema di assiomi (nessuna teoria formale è chiusa, per i teoremi di incompletezza, sotto i principi di riflessione), ci permettono di estendere propriamente il corpo di conoscenze dato. Ma è anche vero, e per lo stesso motivo (nessuna di quelle assunzioni è perciò giustificabile in termini del sistema di partenza), che ciò ci porterebbe a guardare ad una simile estensione come ad un sistema fondato in modo essenziale sulla *convinzione* (non inoppugnabile) nella verità di certi principi, che decresce al crescere della forza logica dell'assunzione fatta.

Se guardiamo a questo fatto da un punto di vista generale (non confinato cioè, all'analisi delle teorie assiomatiche formali), se ne ricava un'indicazione di tipo metodologico-teoretico che, almeno in una certa misura, potrebbe presumibilmente venire legittimata anche da un punto di vista storico: nell'estendere le nostre teorie si fa talvolta ricorso ad ipotesi euristicamente, ma non assolutamente, giustificate, che rappresentano dei tentativi di fornire nuovi assiomi; ovviamente, proprio per il loro carattere essenzialmente ipotetico, torneremmo sui nostri passi, cancellando alcune delle ipotesi siffatte, qualora esse si rivelassero contraddittorie[483].

Si ricava cioè l'impressione che, talvolta, si dia vita ad un processo 'evolutivo' delle conoscenze basato su tentativi ed errori.

Ciò di cui ci vogliamo occupare innanzi tutto nel presente capitolo, è la possibilità di caratterizzare logicamente un tale processo, così da creare un modello che sia più fedele a certe procedure euristiche, come quelle di cui si è appena discusso, messe presumibilmente in atto in ambito scientifico.

Ritorneremo poi, alla luce dei risultati ottenuti in questo senso, sulla valutazione della tesi meccanicista.

5.1 Procedure per tentativi ed errori ed insiemi Δ^0_2

L'indagine sulle procedure per tentativi ed errori risale ai due lavori [Put65] e [Gol65].

L'idea intuitiva di un processo che si dispiega per tentativi ed errori può essere illustrata come segue: supponiamo di avere una procedura di decisione (una macchina di Turing) che, per ogni input, produce una successione di output (per semplificare le cose si può pensare agli output come ad una successione di '`SI`' e '`NO`'), solo l'ultimo dei quali deve essere inteso come la risposta della macchina. Supponiamo inoltre che non disponiamo di una procedura effettiva per determinare se una certa risposta sia o meno quella definitiva (cioè l'ultima di una serie).

La macchina, quindi, detto in altri termini, può 'cambiare idea' un numero finito ma arbitrario di volte, e non è possibile determinare, ad ogni istante, se la macchina sia sul punto di cambiare idea un'altra volta o meno.

Formalmente, la resa di questa idea è racchiusa nella seguente definizione:

[483]Questo fatto in particolare, distingue l'osservazione del testo dalla problematica metamatematica relativa al tipo di enunciati ai quali essa si ispira: nel caso dei principi di riflessione, siamo di fronte infatti ad ipotesi indipendenti dalla teoria assiomatica alla quale si riferiscono, ed il sistema ottenuto da essa per aggiunta di simili enunciati è dunque non contraddittorio.

DEFINIZIONE 5.1.1 *$P(x_1, \dots, x_k)$ è un predicato determinato per tentativi ed errori se e solo se esiste una funzione ricorsiva $k+1$-aria f tale che, per ogni $n_1, \dots, n_k \in \mathbf{N}$:*

$$\begin{aligned} P(n_1, \dots, n_k) &\Leftrightarrow \lim_{y \to \infty} f(n_1, \dots, n_k, y) = 1 \\ \overline{P}(n_1, \dots, n_k) &\Leftrightarrow \lim_{y \to \infty} f(n_1, \dots, n_k, y) = 0 \end{aligned}$$

dove (per ogni $m \in \mathbf{N}$):

$$\lim_{y \to \infty} f(x_1, \dots, x_k, y) = m := \bigvee y \bigwedge z[z \geq y \to f(x_1, \dots, x_k, z) = m]$$

Si considerano cioè quei predicati per i quali (e per il complemento dei quali) esiste una procedura effettiva di decisione, che può sì 'cambiare idea' su ogni dato input, ma che si 'stabilizza' dopo un numero finito di indecisioni[484].

In termini equivalenti, data una successione enumerabile di insiemi ricorsivi di numeri naturali $\langle A_n \rangle_{n \in \mathbf{N}}$ (la successione degli indici della funzione caratteristica dei quali è enumerata dalla f di cui sopra, al variare di y), siamo interessati agli insiemi A tali che vale la seguente:

$$A = \bigcup_{k \in \mathbf{N}} \bigcap_{m \geq k} A_m = \bigcap_{k \in \mathbf{N}} \bigcup_{m \geq k} A_m$$

Mutuando la notazione usata nella definizione, chiamiamo *limiti* gli insiemi $A \subseteq \mathbf{N}$ per i quali vale la condizione precedente (scrivendo $A = \lim_{n \to \infty} A_n$), ed indichiamo con Λ la collezione di essi. Un insieme $A \subseteq \mathbf{N}$ è un *limite inferiore* (*superiore*) se esiste una successione enumerabile di insiemi ricorsivi $\langle A_n \rangle_{n \in \mathbf{N}}$ tale che:

$$A = \underline{\lim}_{n \to \infty} A_n = \bigcup_{k \in \mathbf{N}} \bigcap_{m \geq k} A_m \ (\overline{\lim}_{n \to \infty} A_n = \bigcap_{k \in \mathbf{N}} \bigcup_{m \geq k} A_m)$$

Indichiamo con $\underline{\Lambda}$ ($\overline{\Lambda}$) la classe dei limiti inferiori (superiori).

Il risultato fondamentale relativo ai limiti, è il seguente teorema[485] che permette di darne una caratterizzazione in relazione alla gerarchia aritmetica degli insiemi di numeri naturali:

[484] Rispetto all'idea intuitiva di partenza, si ha, in questo caso, una successione infinita di `SI` e `NO` (che, come si è detto, è costituita da tutti `SI` da un certo punto in poi), piuttosto che una successione finita di output. Come risulterà anche dal prosieguo del paragrafo, le due definizioni (rispetto a successioni di output finite ed infinite) sono equivalenti.

[485] Cfr. [Put65], p. 51, Teor. 1.

TEOREMA 5.1.2 (1) *$A \subseteq \mathbf{N}$ è un limite se e solo se $A \in \Delta^0_2$.*
(2) *$A \subseteq \mathbf{N}$ è un limite inferiore se e solo se $A \in \Sigma^0_2$.*
(3) *$A \subseteq \mathbf{N}$ è un limite superiore se e solo se $A \in \Pi^0_2$*[486].

È utile richiamare in questa sede, alcuni risultati standard che permettono di apprezzare meglio il significato di una simile caratterizzazione. Le proposizioni che abbiamo in mente, sono legate all'indagine sui gradi di irresolubilità dei problemi, ed in particolare alle nozioni di riduzione tra insiemi di numeri naturali ad esse connesse[487].

Diamo intanto la seguente definizione:

DEFINIZIONE 5.1.3 (i) *$A \subseteq \mathbf{N}$ è 1-riducibile ad un insieme $B \subseteq \mathbf{N}$ (in simboli $A \leq_1 B$) se e solo se esiste una funzione ricorsiva f iniettiva tale che:*

$$\forall x[x \in A \Leftrightarrow f(x) \in B]$$

(ii) *$A \subseteq \mathbf{N}$ è Turing-riducibile all'insieme $B \subseteq \mathbf{N}$ (si scrive $A \leq_T B$) se e solo se la funzione caratteristica c_A di A è B-ricorsiva*[488].

Si osserva facilmente che vale:

$$A \leq_1 B \Rightarrow A \leq_T B$$

ma che non vale il viceversa[489].

Di particolare importanza, nello studio della riducibilità tra insiemi, risulta essere la cosiddetta 'gerarchia del salto', costruita a partire dall'operazione insiemistica $j(A)$, definita da (per ogni $A \subseteq \mathbf{N}$), $j(A) := \{x \mid \bigvee y T^A_1(x,x,y)\}$.

[486]Risultati analoghi valgono ovviamente anche se formulati in termini di predicati $P(x_1, \ldots, x_n)$ determinati per tentativi ed errori (ed è in questa forma che il teorema viene enunciato e dimostrato da Putnam).

[487]Non pretendiamo affatto di offrire un quadro esaustivo sull'argomento, peraltro vastissimo. Ci limiteremo piuttosto ad un brevissimo accenno ad alcune nozioni di base, rimandando ad altri testi (ad esempio, [Rog67], capp. 6 e segg.) per una trattazione sistematica della problematica.

[488]Una funzione f è A-ricorsiva, per A insieme, se e solo se essa è ottenuta per composizione, sostituzione, recursione primitiva e minimalizzazione dalle usuali funzioni iniziali e, in più, dalla funzione caratteristica di A, $c_A(x)$.

[489]Preso l'insieme $K := \{x \mid \bigvee y T_1(x,x,y)\}$, ossia l'insieme degli indici di funzioni ricorsive parziali che sono definite sul proprio indice, ed il suo complemento $\overline{K}$, è ovvio che valga $K \equiv_T \overline{K} := K \leq_T \overline{K}$ & $\overline{K} \leq_T K$. Questi stessi insiemi risultano però addirittura incomparabili rispetto alla 1-riducibilità, come si può facilmente verificare dal fatto che valgono: (i) $K \leq_1 \overline{K} \Rightarrow \overline{K} \leq_1 K$, (ii) $\overline{K} \leq_1 K \Rightarrow \overline{K}$ ricorsivamente enumerabile, (iii) $\overline{K}$ non ricorsivamente enumerabile.

Si definisce di conseguenza la seguente successione:

$$\begin{aligned}
\emptyset^{(0)} &:= \emptyset \\
\emptyset^{(n+1)} &:= j(\emptyset^{(n)}) \\
\emptyset^{(\omega)} &:= \{< x, y >|\; x \in \emptyset^{(y)}\}
\end{aligned}$$

La collezione così definita svolge, come si diceva, un ruolo cruciale per la decidibilità o risolubilità dei problemi, come dimostrano i risultati seguenti[490]. Gli stessi risultati ci consentono di approfondire ulteriormente il senso della caratterizzazione dei limiti data in precedenza[491]:

TEOREMA 5.1.4 1. *Per ogni $n \geq 0$ ed $A \subseteq \mathbf{N}$:*

$$A \in \Sigma^0_{n+1} \cap \Pi^0_{n+1} \Leftrightarrow A \leq_T \emptyset^{(n)}$$

2. *Per ogni $n \geq 0$ ed $A \subseteq \mathbf{N}$:*

$$\begin{aligned}
A \in \Sigma^0_n &\Leftrightarrow A \leq_1 \emptyset^{(n)} \\
A \in \Pi^0_n &\Leftrightarrow A \leq_1 \overline{\emptyset}^{(n)}
\end{aligned}$$

Da essi (e dal fatto che $\Delta^0_2 := \Sigma^0_2 \cap \Pi^0_2$), osservando che $K = \{x \mid \bigvee y T_1(x, x, y)\} = \emptyset^{(1)}$, si ricava che:

$$\begin{aligned}
A \in \Delta^0_2 &\Leftrightarrow A \leq_T \emptyset^{(1)} = K \\
A \in \Sigma^0_2 &\Leftrightarrow A \leq_1 \emptyset^{(2)} = j(K) \\
A \in \Pi^0_2 &\Leftrightarrow A \leq_1 \overline{\emptyset}^{(2)} = j(\overline{K})
\end{aligned}$$

Si può quindi apprezzare in modo immediato il significato del passaggio dalla considerazione degli insiemi ricorsivamente enumerabili (cioè Σ^0_1) come nel caso delle usuali teorie assiomatiche formali, alla classe Λ dei limiti, o a sottoinsiemi di quest'ultima. Ma ciò consente anche di fornire una lettura della collezione Δ^0_2 in termine di risolubilità di problemi: ad essa appartengono tutti e soli gli insiemi di numeri naturali, il problema dell'appartenenza ai quali è riducibile (nel senso specificato) al problema dell'arresto (rappresentato dall'appartenenza a K).

---

[490]Cfr. [Rog67], pp. 314-316, per i dettagli.

[491]Si tratta, come indicato, della versione speciale di due teoremi che valgono, in ovvia riformulazione, anche in versione generalizzata ossia rispetto alla computabilità relativizzata (cfr. *ibid.*, *loc. cit.*).

Proprio questi risultati, offrono la possibilità di caratterizzare la classe Λ dei limiti in un modo alternativo alla definizione di Putnam. In sostanza, anziché considerare i limiti di successioni enumerabili di insiemi decidibili, è sufficiente dare la definizione rispetto ad una successione enumerabile (relativamente agli indici canonici) di insiemi finiti[492].

Ciò significa porre:

DEFINIZIONE 5.1.5 *$A \subseteq \mathbf{N}$ è un limite se e solo se esiste una funzione ricorsiva f tale che $A = \lim_{n\to\infty} D_{f(n)}$.*

Indicata con Λ^f la classe dei limiti così definita, si osserva ovviamente che $\Lambda^f \subseteq \Delta^0_2$, e che se $A \subseteq \mathbf{N}$ e A è ricorsivamente enumerabile allora $A \in \Lambda^f$[493]. È dimostrato poi in [Mag74] che vale[494]: dati $A, B \subseteq \mathbf{N}$, se $B \in \Lambda^f$ e $A \leq_T B$ allora $A \in \Lambda^f$.

Dal ché si ricava immediatamente, essendo K ricorsivamente enumerabile e quindi un limite nel senso della definizione 5.1.5 e valendo il teorema di Post (cioè 5.1.4, 1), che $\Lambda^f = \Delta^0_2 = \Lambda$.

5.2 I Sistemi Dialettici di Roberto Magari

Torniamo al problema di partenza: vorremmo riuscire a caratterizzare logicamente in modo soddisfacente il processo mediante il quale, in certe circostanze, si estende un corpo di conoscenze dato mediante enunciati che rappresentano possibili nuovi assiomi; le conseguenze di essi vengono poi cercate, ed alcune dei principi assunti vengono cancellati, qualora rivelino la loro contraddittorietà.

Ci sono quindi *due* eventi di cui si deve dare conto, essendo l'uno l'aggiunta di nuovi assiomi ad un corpo di conoscenze dato, l'altro la cancellazione

---

[492] La terminologia e la notazione è mutuata da Rogers (cfr. [Rog67], p. 70): dato un insieme $A \subseteq \mathbf{N}$ finito e non vuoto (con $A = \{x_0, \ldots, x_n\}$) si chiama $x := 2^{x_0} + \ldots + 2^{x_n}$ l'indice canonico di A (e si scrive $A = D_x$), essendo 0 l'indice canonico di A, qualora $A = \emptyset$.

[493] Assunto che valga, per qualche $e \in \mathbf{N}$, $A = W_e := \{x \mid \bigvee y T_1(e, x, y)\}$ (A, essendo ricorsivamente enumerabile, coincide cioè con il dominio di una funzione ricorsiva parziale indiciata da e), si consideri la successione di insiemi finiti $\langle M^e_n \rangle_{n\in\mathbf{N}}$, dove, per ogni $n \in \mathbf{N}$, sia $M^e_n := \{x < n \mid \bigvee z T_1(e, x, z) \ \&\ Lgh(z) \leq n\}$. Si osserva allora che esiste ovviamente una funzione ricorsiva f tale che, per ogni $n \in \mathbf{N}$, $D_{f(n)} = M^e_n$, e si verifica facilmente che si ha anche $\bigvee x \bigwedge y \geq x (z \in M^e_y) \Rightarrow \bigwedge x \bigvee y \geq x (z \in M^e_y) \Rightarrow z \in A \Rightarrow \bigvee x \bigwedge y \geq x (z \in M^e_y)$.

[494] Cfr. [Mag74], p. 132, Lemma 5.

delle ipotesi inconsistenti con esso. Si vuole inoltre misurare, trattandosi del tema principale all'attenzione nel presente lavoro, la portata delle limitazioni logiche rispetto al modello così individuato.

Questo compito è stato svolto, in un modo che ci pare meritorio di essere analizzato più in dettaglio, dal matematico Roberto Magari intorno alla metà degli anni '70. A causa dell'approccio prevalentemente algebrico alla teoria dei sistemi d'assiomi di Magari, una brevissima introduzione alle nozioni di base risulta essere doverosa[495].

Il concetto fondamentale è quello di *operatore di chiusura algebrico* su un certo universo V.

Un'operazione Φ che associa ad ogni $A \subseteq V$ un $\Phi(A) \subseteq V$, è un *operatore di chiusura* (o *di Moore*) se soddisfa le seguenti condizioni:

(1) $A \subseteq \Phi(A)$ (Φ è *inclusivo*)

(2) $A \subseteq B \Rightarrow \Phi(A) \subseteq \Phi(B)$ (Φ è *monotono*)

(3) $\Phi(\Phi(A)) = \Phi(A)$ (Φ è *idempotente*)

Un operatore di chiusura Φ è *algebrico* se, in più, soddisfa:

(4) $\Phi(A) = \bigcup\{\Phi(F) \mid F \subseteq A \ \& \ |F| < \aleph_0\}$

Una terna $\langle P, \Phi, c\rangle$ è un *sistema formale* se e solo se valgono:

(i) P è un insieme decidibile.

(ii) Φ è un operatore di chiusura algebrico che associa a sottoinsiemi $X \subseteq P$ ricorsivamente enumerabili, sottoinsiemi $\Phi(X) \subseteq P$ ancora ricorsivamente enumerabili.

(iii) $c \in P$ e $\Phi(\{c\}) = P$.

Non sarà sfuggita l'analogia diretta con l'approccio assiomatico formale più tradizionale: P è l'insieme delle proposizioni del sistema, Φ è l'operatore di deduzione e c rappresenta una contraddizione.

Dato un sistema formale $\mathsf{S} = \langle P, \Phi, c\rangle$, indichiamo allora con $\Phi(\emptyset)$ l'insieme dei suoi teoremi, e, per ogni sottoinsieme enumerabile A di P, con $\Phi(A)$

[495] Ancora una volta non intendiamo fare altro che introdurre i pochi concetti utili alla migliore comprensione di quanto si andrà dicendo. Si rimanda ad altri testi (ad esempio [Hal62]) per una trattazione esaustiva dell'argomento.

l'insieme dei teoremi della teoria che ha gli elementi di A come assiomi[496]. Nel seguito, si considererà il caso in cui l'insieme delle proposizioni P coincide con $\mathbf{N}$[497]. Ciò non costituisce ovviamente alcuna restrizione essenziale alla generalità di quanto diremo.

La proposta di Magari al fine di dare conto dell'approccio 'per tentativi ed errori' al problema dell'estensione di un corpo di conoscenze dato di cui si è detto, consiste nel passare dai sistemi formali ordinari $\mathsf{S} = \langle \Phi, c \rangle$ come sopra, a *sistemi dialettici*[498] $\mathsf{S}_d = \langle h, f, c \rangle$ definiti nel modo seguente[499]:

DEFINIZIONE 5.2.1 *Una terna* $\mathsf{S}_d = \langle h, f, c \rangle$ *è un sistema dialettico se valgono le seguenti condizioni:*

(1) h, f *sono funzioni ricorsive totali e* f *è una biezione da* $\mathbf{N}$ *in* $\mathbf{N}$*;*

(2) $c \in \mathbf{N}$ *e* $W_{h(2^c)} = \mathbf{N}$[500]*;*

(3) $W_{h(0)} \neq \emptyset$[501]*;*

Se $\mathsf{S}_d = \langle h, f, c \rangle$ è un sistema dialettico, la funzione h verrà detta, rispetto al processo che si intende descrivere, la *funzione di deduzione* di S_d (che associa in particolare ad ogni sottoinsieme finito X di $\mathbf{N}$ un sottoinsieme ricorsivamente enumerabile - quello dei teoremi di X), f è la *funzione proponente* (che ad ogni stadio del processo fornisce una nuova ipotesi provvisoria che estende quelle precedentemente assunte) e c è la contraddizione.

Il passo successivo consiste nel definire una nozione di dimostrabilità per un sistema dialettico. L'idea fondamentale, mediante la quale si intende

---

[496] Si può dunque leggere, per ogni $A \subseteq P$ ricorsivamente enumerabile, $\alpha \in \Phi(A)$ come $A \vdash \alpha$.

[497] In particolare, eviteremo di indicare i sistemi formali come terne, scrivendo semplicemente $\mathsf{S} = \langle \Phi, c \rangle$.

[498] La scelta del nome non ha alcun legame con il significato dell'espressione 'dialettico' in ambito filosofico. Esso deve essere ricondotto piuttosto all'idea dalla quale siamo partiti: il problema di caratterizzare il processo di aggiunta e cancellazione di nuovi assiomi.

[499] Come in precedenza, si ricorre alla notazione di Rogers in [Rog67], e si indica, per ogni $n \in \mathbf{N}$, con W_n il dominio della funzione indiciata da n in una qualche codifica standard delle funzioni ricorsive parziali.

[500] Si ricordi che, sulla base della notazione assunta, 2^c è l'indice canonico dell'insieme finito che contiene c come unico elemento (ossia, vale $D_{2^c} = \{c\}$), e che W_n sta per il dominio della funzione indiciata da n. Questa condizione equivale allora a dire che la funzione indiciata da $h(2^c)$ è totale.

[501] Si ricordi che vale $D_0 = \emptyset$.

catturare il concetto informale di una procedura inferenziale che procede per tentativi ed errori, consiste nello stratificare (in una successione di insiemi finiti) il processo di dimostrazione che, nel caso dei sistemi formali usuali, consiste nell'enumerarne in modo effettivo i teoremi. Ciò è reso possibile dal considerare, ad un certo 'livello' $n \in \mathbf{N}$, solo gli enunciati enumerati, tramite una procedura effettiva determinata dal sistema dialettico di partenza, in corrispondenza ad input minori di n e in un numero di passi (ossia mediante una computazione di lunghezza) minore di n. Si tratta, poi, di identificare l'insieme delle conseguenze di assunzioni date, con quanto si ottiene da una simile procedura, per ogni $n \in \mathbf{N}$.

La seguente definizione, ripresa da Magari, non fa altro che ribadire in termini più rigorosi quanto si è cercato di spiegare informalmente:

DEFINIZIONE 5.2.2 *Dato un sistema dialettico $\mathsf{S}_d = \langle h, f, c\rangle$ si definiscono per ogni $X \subseteq \mathbf{N}$ con $X = D_m$ per qualche $m \in \mathbf{N}$ (ossia X insieme finito di numeri naturali):*

(i) $\Phi^h_n(X) := \{y < n \mid \bigvee z T_1(h(x), y, z) \;\&\; Lgh(z) \leq n\} \cup X$, *(dove $Lgh(z)$ indica la lunghezza della computazione codificata da z);*

(ii)

$$\begin{aligned} K^h_0(X) &:= X \\ K^h_{n+1}(X) &:= \bigcup\{\Phi^h_{n+1}(Y) \mid Y \subseteq K^h_n(X)\} \end{aligned}$$

(iii) *per ogni $Z \subseteq \mathbf{N}$:*

$$\Phi^h(Z) := \bigcup_{n \in \mathbf{N}} \{K_n(X) \mid X \subseteq Z \;\&\; |X| < \aleph_0\}$$

Fatto ciò, si è così in grado di definire le nozioni cruciali per la costruzione di Magari:

DEFINIZIONE 5.2.3 *Dato un sistema dialettico $\mathsf{S}_d = \langle h, f, c\rangle$, si definiscono una successione $\langle s_n\rangle_{n\in\mathbf{N}}$ di* stati provvisori *del sistema (dove, per ogni $n \in \mathbf{N}$ $s_n := \langle (x_0, X_0), \ldots, (x_r, X_r)\rangle$ con $x_i \in \mathbf{N}$ e $X_i \subseteq \mathbf{N}$)) e una successione $\langle A_m\rangle_{m\in\mathbf{N}}$ di* assiomi provvisori *di S_d, come segue:*

$$s_0 := \langle (0, \{f(0)\})\rangle$$

dato $s_{n-1} := \langle (x_0, X_0), \ldots, (x_r, X_r)\rangle$, *e posti:*

$$k \;:=\; \begin{cases} \min i \leq r (c \in X_i), & \textit{se esiste un } i \leq r \textit{ tale che } c \in X_i \\ r+1, & \textit{altrimenti} \end{cases}$$

$$h \;:=\; \begin{cases} k, & \textit{se } k \leq r \\ r, & \textit{altrimenti} \end{cases}$$

Si definiscono:

$$A_{n-1} := \bigcup_{i<k} X_i$$

e:

$$s_n := \langle (x_0, K_n^h(X_0)), (x_1, K_n^h(X_0 \cup X_1), \ldots, (x_{k-1}, K_n^h(\bigcup_{i<k} X_i)), (x_h{+}1, \{f(x_h{+}1)\})\rangle$$

DEFINIZIONE 5.2.4 *Dato un sistema dialettico* $\mathsf{S}_d := \langle h, f, c\rangle$*, l'insieme* A_d *dei* teoremi definitivi *di* S_d *è definito da:*

$$A_d := \underline{\lim}_{n\to\infty} A_n = \{x \mid \bigvee k \bigwedge m \geq k (x \in A_m)\}$$

I due risultati seguenti, dimostrati da Magari[502], ci consentono di riconciliare le definizioni date con l'idea intuitiva di partenza, svelando, ad un tempo, il criterio che è all'origine della costruzione dei sistemi dialettici:

PROPOSIZIONE 5.2.5 *Dato un sistema dialettico* $\mathsf{S}_d = \langle h, f, c\rangle$ *e definiti come in 5.2.2* Φ_n^h, K_n^h, Φ^h*, il sistema* $\mathsf{S}^d = \langle \Phi^h, c\rangle$ *è un sistema formale (che chiameremo sistema formale associato a* S_d*).*

La dimostrazione si riduce ovviamente a far vedere che Φ^h è un operatore di chiusura algebrico che associa, a sottoinsiemi numerabili di $\mathbf{N}$, sottoinsiemi ancora enumerabili dello stesso insieme.

Questo fatto è all'origine anche del risultato seguente e del significato che è possibile attribuire ad esso:

PROPOSIZIONE 5.2.6 *Dato un sistema dialettico* $\mathsf{S}_d = \langle h, f, c\rangle$*, si consideri l'insieme* B_d *definito da:*

$$\begin{aligned} f(0) \in B_d &\iff c \notin \Phi^h(\{f(0)\}) \\ f(n) \in B_d &\iff c \notin \Phi^h(\{f(n)\} \cup (B_d \cap \{f(i) \mid i < n\})) \end{aligned}$$

(dove $n > 0$*). Vale:*

$$A_d = B_d$$

[502]Cfr. [Mag74], p. 125, Lemma 1 e pp. 126-127, Lemma 2, rispettivamente.

Sulla base di 5.2.5 siamo in un certo qual modo legittimati a vedere in h, come avevamo anticipato, la funzione di deduzione, essendo l'operatore Φ^h definito a partire da essa, effettivamente un operatore di chiusura algebrico su $\mathbf{N}$.

L'idea sulla quale si fonda la definizione della successione di stati provvisori di un sistema dialettico S_d e di quella degli insiemi di assiomi provvisori del sistema, può essere dunque ricostruita nel modo seguente: all'origine vi è una collezione di teoremi, rappresentata dall'insieme $\Phi^h(\emptyset)$ (e si noti che, sulla base di quanto stabilito al punto (3) della definizione 5.2.1, vale $\Phi^h(\emptyset) \neq \emptyset$); in relazione alle 'proposte' della funzione f, ossia di una scelta di possibili nuovi assiomi per il sistema di partenza, ad ogni stadio provvisorio se ne calcolano progressivamente alcune tra le conseguenze logiche (nel senso della definizione dei K_n^h di cui si fa uso), essendo il criterio guida, come dimostra in modo anche più immediato la definizione induttiva dell'insieme B_d di 5.2.6, quello di procedere escludendo di volta in volta gli enunciati contraddittori; si considerano poi, come tesi definitive del sistema, tutti e soli gli enunciati che appartengono in modo 'stabile' agli insiemi degli assiomi provvisori.

Il risultato 5.2.6, dunque, contribuisce, in connessione con 5.2.5, a chiarire la natura dei sistemi dialettici. Ma esso ne rivela altresì il principio di fondo, mostrando come un sistema dialettico non sia altro che una particolare istanza dei *completamenti ricorsivi alla Lindembaum* di una teoria formale S[503], i quali, in generale, sono teorie $\mathsf{T_S}$ definite, a partire da un'enumerazione ricorsiva $\{\phi_n \mid n \in \mathbf{N}\}$ di tutti gli enunciati del linguaggio di S, sulla base delle seguenti clausole:

$$\begin{aligned}
\mathsf{T}_0 &:= \mathsf{S} \\
\mathsf{T}_{n+1} &:= \begin{cases} \mathsf{T}_n \cup \{\phi_n\}, & \text{se } \mathsf{T}_n \cup \{\phi_n\} \text{ è consistente} \\ \mathsf{T}_n \cup \{\neg\phi_n\}, & \text{altrimenti} \end{cases} \\
\mathsf{T_S} &:= \bigcup_{n\in\mathbf{N}} \mathsf{T}_n
\end{aligned}$$

Proprio questo fatto, allora, permette di comprendere i motivi per i quali con i sistemi dialettici si riesce ad evitare certe limitazioni dei sistemi usua-

[503]Il risultato indicato nel testo, peraltro, non costituisce una sorpresa: segue infatti dal cosiddetto Teorema della Base di Kreisel (cfr. [Sho67], p. 187) e dai risultati indicati nel §5.1, che la complessità logica dei completamenti alla Lindembaum di una teoria formale (ovvero, la complessità di un ramo infinito nell'albero dei completamenti) è proprio Δ^0_2.

li, rendendo i risultati relativi, dei quali intendiamo adesso occuparci, non particolarmente sorprendenti[504].

5.2.1 Sistemi dialettici, teorie "meta-formali" e risultati limitativi

I sistemi dialettici di Magari vengono dunque definiti seguendo le indicazioni euristiche di partenza. Essi aspirano cioè a cogliere il processo sulla base del quale nuove ipotesi vengono aggiunte ad un corpo consolidato di conoscenze, in forma di tentativo, ed altre, quelle rivelatesi contraddittorie, sono invece eliminate.

Come è ovvio, il proposito di precisare formalmente un tale processo comporta delle scelte che sono opinabili e si presta, per questo motivo, a delle critiche. Piuttosto che soffermarci su queste ultime, crediamo sia meglio procedere a considerare il secondo aspetto della questione, ossia la valutazione dei teoremi limitativi rispetto al concetto individuato, che costituisce il presupposto per un ritorno all'argomento principale di questa parte del lavoro: la disamina della valenza filosofica in generale, e anti-meccanicista in particolare, dei teoremi di incompletezza.

Seguendo le indicazioni dello stesso Magari contenute nei due lavori di riferimento ([Mag74] e [Mag75]), riteniamo opportuno, anche per ottemperare

[504] I sistemi dialettici di Magari, in ogni caso, non sono l'unico esempio di nozione scaturita dall'analisi delle procedure per tentativi ed errori. Negli stessi anni, ma in modo del tutto indipendente dall'algebrista italiano, Jeroslow ha considerato, in [Jer75], il concetto di Logica Sperimentale, definito a partire da una relazione ternaria $H(t, x, y)$ che si legge: "all'istante t, y è una dimostrazione riconosciuta di x". I *teoremi* di una logica speritmentale sono identificati con le formule *ricorrenti*, definite da:

$$Rec_H(x) := \bigwedge t \bigvee s \geq t \bigvee y(H(s, x, y))$$

Una condizione ulteriore è quella delle formule *stabili* (legata al concetto di tesi definitiva del caso di Magari), definita da:

$$Stbl_H(x) := \bigvee t \bigvee y \bigwedge s \geq t(H(s, x, y))$$

L'analisi si concentra sulle logiche sperimentali *convergenti*, per le quali vale $Rec_H(x) \Rightarrow Stbl_H(x)$, e che si dimostrano coincidere, mediante un'applicazione del risultato di Putnam citato, con gli insiemi Δ^0_2. Riformulazioni dei risultati limitativi vengono date e analizzate. Un ulteriore approfondimento e raffinamento del tema, è quello di Hájek in [Háj77], che indaga certe caratterizzazioni degli insiemi Π^0_2 e Π^0_3 basati sulle logiche sperimentali.

ad una completezza illustrativa che ci pare doverosa, suddividere la trattazione del rapporto con i risultati limitativi considerando, da un lato, i teoremi di Gödel relativi ai limiti *deduttivi* dei sistemi formali, dall'altro, il risultato di Tarski sull'indefinibilità aritmetica del predicato di verità, che mostra un'essenziale carenza *espressiva* delle teorie assiomatiche formali.

Trattandosi di questioni la cui controparte tecnica viene ampiamente ed esaurientemente discussa da Magari, ci limiteremo, ove necessario, a richiamare le nozioni e le dimostrazioni principali (senza riprodurle nel dettaglio, talvolta, anche per ragioni di concisione), privilegiando, invece, l'analisi di quegli aspetti che favoriscono una più proficua riflessione nel quadro delle tematiche all'attenzione del presente lavoro.

I sistemi dialettici e l'incompletezza sintattica

Come si è anticipato in chiusura del §5.2, tenendo conto del risultato 5.2.6, quei corollari che permettano di apprezzare come le limitazioni sancite dai teoremi di incompletezza non valgano per i sistemi dialettici sono, in un certo senso, banali. Il modo con cui ciò avvenga precisamente, lo si può brevemente illustrare come segue.

Innanzi tutto, da 5.2.6 si ricavano in modo immediato i due corollari seguenti:

COROLLARIO 5.2.7 *Se il sistema $\mathsf{S}^d := \langle \Phi^h, c \rangle$ associato ad un sistema dialettico S_d è consistente (ossia se vale $\Phi^h(\emptyset) \neq \mathbf{N}$), allora $\Phi^h(\emptyset) \subseteq A_d$ (cioè i teoremi di S^d sono teoremi definitivi di S_d).*

Dim.: Se S^d è consistente, allora $c \notin \Phi^h(\emptyset)$, da cui segue, per definizione di B_d di 5.2.6, $\Phi^h(\emptyset) \subseteq B_d$. Il risultato si ricava allora dalla proposizione 5.2.6. ∎

COROLLARIO 5.2.8 *Se il sistema $\mathsf{S}^d := \langle \Phi^h, c \rangle$ associato a S_d è consistente, allora A_d è un'estensione non contraddittoria ($c \notin A_d$) e massimale di $\Phi^h(\emptyset)$.*

Dim.: Immediato, dalla definizione di B_d e da 5.2.6. ∎

Si ponga inoltre:

DEFINIZIONE 5.2.9 *Un sistema dialettico $\mathsf{S}_d := \langle h, f, c \rangle$ è consistente se e solo se $A_d \neq \emptyset$.*

Vale, allora:

COROLLARIO 5.2.10 *Un sistema dialettico è consistente se e solo se lo è il sistema formale ad esso associato.*

Dim.: Dato un sistema dialettico $\mathsf{S}_d := \langle h, f, c\rangle$, se $\Phi^h(\emptyset) = \mathbf{N}$, si ha, in particolare che, se $c \in \Phi^h(\emptyset)$, segue, ovviamente, $B_d = \emptyset$.
Viceversa, se $\Phi^h(\emptyset) \neq \mathbf{N}$, considerando che dal punto (3) della definizione 5.2.1 si ha che $\Phi^h(\emptyset) \neq \emptyset$ e che, per il corollario 5.2.7, $\Phi^h(\emptyset) \subseteq A_d$, vale $A_d \neq \emptyset$. ■

Per apprezzare i risultati che sanciscono il fatto che i sistemi dialettici non cadono sotto le usuali limitazioni deduttive dei sistemi formali d'assiomi, occorre, innanzi tutto, arricchire il linguaggio fin qui usato. Infatti, si sarà forse notato come non si sia specificato affatto se un sistema dialettico sia dotato o meno degli usuali operatori logici.

In questo contesto, ciò significa che, dato un sistema dialettico $\mathsf{S}_d = \langle h, f, c\rangle$, esistono due funzioni primitive ricorsive, $\neg x$ e $x \vee y$, tali che valgono, per ogni $x, y \in \mathbf{N}$ e $X \subseteq \mathbf{N}$, le seguenti condizioni (che corrispondono ad usuali proprietà dei connettivi logici):

$$\begin{aligned}
\Phi^h(\neg\neg x) &= \Phi^h(x) && (5.1)\\
c &\in \Phi^h(x, \neg x) && (5.2)\\
x \vee \neg x &\in \Phi^h(\emptyset) && (5.3)\\
\Phi^h(X \cup \{x \vee y\}) &= \Phi^h(X \cup \{x\}) \cap \Phi^h(X \cup \{y\}) && (5.4)
\end{aligned}$$

Si verifica facilmente che vale:

COROLLARIO 5.2.11 *Sia $\mathsf{S}_d = \langle h, f, c\rangle$ un sistema dialettico siffatto, che sia, in più, consistente. Allora, per ogni $x \in \mathbf{N}$, $x \in A_d$ se e solo se $\neg x \notin A_d$.*

Dim.: Tenendo presente 5.2.6 e (5.2), sia ha che, per ogni $x \in \mathbf{N}$, $x \in B_d$ e $\neg x \in B_d$ si escludono vicendevolmente.
Supponiamo adesso che si abbia $x \notin A_d$ e $\neg x \notin A_d$. Sulla base di 5.2.6, ciò significa che, dati $m, n \in \mathbf{N}$ tali che $f(m) = x$ e $f(n) = \neg x$, per qualche $M \subseteq (A_d \cap \{f(i) \mid i < m\})$ e $N \subseteq (A_d \cap \{f(j) \mid j < n\})$, valgono:

$$\begin{aligned}
c &\in \Phi^h(\{f(m)\} \cup M)\\
c &\in \Phi^h(\{f(n)\} \cup N)
\end{aligned}$$

Dal ché segue, per $P = M \cup N$:

$$c \in (P \cup \{f(m)\}) \cap (P \cup \{f(n)\}) \stackrel{(5.4)}{=} (P \cup \{f(m) \vee f(n)\}) \stackrel{(5.3)}{\subseteq} A_d$$

Ma ciò contraddice quanto sancisce il corollario 5.2.8. Dunque, vale $x \in A_d$ o $\neg x \in A_d$. ■

Altrettanto semplice, per quanto possa apparire artificioso, è il seguente:

COROLLARIO 5.2.12 *Dato un sistema formale* $\mathsf{S} = \langle \Phi, c \rangle$ *consistente e sufficientemente potente, è possibile costruire un sistema dialettico* $\mathsf{S}_d = \langle h, f, c \rangle$ *tale che* $\mathsf{S} = \mathsf{S}^d$ *(ossia,* S *è il sistema associato a* S_d*) e, se* $Con_{\mathsf{S}} \in \mathbf{N}$ *è (il codice de) l'enunciato che esprime la consistenza di* S*,* $Con_{\mathsf{S}} \in A_d$*.*

Dim.: Si noti, innanzi tutto, che dato un sistema formale S usuale, è sempre possibile costruire un sistema dialettico S_d tale che $\mathsf{S} = \mathsf{S}^d$ valga.
Se infatti $\mathsf{S} = \langle \Phi, c \rangle$, si ha che esiste una procedura effettiva h' che associa ad ogni insieme ricorsivamente enumerabile $A \subseteq \mathbf{N}$ (o meglio all'indice della funzione che lo enumera), una procedura effettiva di enumerazione dell'insieme $\Phi(A)$. Dunque, in particolare, si ha che esiste una h che associa all'indice canonico di ogni insieme finito $X \subseteq \mathbf{N}$, una funzione che enumera $\Phi(X)$. Scelta allora una biezione ricorsiva f, si ha che $\mathsf{S}_d := \langle h, f, c \rangle$ è un sistema dialettico e che vale $\mathsf{S} = \mathsf{S}^d$ [505].
Supponiamo, in più, di scegliere la f in modo tale che $f(0) = Con_{\mathsf{S}}$. Dal secondo teorema di Gödel si ha che $\neg Con_{\mathsf{S}} \notin \Phi(\emptyset)$. Inoltre, dal momento che $f(0) = Con_{\mathsf{S}}$, vale anche (per (5.2)) che, dato $m \in \mathbf{N}$ tale che $f(m) = \neg Con_{\mathsf{S}}$:

$$c \in \Phi(\{f(m)\} \cup (B_d \cap \{f(i) \mid i < m\}))$$

Ergo, per 5.2.6, $\neg Con_{\mathsf{S}} \notin A_d$, dal ché, per il corollario 5.2.11, $Con_{\mathsf{S}} \in A_d$. ■

[505] Che si dimostri, sulla base delle definizioni, che per ogni $Z \subseteq \mathbf{N}$:

$$\Phi(Z) = \bigcup_{n \in \mathbf{N}} \{K_n(X) \mid X \subseteq Z \ \& \ |X| < \aleph_0\} =: \Phi^h(Z)$$

segue da note proprietà logiche.

Tenendo presente il risultato 5.2.10, sulla base del quale un sistema dialettico è consistente se e solo se lo è il sistema formale associato, la proposizione Con_{S}, che esprime la consistenza di un sistema formale S, esprime anche la consistenza di un sistema dialettico S_d per il quale si abbia $\mathsf{S} = \mathsf{S}^d$. Dunque, sotto le ipotesi del risultato precedente, un sistema dialettico consistente è in grado di dimostrare la propria consistenza[506].

Una teoria non tarskiana della verità

Si è appena visto che passando da teorie assiomatiche usuali, rispetto alla quale l'insieme dei teoremi è ricorsivamente enumerabile (cioè Σ_1^0), a sistemi il cui insieme di teoremi è, invece, un limite inferiore (e quindi Σ_2^0), si riesce ad ovviare, in una qualche misura, alle limitazioni deduttive dei primi. Da ciò, si potrebbe trarre l'idea che, mediante una strategia analoga, si possa ovviare anche alle limitazioni espressive che le teorie formali presentano.

Sappiamo infatti che, considerato il sistema per l'aritmetica di Peano PA, non è possibile definire, nel linguaggio sul quale esso è basato, un predicato $T(x)$ tale che si abbia, per ogni formula ϕ:

$$\models T(\ulcorner \phi \urcorner) \Leftrightarrow \models \phi$$

(dove $\models \phi$ è un'abbreviazione per $\phi \in \mathbf{V}$, e $\mathbf{V}$ è l'insieme delle formule di PA vere, sotto l'interpretazione intesa dei termini, in senso tarskiano).

Una via per risolvere il problema, può essere quella di riformulare le nozioni in gioco (sulla base dell'idea che il concetto che isoliamo mediante la collezione $\mathbf{V}$, per quanto 'naturale', sia fin troppo esteso), il ché conduce a chiedersi se è possibile rispondere affermativamente al seguente quesito: è possibile, indicato con P_{PA} l'insieme delle formule di PA, isolare un $S \subseteq P_{\mathsf{PA}}$ (l'insieme delle proposizioni *sensate*) rispetto al quale definire una collezione V^S (delle proposizioni S-vere) ed un predicato $V^S(x)$ in PA, tale che valga, per ogni $\phi \in S$, $\models V^S(\ulcorner \phi \urcorner)$ se e solo se $\models \phi$?

[506] Il richiamo al corollario 5.2.10 potrebbe spingere a vedere l'interesse del risultato conseguito, come limitato: in fin dei conti si mostra la dimostrabilità di un enunciato che è al più equivalente a quello che esprime la consistenza di un sistema dialettico. Magari, consapevole di ciò, offre ([Mag74], pp. 141 e segg.) un raffinamento del risultato precedente, mostrando come sia possibile costruire un sistema dialettico che dimostra un enunciato che ne esprime, in modo "naturale", la consistenza. In questa sede, abbiamo tuttavia preferito soprassedere su quest'analisi ulteriore della questione, per privilegiare, nel seguito, quanto ci pare più significativo dire sui teoremi limitativi rispetto ai sistemi dialettici nel contesto del presente lavoro.

È chiaro che una soluzione la si ha già rispetto alle usuali teorie assiomatiche, considerando che, dato $\mathbf{T}_{\mathsf{PA}} := \{\phi \mid \mathsf{PA} \vdash \phi\}$ e posto $S := \mathbf{T}_{\mathsf{PA}} \cup \{\neg\phi \mid \phi \in \mathbf{T}_{\mathsf{PA}}\} \subseteq P_{\mathsf{PA}}$, si ha che, per ogni $\phi \in S$, vale (supposto $\mathbf{T}_{\mathsf{PA}} \subseteq \mathbf{V}$)[507]:

$$\begin{array}{ccc} \mathsf{PA} \vdash \phi & \overset{(1)}{\Leftrightarrow} & \mathsf{PA} \vdash Teor_{\mathsf{PA}}(\ulcorner\phi\urcorner) \\ {\scriptstyle(2)}\Updownarrow & & \Updownarrow{\scriptstyle(4)} \\ \models Teor_{\mathsf{PA}}(\ulcorner\phi\urcorner) & \overset{(3)}{\Leftrightarrow} & \models \phi \end{array}$$

È dunque possibile, facendo coincidere $\mathbf{T}_{\mathsf{PA}}$ con l'insieme delle formule S-vere, ovviare alle ordinare limitazioni espressive dei sistemi d'assiomi.

In questo caso, si avrebbe una teoria formale $\langle P_{\mathsf{PA}}, S, V^S\rangle$ con S, V^S insiemi ricorsivamente enumerabili. Tuttavia, non sembrano esserci motivazioni indipendenti, che non si riducano, cioè, banalmente all'idea di voler ovviare alle limitazioni suddette per questa via *ad hoc*, per giustificare una simile scelta.

Magari si propone, in [Mag75], di costruire invece una teoria $\langle P_{\mathsf{PA}}, S_0, V_0^S\rangle$, che egli chiama, per estensione della terminologia precedente, "meta-formale", nella quale si abbia, invece, $S_0, V_0^S \in \Sigma_2^0$ (nella quale, cioè, l'insieme delle proposizioni sensate e di quelle sensatamente vere, sono limiti inferiori[508]), e rispetto alla quale, soprattutto, si possano proporre degli argomenti fondati per la scelta di tali S_0 e V_0^S.

La definizione dei due insiemi è la seguente:

DEFINIZIONE 5.2.13 (1) *$V_0^S \subseteq \mathbf{N}$ è il più piccolo insieme tale che (per ogni $\phi \in P_{\mathsf{PA}}$):*

---

[507] L'equivalenza (1), segue dalle note proprietà del predicato $Teor_{\mathsf{PA}}$, avendo qui supposto la correttezza di PA. (2) si ricava da (1) e dall'ipotesi $\mathbf{T}_{\mathsf{PA}} \subseteq \mathbf{V}$, in una direzione (soprasotto); nell'altra, è immediata. Per quel che riguarda (3), essa segue ancora dall'ipotesi di correttezza in una direzione: $\models Teor_{\mathsf{PA}}(\ulcorner\phi\urcorner) \Rightarrow \mathsf{PA} \vdash \phi \Rightarrow \models \phi$. Se invece vale $\models \phi$, si ha che, poiché $\mathbf{T}_{\mathsf{PA}} \subseteq \mathbf{V}$, $\mathsf{PA} \nvdash \neg\phi$, da cui, essendo $\phi \in S$, deve essere $\mathsf{PA} \vdash \phi$; allora, da (1) più l'ipotesi di correttezza, $\models Teor_{\mathsf{PA}}(\ulcorner\phi\urcorner)$. Infine, (4) è un'ovvia conseguenza delle altre equivalenze.

[508] Implicitamente, siamo così passati da considerare insiemi di proposizioni, ai corrispondenti insiemi dei codici (attraverso l'assunzione nascosta dell'esistenza di una procedura di aritmetizzazione standard del linguaggio formale). Nel caso generale, una terna $\langle P, S, V\rangle$ è una teoria meta-formale se e solo se (i) P è un insieme decidibile di proposizioni, (ii) c'è una corrispondenza biunivoca computabile da $S, V \subseteq P$ in due limiti inferiori. Nel seguito, lavoreremo direttamente, come si è fatto nel paragrafo precedente, con il caso in cui $P = \mathbf{N}$.

(i) $\mathbf{T}_{\mathsf{PA}} \subseteq V_0^S$;

(ii) *se* $\phi \in V_0^S$ *e* $\phi \to \psi \in V_0^S$, *allora* $\psi \in V_0^S$;

(iii) *se* $\neg\phi \to Teor_{\mathsf{PA}}(\ulcorner\neg\phi\urcorner) \in V_0^S$ *e* $\neg\phi \notin \mathbf{T}_{\mathsf{PA}}$, *allora* $\phi \in V_0^S$.

(2) $S_0 = V_0^S \cup \{\neg\phi \mid \phi \in V_0^S\}$.

Vale la pena soffermarci brevemente sull'idea che sta alla base di una tale proposta.

È intanto ovvio come, nella definizione di V_0^S, la clausola fondamentale sia la seconda. Per comprenderne a fondo il significato, si può riprendere la terna $\langle P_{\mathsf{PA}}, S, \mathbf{T}_{\mathsf{PA}}\rangle$, dove $S = \mathbf{T}_{PA} \cup \{\neg\phi \mid \phi \in \mathbf{T}_{\mathsf{PA}}\}$, e considerare come, tra i motivi di maggior rilievo per scartare una simile definizione, ci sia il fatto che, non solo le proposizioni indecidibili in PA risultano essere ovviamente insensate, ma che, tra queste, lo sono anche quante possiedono, in PA, un 'criterio di falsificazione'.

Risultano infatti insensate proposizioni ϕ per le quali si abbia:

$$\models \phi,\ \mathsf{PA} \nvdash \neg\phi,\ \mathsf{PA} \vdash \neg\phi \to Teor_{\mathsf{PA}}(\ulcorner\neg\phi\urcorner)$$

È il caso, ad esempio, dell'enunciato indecidibile di Gödel. Tuttavia, in una situazione simile, si potrebbe contestare l'insensatezza di ϕ: in fin dei conti, per quanto PA non refuti di fatto ϕ, la stessa teoria ci rende consapevoli del fatto che, se $\neg\phi$ fosse vera, essa sarebbe dimostrabile in PA e, di conseguenza, ϕ risulterebbe falsificabile (per quanto non falsificata, disponiamo cioè di un 'metodo' per falsificarla).

Questa è l'idea cruciale alla quale Magari ricorre per estendere $\mathbf{T}_{\mathsf{PA}}$, inteso come collezione di enunciati verificabilmente (cioè dimostrabilmente) veri (sotto l'ipotesi di correttezza dell'aritmetica di Peano).

A monte[509], c'è l'idea che la metamatematica debba essere considerata, nei confronti delle teorie assiomatiche, una 'scienza', nello stesso senso in cui la fisica, ad esempio, è vista come la scienza dei fenomeni naturali, e che, quindi, sia lecito adottare, nel caso della prima, quei criteri che, in ambito teoretico, sono stati proposti per discipline come la seconda (è il caso del criterio 'popperiano' - una tesi è sensata se falsificabile - utilizzato per la definizione di V_0^S).

[509]Si vedano in particolare, di [Mag75], le pp. 343-345.

Il risultato fondamentale che, rispetto a S_0 e V_0^S, Magari presenta[510], dal quale discende che $\langle P_{\mathsf{PA}}, S_0, V_0^S\rangle$ è una teoria meta-formale che estende PA (vale, infatti, $\mathbf{T}_{\mathsf{PA}} \subseteq V_0^S$ per definizione), è la seguente:

PROPOSIZIONE 5.2.14 *S_0 e V_0^S sono limiti inferiori.*

Nel corso della dimostrazione di 5.2.14, si isola in particolare una relazione primitiva ricorsiva ternaria $R_0(x,y,z)$, tale che valgono:

$$\begin{array}{rcl} S_0 & := & \{\phi \mid \bigvee x[\bigvee y \bigwedge z \geq y(R_0(\ulcorner\phi\urcorner, x, z)) \vee \bigvee y \bigwedge z \geq y(R_0(\ulcorner\neg\phi\urcorner, x, z))]\} \\ V_0^S & := & \{\phi \mid \bigvee x \bigvee y \bigwedge z \geq y(R_0(\ulcorner\phi\urcorner, x, z))\} \end{array}$$

Sfruttando, quindi, le usuali proprietà di rappresentabilità in PA delle relazioni primitive ricorsive (e dimostrando alcuni lemmi supplementari[511]), si ricava[512] (supponendo di scrivere $|\models \phi$ per $\phi \in V_0^S$), la seguente:

PROPOSIZIONE 5.2.15 (1) *Esiste una formula $\sigma_0 \in P_{\mathsf{PA}}$ tale che, per ogni $\phi \in P_{\mathsf{PA}}$:*

$$\phi \in S_0 \Leftrightarrow |\models \sigma_0(\ulcorner\phi\urcorner) \Leftrightarrow \models \sigma_0(\ulcorner\phi\urcorner)$$

(2) *Esiste una formula $\theta_0 \in P_{\mathsf{PA}}$ tale che, per ogni $\phi \in P_{\mathsf{PA}}$:*

$$|\models \phi \Leftrightarrow |\models \theta_0(\ulcorner\phi\urcorner) \Leftrightarrow \models \theta_0(\ulcorner\phi\urcorner)$$

Dal risultato discende, allora, la possibilità di ovviare, con la meta-teoria $\langle P_{\mathsf{PA}}, S_0, V_0^S\rangle$, alle limitazioni espressive delle usuali teorie assiomatiche (nel senso specificato e sulla base delle considerazioni fatte, in particolare quelle che intendono giustificare la restrizione al sottoinsieme di $\mathbf{V}$ rappresentato da V_0^S[513]).

5.2.2 Limitazioni dei sistemi dialettici e delle teorie meta-formali

Nonostante i risultati testé illustrati, non possiamo certo aspettarci di essere in grado di evitare, se non nel senso tutto particolare legato ad alcune osservazioni fatte in precedenza, la possibilità di riprodurre, anche per le

[510] Cfr. *Ibid.*, pp. 354-355, Lemma 1.
[511] Cfr. *Ibid.*, pp. 355-358, in particolare Lemma 3, (ii).
[512] *Ibid.*, p. 358.
[513] Come risulta dalla proposizione 5.2.15.

nuove cornici formali che siamo venuti considerando, gli stessi argomenti che conducono ad indicare le limitazioni delle usuali teorie assiomatiche (anche se, come vedremo, con delle novità decisive rispetto ad un certo uso che, di questi risultati, si è cercato di fare).

Il teorema noto, che palesa in modo inequivocabile le carenze degli accorgimenti adottati è il seguente[514]:

TEOREMA 5.2.16 *La collezione* $\mathbf{V}$ *delle proposizioni aritmetiche vere, è ricorsivamente isomorfa a* $\emptyset^{(\omega)}$*, ossia esiste una biezione ricorsiva* f *tale che,* $f(\mathbf{V}) = \emptyset^{(\omega)}$*.*

Segue quindi[515], che $\mathbf{V}$ è Turing riducibile (anzi, Turing equivalente) a $\emptyset^{(\omega)}$ e, dunque, per ogni $n \in \mathbf{N}$, $\mathbf{V} \not\leq_T \emptyset^{(n)}$. Sulla base di 5.1.4, 1 e 2, e delle rispettive conseguenze, concludiamo quindi che $\mathbf{V}$ non può essere un limite e, *a fortiori*, neanche l'insieme delle tesi definitive di un sistema dialettico[516].

Queste considerazioni, che di per sé sono già sufficienti per poter parlare di limitazioni per i sistemi dialettici, possono anche essere rese con maggior precisione.

Posto infatti:

DEFINIZIONE 5.2.17 *Un insieme* $A \subseteq \mathbf{N}$ *è completamente produttivo se e solo se esiste una funzione ricorsiva (totale)* f *tale che, per ogni* $n \in \mathbf{N}$*,* $f(n) \in A \dot{+} W_n = (A \setminus W_n) \cup (W_n \setminus A)$*.*

è noto che si dimostra[517] che vale:

PROPOSIZIONE 5.2.18 $\mathbf{V}$ *è completamente produttivo.*

---

[514]Cfr. [Rog67], p. 318.

[515]Si dimostra infatti (cfr. *ibid.*, p. 85, Teorema VI) che due insiemi $A, B \subseteq \mathbf{N}$ sono ricorsivamente isomorfi se e solo se sono 1-equivalenti (ossia, se e solo se vale $A \leq_1 B$ & $B \leq_1 A$).

[516]Si ricava cioè che $\mathbf{V}$ è un insieme iperaritmetico che si situa al primo livello transfinito della gerarchia del salto.

[517]Cfr., ad esempio, [Rog67], p. 97. Più precisamente, si dimostra che $\mathbf{V}$ è *produttivo*, ossia che esiste una funzione ricorsiva parziale (cioè un $e \in \mathbf{N}$), tale che:

$$\bigwedge x[W_x \subset \mathbf{V} \Rightarrow \bigvee y(T_1(e,x,y) \; \& \; \varphi_e(x) \in \mathbf{V} \dot{+} W_x)]$$

e si sfrutta poi, per giungere alla conclusione indicata nel testo, l'equivalenza tra la produttività e la completa produttività di un insieme (*ibid.*, p. 183, Teorema VI).

Ciò implica in particolare, che esiste una procedura effettiva che in corrispondenza ad ogni sistema formale S (identificando un sistema con la chiusura deduttiva di un insieme ricorsivamente enumerabile di assiomi), associa un enunciato che, se il sistema è consistente, è vero, ma indimostrabile in S.

Risultati che, come questo, permettono di esprimere l'essenza dei teoremi di incompletezza in modo così inequivocabile, sembrano essere particolarmente adatti per offrire un sostegno a quegli argomenti che intendono affermare una superiorità del matematico umano sulle macchine: esiste un metodo effettivo ed uniforme che ci permette di esibire un enunciato che sappiamo essere vero, ma anche indimostrabilmente tale mediante una procedura effettiva di computo.

Si potrebbe dunque ritenere un fatto decisamente negativo per gli scopi che ci siamo prefissi, la possibilità di riproporre un risultato analogo, sotto un'appropriata riformulazione delle nozioni in gioco, rispetto ai sistemi dialettici.

Supponiamo di definire un insieme $A \subseteq \mathbf{N}$ *completamente limite-produttivo*, se (indicato con λx il limite della successione degli insiemi finiti generati dalla funzione indiciata da x[518]), esiste una funzione ricorsiva f tale che $f(x) \in A \dot{+} \lambda x$[519].

L'insieme A sarebbe dunque tale che è possibile esibire in modo effettivo un elemento che lo distingue da ogni limite.

Magari dimostra[520] che, se $A, B \subseteq \mathbf{N}$ sono tali che A è ricorsivo e $A \cap B$ completamente limite-produttivo, allora anche B è completamente limite produttivo.

Esaminando poi l'insieme M degli indici di funzioni ricorsive totali che non appartengono al limite da essi generato, ossia, nella notazione adottata, la collezione definita da:

$$x \in M \Leftrightarrow \begin{array}{c} \bigwedge y \bigvee z (T_1(x, y, z)) \ \& \\ \underline{\lim}_{n\to\infty} D_{\varphi_x(n)} = \overline{\lim}_{n\to\infty} D_{\varphi_x(n)} \ \& \\ x \notin \lambda x \end{array}$$

si osserva che, per ogni $x \in \mathbf{N}$ tale che $\underline{\lim}_{n\to\infty} D_{\varphi_x(n)} = \overline{\lim}_{n\to\infty} D_{\varphi_x(n)}$, vale:

$$x \in \lambda x \Leftrightarrow x \notin M$$

518 Nella notazione adottata, $\lambda x = \lim_{n\to\infty} D_{\varphi_x(n)}$.

519 Ovviamente, si sottintende che ciò valga per ogni $x \in \mathbf{N}$ che 'ammette limite', ossia per il quale si abbia $\underline{\lim}_{n\to\infty} D_{\varphi_x(n)} = \overline{\lim}_{n\to\infty} D_{\varphi_x(n)}$.

520 [Mag74], p. 140.

(ergo $x \in M \dot{+} \lambda x$ e, quindi, M è completamente limite-produttivo).

Vale inoltre chiaramente[521], che esiste un predicato aritmetico $P(x)$, ossia, in sostanza, un'espressione unaria del linguaggio del sistema PA, tale che, per ogni $n \in \mathbf{N}$:

$$P(\overline{n}) \in \mathbf{V} \Leftrightarrow n \in M \tag{5.5}$$

Considerando poi che $A = \{P(\overline{n}) \mid n \in \mathbf{N}\}$ è decidibile e che $\mathbf{V} \cap A$ risulta essere, valendo (5.5), ricorsivamente isomorfo a M, Magari ricava infine, sulla base di certi risultati che permettono di associare limiti a sistemi dialettici in modo effettivo ed uniforme[522], il seguente:

TEOREMA 5.2.19 *Esiste una procedura effettiva che dato un qualsiasi sistema dialettico S_d che contenga, tra le proprie proposizioni, quelle dell'aritmetica di Peano, fornisce una proposizione che è vera se e solo se non appartiene all'insieme A_d delle tesi definitive di S_d.*

Un simile risultato, allora, ci porterebbe a riconoscere come, non solo, sulla base delle osservazioni precedenti, non siamo in grado, passando dalle usuali teorie assiomatiche formali ai sistemi dialettici, ad ovviare a quell'aspetto che rende rilevanti i risultati di indecidibilità, l'esistenza di una differenza essenziale tra la verità delle proposizioni della teoria dei numeri, e la loro verificabilità effettiva, ma non si riesce neanche ad evitare la possibilità di formulare quegli stessi risultati in quella forma che ha dato adito all'idea che, in corrispondenza del primo fatto (la discrasia tra verità e dimostrabilità formale), si possano individuare gli estremi per una differenza, parimenti essenziale, tra le capacità di un qualsiasi matematico umano e quelle di una macchina.

Come se ciò non bastasse a creare il sospetto che si possa anche in questo contesto avanzare certe conseguenze dei teoremi limitativi, è possibile giungere a conclusioni del tutto analoghe anche percorrendo la strada delle teorie meta-formali, ossia ridefinendo, sulla base di certe considerazioni di adeguatezza ad una concezione di verità alternativa a quella più tradizionale, la collezione delle proposizioni vere (o meglio sensatamente tali) dell'aritmetica.

Se il risultato 5.2.16 citato in apertura è di per sé sufficiente a concludere che l'insieme $\mathbf{V} \setminus V_0^S$ non è vuoto, questo stesso fatto può essere rivelato mediante una costruzione che ricalca da vicino l'usuale metodologia di prova

[521] Si ricordi che l'indice caratteristico di un insieme finito $A = \{n_1, \ldots, n_k\} \subseteq \mathbf{N}$, è $n = 2^{n_1} + \ldots + 2^{n_k}$.

[522] Cfr., in particolare, [Mag74], p. 131, Teorema 2.

dei risultati di Gödel. In più, si dimostra però che gli enunciati ottenuti in questo modo sono insensati, confermando quindi, allo stesso tempo, la valenza delle definizioni date.

Analizzando le proprietà basilari degli insiemi S_0 e V_0^S in relazione alla possibilità di fornirne, grazie alle formule $\sigma_0(x)$ e $\theta_0(x)$ di 5.2.15, una versione aritmetizzata, Magari dimostra[523], in particolare, che vale la seguente[524] (si ricordi che $|\models \alpha$ sta per $\alpha \in V_0^S$):

PROPOSIZIONE 5.2.20 (i) *Se* PA *è consistente, valgono, per ogni* $\phi, \psi \in P_{\mathsf{PA}}$*:*

$$|\models \theta_0(\ulcorner \phi \to \psi \urcorner) \to (\theta_0(\ulcorner \phi \urcorner) \to \theta_0(\ulcorner \psi \urcorner)) \tag{5.6}$$
$$\text{Se } |\models \theta_0(\ulcorner \phi \to \psi \urcorner) \text{ allora } |\models \theta_0(\ulcorner \phi \urcorner) \to \theta_0(\ulcorner \psi \urcorner) \tag{5.7}$$
$$|\models \theta_0(\ulcorner \phi \wedge \psi \urcorner) \leftrightarrow \theta_0(\ulcorner \phi \urcorner) \wedge \theta_0(\ulcorner \psi \urcorner) \tag{5.8}$$
$$|\models \theta_0(\ulcorner \phi \urcorner) \to \theta_0(\ulcorner \theta_0(\ulcorner \phi \urcorner) \urcorner) \tag{5.9}$$
$$\text{se } |\models \phi \text{ allora } |\models \theta_0(\ulcorner \phi \urcorner) \tag{5.10}$$

(ii) *Se* PA *è* ω*-consistente, allora:*

$$\text{se } |\models \theta_0(\ulcorner \phi \urcorner) \text{ allora } |\models \phi \tag{5.11}$$

Si possono così riformulare, rispetto alla teoria meta-formale $\langle P_{\mathsf{PA}}, S_0, V_0^S \rangle$, i due risultati gödeliani:

TEOREMA 5.2.21 *Se* PA *è* ω*-consistente, esistono proposizioni vere e insensate.*

Dim.: Sia ϕ l'enunciato rispetto al quale vale, sulla base del lemma di diagonalizzazione:

$$\mathsf{PA} \vdash \phi \leftrightarrow \neg\theta_0(\ulcorner \phi \urcorner) \tag{5.12}$$

Se valesse $\phi \in V_0^S$, si avrebbe anche, per (5.10), $\theta_0(\ulcorner \phi \urcorner) \in V_0^S$ da cui, per (5.12), $\neg\phi \in V_0^S$, il ché è assurdo considerato che vale $V_0^S \subseteq \mathbf{V}$. Se invece fosse $\neg\phi \in V_0^S$, si avrebbe, per (5.12), $\theta_0(\ulcorner \phi \urcorner) \in V_0^S$ dal ché si potrebbe concludere, tenuto conto di (5.11), $\phi \in V_0^S$.

[523] Cfr. [Mag75], p. 364, Teorema 10.

[524] Alcuni enunciati sono una semplice conseguenza di risultati che abbiamo già citato, ma che si è detto valere sotto l'assunzione $\mathbf{T}_{\mathsf{PA}} \subseteq \mathbf{V}$. La proposizione seguente, indica dunque come è possibile indebolire la suddetta ipotesi.

Ergo, $\phi \notin S_0$ e $\phi, \neg\phi \notin V_0^S$, da cui si ricava in particolare, sulla base di 5.2.15 (2), che vale $\neg\theta_0(\ulcorner\phi\urcorner) \in \mathbf{V}$ e quindi, ancora per (5.12), $\phi \in \mathbf{V} \setminus V_0^S$. ■

TEOREMA 5.2.22 *Assumiamo* $\mathbf{T}_{\mathsf{PA}} \subseteq \mathbf{V}$. *Ogni formulazione* χ *della consistenza di* $\langle P_{\mathsf{PA}}, S_0, V_0^S\rangle$ *tale che* $|\models \chi \to \neg\theta_0(\ulcorner 0 = 1\urcorner)$ *valga, è insensata.*

Dim.: Sia ϕ l'enunciato che, come nel caso del teorema precedente, verifica la (5.12). Da questa, tenendo presente l'ipotesi di correttezza, e dalla (5.9), si ricava per logica $|\models \theta_0(\ulcorner\neg\phi\urcorner) \leftrightarrow \theta_0(\ulcorner\theta_0(\ulcorner\phi\urcorner)\urcorner)$. Ancora dalla (5.9), vale allora:

$$|\models \theta_0(\ulcorner\phi\urcorner) \to \theta_0(\ulcorner\neg\phi\urcorner) \tag{5.13}$$

Sia χ come da ipotesi. Se valesse $\chi \in V_0^S$, si avrebbe anche $\neg\theta_0(\ulcorner 0 = 1\urcorner) \in V_0^S$, e quindi (sempre ricordando l'assunzione $\mathbf{T}_{\mathsf{PA}} \subseteq \mathbf{V}$), $\neg\theta_0(\ulcorner\phi \wedge \neg\phi\urcorner) \in V_0^S$. Ma allora, poiché vale (5.8), varrebbe anche:

$$|\models \neg(\theta_0(\ulcorner\phi\urcorner) \wedge \theta_0(\ulcorner\neg\phi\urcorner))$$

Per logica, segue allora $\theta_0(\ulcorner\phi\urcorner) \to \neg\theta_0(\ulcorner\neg\phi\urcorner) \in V_0^S$ da cui, per (5.13), $\neg\theta_0(\ulcorner\phi\urcorner) \in V_0^S$. Da ciò si concluderebbe allora, per (5.12), $\phi \in V_0^S$, che abbiamo visto essere assurdo.
Tenendo presente che $V_0^S \subset \mathbf{V}$ vale e che quindi $\neg\chi$ non può essere in V_0^S, si ricava così la tesi. ■

La possibilità di riformulare i risultati di incompletezza, tanto nel contesto dei sistemi dialettici quanto in quello delle teorie meta-formali, parrebbe rendere limitato, almeno per ciò che riguarda gli scopi che ci siamo qui prefissi, l'interesse del passaggio a simili cornici formali. Si profila dunque la necessità di un chiarimento, soprattutto per i sistemi dialettici che incarnano l'idea di una maggiore plausibilità euristica dalla quale siamo partiti, e soprattutto in relazione a quei risultati che (come 5.2.19) sembrano rendere plausibile la riproposizione delle stesse obiezioni concettuali avanzate già nei confronti delle usuali teorie assiomatiche come modelli dell'agire umano. Occorre chiarire quindi in quale senso, di fronte ai formalismi specificati in precedenza, possiamo sostenere di essere in una posizione di vantaggio rispetto alle argomentazioni come quelle di Lucas e Penrose, le quali, si suppone, si applicano ai sistemi formali più tradizionali[525].

[525]Meritano di essere menzionati, per quanto estranei alle problematiche del presente

5.3 Siamo sistemi dialettici?

Un confronto tra gli usuali sistemi formali d'assiomi ed i sistemi dialettici lo si può fare, seguendo le indicazioni dello stesso Magari[526], sulla base di un gioco che, senza troppe difficoltà, si può realizzare come esso ricalchi, in larga misura, le condizioni del Gioco dell'Imitazione ideato da Turing. Poiché una simile esemplificazione permette inoltre di apprezzare in modo inequivocabile i vantaggi dei sistemi proposti dall'algebrista italiano sulle teorie assiomatiche tradizionali (almeno per ciò che concerne il fenomeno dell'incompletezza che all'attenzione della presente indagine), crediamo sia doveroso, ancor prima che opportuno, ricorrere ad una simile spiegazione.

Se supponiamo di prendere le mosse da un sistema formale come PA, ci possiamo porre come obiettivo quello di minimizzare l'insieme $\mathbf{V} \dot{+} \mathbf{T}_{\mathsf{PA}}$ che, in ogni caso, sappiamo essere non vuoto per i risultati ordinari ai quali si è accennato in precedenza.

Concentriamo a tal fine l'attenzione, in particolare, su quegli enunciati ϕ che ne fanno parte e per i quali valga:

$$\mathsf{PA} \nvdash \phi, \mathsf{PA} \nvdash \neg\phi, \mathsf{PA} \vdash \neg\phi \rightarrow Teor_{\mathsf{PA}}(\ulcorner\phi\urcorner)$$

Questa è la situazione, ad esempio, dell'enunciato indecidibile di Gödel.

Si consideri un sistema formale corretto $\mathsf{S} = \langle \Phi, c \rangle$ che estenda PA, ossia tale che valga $\mathbf{T}_{\mathsf{PA}} \subseteq \Phi(\emptyset) \subseteq \mathbf{V}$, e costruiamo un sistema dialettico $\mathsf{S}_d = \langle h, f, c \rangle$ in modo che siano verificate:

(i) S è il sistema formale associato di S_d[527];

---

lavoro, alcuni contributi successivi ispirati dalle indagini di Magari. In particolare, i due lavori di Aldo Ursini ([Urs77] e [Urs78]), il primo dei quali contiene la definizione di una progressione transfinita di teorie assiomatiche 'alla Feferman' (cfr. [Fef62]), ispirata alle condizioni soddisfatte dall'insieme V_0^S: ad ogni stadio si aggiungono cioè gli enunciati non refutabili ma dimostrabilmente falsificabili allo stadio precedente. Si dimostra che la riunione sulla progressione così definita coincide con l'insieme delle formule vere nel modello standard dell'aritmetica. Nel secondo lavoro, più breve, Ursini fornisce una dimostrazione della Σ_2^0 completezza di V_0^S in PA (un risultato che Magari cita, ma non dimostra, e che è alla base della proposizione 5.2.20), rispetto ad una formula che rappresenta V_0^S diversa rispetto a quella da noi indicata con θ_0 e che corrisponde, in [Mag75], a $\dot{V}_0$. Vale la pena citare anche, essendo l'ultimo contributo a noi noto, in ordine di tempo, che si ispira a [Mag74] e [Mag75], l'articolo [MSS96].

[526] Cfr. [Mag74], pp. 146-149.

[527] Come si è osservato in relazione alla dimostrazione del corollario 5.2.12, è sempre possibile verificare una simile condizione.

(ii) la f sia una biezione computabile scelta in modo tale che, dato $m \in \mathbf{N}$ per il quale si abbia $f(m) = \phi$ (dove ϕ è un enunciato che verifica le condizioni suddette), si abbia: (†) se $n \in \mathbf{N}$ è tale che $f(n) = \neg\phi$, allora $m < n$; (‡) per ogni $r \in \mathbf{N}$ tale che $f(r) = \psi$ e $\psi \to \neg\phi \in \Phi(\emptyset)$, sia $m < r$.

Vale allora:

$$\phi \in \mathbf{V} \Leftrightarrow \phi \in A_d \tag{5.14}$$

Infatti si ha che, in un senso, poiché valgono $\phi \notin \mathbf{V} \Leftrightarrow \neg\phi \in \mathbf{V}$, $\neg\phi \to Teor_{\mathsf{PA}}(\ulcorner\neg\phi\urcorner) \in \Phi(\emptyset)$ e $\mathbf{T}_{\mathsf{PA}} \subseteq \Phi(\emptyset) \subseteq \mathbf{V}$, vale anche:

$$\phi \notin \mathbf{V} \Rightarrow \neg\phi \in \Phi(\emptyset) \subseteq A_d \overset{5.2.8}{\Rightarrow} \phi \notin A_d$$

Viceversa, se $\phi \in \mathbf{V}$ allora $\neg\phi \notin \Phi(\emptyset)$ e, verificando la f le condizioni (i) e (ii) suddette, segue, per $n \in \mathbf{N}$ come da (†), e per ogni $r \in \mathbf{N}$ come da (‡):

$$\begin{array}{lcl} c & \in & \Phi(\{f(n)\} \cup (A_d \cap \{f(i) \mid i < n\})) \\ c & \in & \Phi(\{f(r)\} \cup (A_d \cap \{f(j) \mid j < r\})) \end{array}$$

dal ché si può concludere (sulla base di 5.2.6) che $\neg\phi \notin A_d$. Ergo, per 5.2.11, $\phi \in A_d$.

Immaginiamo adesso di condurre un gioco, costruito sulla falsariga di quello di Turing e che prevede perciò due giocatori A e B ed un interrogatore C, con la sola variante, rispetto alla situazione indicata dal matematico britannico, che, anziché venire smascherati da C, A e B acquistino punteggio (un punto, in particolare) per ogni risposta corretta alle domande poste da C, non guadagnino nulla per le mancate risposte e perdano in egual misura (sempre un punto, cioè) per le risposte sbagliate.

Se supponiamo allora di far giocare una partita (come A e B, rispettivamente) al sistema S di cui sopra da un lato, ed al sistema S_d costruito in precedenza dall'altro, assumendo di indicare con r_n^{A} (r_n^{B}) la risposta di A (B) all'n-esimo quesito e supponendo infine che C si limiti a chiedere 'ϕ è vera?' (per una certa ϕ per la quale la costruzione precedente si realizza), otteniamo il seguente quadro[528].

[528]Si potrebbe descrivere la situazione in termini equivalenti, ma, forse, più verosimili, pensando, come fa Magari, ad A e B come due giocatori (umani) i quali decidono, come strategia di gioco, di riferirsi, per l'uno, alla presenza o meno di una dimostrazione di ϕ in S (e di rispondere solo nel caso affermativo, di astenersi in caso ϕ sia indecidibile), di riferirsi, all'n-esima domanda, all'insieme A_n delle tesi provvisorie allo stadio n-esimo di S_d, l'altro.

Distinguendo i vari casi, si ha:

(i) $\phi \notin \mathbf{V}$. Dunque, $\neg\phi \in \mathbf{V}$ e, poichè $\neg\phi \to Teor_{\mathsf{PA}}(\ulcorner\neg\phi\urcorner) \in \Phi(\emptyset)$ e $\Phi(\emptyset) \subseteq \mathbf{V}$ valgono, segue:

$$\lim_{n\to+\infty} r_n^A = +\infty$$

Per quel che riguarda B, dal momento che $\phi = f(m) < f(n) = \neg\phi$, si ha che, per certi $i \in \mathbf{N}$ iniziali, vale $\phi \in A_i$ e B perde dunque dei punti. Da un certo momento in poi, varrà invece $\phi, \neg\phi \notin A_j$ e B rimane dunque stabile nel proprio punteggio. Ma, infine, accadrà anche che, in relazione ad un $p_0 = \min p(f(p) = \psi \ \& \ \psi \to \neg\phi \in \Phi(\emptyset))$, da un certo $k \in \mathbf{N}$ in poi, $\neg\phi \in A_k$. Dunque:

$$\lim_{n\to+\infty} r_n^B = +\infty$$

(ii) $\phi \in \Phi(\emptyset)$ e, quindi, $\phi \in \mathbf{V}$. In tal caso:

$$\lim_{n\to+\infty} r_n^A = +\infty$$

e, per 5.2.7:

$$\lim_{n\to+\infty} r_n^B = +\infty$$

(iii) $\phi \in \mathbf{V}$ ma $\phi \notin \Phi(\emptyset)$. Sulla base delle ipotesi fatte:

$$\lim_{n\to+\infty} r_n^A = 0$$

Valendo tuttavia (5.14), si ha invece:

$$\lim_{n\to+\infty} r_n^B = +\infty$$

La situazione così descritta mostra in modo chiaro il vantaggio rappresentato dalla strategia per tentativi ed errori alla base di un sistema dialettico, rispetto ai più tradizionali sistemi formali.

Se vista alla luce degli scopi che ci siamo prefissi nel presente lavoro, tuttavia, una simile descrizione si presta ad almeno due obiezioni:

1. l'intera costruzione dipende in modo essenziale dalla scelta di una biezione computabile f che soddisfi le condizioni indicate in precedenza, e dalle altre ipotesi fatte. Ciò sembra rendere plausibile l'esistenza di una strategia ulteriormente migliorativa di quella seguita da B (che tenga conto cioè di condizioni meno specifiche), e che, qualora la partita fosse giocata tra un sistema dialettico ed un matematico, potrebbe consentire comunque a quest'ultimo di esibire la superiorità delle proprie capacità;

2. poiché, sulla scorta di un risultato come 5.2.19 siamo in grado di esibire in modo effettivo un enunciato che è vero se e solo se non fa parte delle tesi definitive di un sistema dialettico, la situazione di superiorità alla quale allude l'obiezione precedente, si dovrebbe verificare, in particolare, in relazione ad un enunciato siffatto. Se in un caso come quello descritto nella simulazione di partita, siamo dunque in grado di sfruttare il fatto che $(\mathbf{V}\dot{+}A_d) \subset (\mathbf{V}\dot{+}\Phi(\emptyset))$, un matematico si trova comunque dimostrabilmente in una posizione di vantaggio rispetto ad un qualsiasi sistema dialettico (e dunque non può essere identificato con nessuno di essi).

Potendo chiaramente essere viste come due aspetti di una stessa posizione, le due obiezioni verranno considerate come una sola.

Si suppone quindi (come farebbe un propugnatore dell' "obiezione matematica") che al di là del vantaggio che essi rivelano rispetto alle teorie tradizionali, i sistemi dialettici, valendo per essi un'appropriata riformulazione delle ordinarie limitazioni, mostrerebbero le stesse carenze se confrontate con i matematici. Si potrebbe da ciò dedurre, quindi, l'idea che l'elemento euristico che ne costituisce la ragion d'essere e, ad un tempo, l'aspetto di maggior novità, non ha niente a che vedere con la questione del confronto, alla luce dei teoremi limitativi, tra la mente umana e le macchine.

Tuttavia, non sarà sfuggito il fatto che un teorema come 5.2.19, che sancisce l'esistenza di una procedura effettiva che in relazione ad ogni sistema dialettico S_d permette di esibire una formula vera *se e solo se essa non sta in* A_d, non garantisce affatto (o almeno non lo fa in modo banale) che le cose stiano come si dice nell'obiezione in questione: non abbiamo, infatti, un metodo che ci permette di stabilire se un dato enunciato sia (o meno) una tesi definitiva, e, dunque, dal risultato suddetto non segue affatto in modo immediato l'esistenza di una procedura che ci permette, allo stesso tempo, di 'riconoscere come vera' una proposizione esibita in tal modo.

Supposto che nel caso delle teorie ordinarie argomenti come l' "obiezione matematica" mostrino la superiorità dell'uomo sulle macchine (e quindi la diversità tra essi), non abbiamo affatto, nel caso dei sistemi dialettici (che pure rivelano, in certi casi notevoli, quella stessa superiorità), una *prova* che ciò continui a valere (come i sostenitori della prima posizione ritengono, probabilmente a torto, nel caso dei sistemi formali).

È pur sempre vero, che rimane in piedi la possibilità che esista una strategia migliore rispetto a quella 'adottata' da un sistema dialettico, e che proprio questa si riveli essere alla base del *modus operandi* dei matematici. Ma questa eventualità che, intanto, sarebbe tutt'altro che sorprendente, considerato che, senza uscire dalla ricerca ordinaria, di strategie migliorative se ne possono fornire molteplici esempi, non dice nulla sull'impossibilità di scoprire poi, anche per essa, l'esistenza di un modello meccanico.

In altre parole, per usare la metafora della cipolla di Turing (della quale si è detto al §3.4), una volta strappata la 'buccia' delle procedure per tentativi ed errori, non abbiamo ragioni logico-matematiche per affermare con certezza che lo strato successivo non possa anch'esso essere strappato via, a seguito di un'attenta analisi.

5.4 La questione euristica ed il rapporto menti/macchine

Come sarà già risultato chiaro dalla discussione conclusiva del paragrafo precedente, la trattazione sistematica da parte di Magari dell'idea di un modello euristicamente fedele a certe strategie adottate dai matematici, ci pare fornisca alcuni spunti di riflessione i quali, aggiungendosi alle considerazioni da parte di Gödel (nella più volte citata *Gibbs Lecture*) di certe 'ragioni concettuali' derivanti dalla necessità di precisazioni di carattere terminologico, ed alla disamina delle 'ragioni logiche e metamatematiche' che abbiamo visto nel capitolo precedente, concorrono a completare il quadro della critica all' "obiezione matematica".

L'esito al quale si giunge, sembra inevitabilmente essere il seguente: se si assume la prospettiva che il confronto tra le capacità razionali di un matematico e quelle di una macchina debba, innanzi tutto, passare attraverso la costruzione di un modello matematico dell'agire del primo che sia conforme a certe caratteristiche del suo operare, si è in grado di definire, in modo logicamente ineccepibile, un concetto che risponde (non senza certe inevitabili semplificazioni e limitatamente a particolari strategie conoscitive individuate) al requisito richiesto, il quale, non solo presenta considerevoli vantaggi (in relazione allo scopo prefissato) rispetto ai sistemi formali usuali, ma che non pare 'offrire il fianco' agli argomenti elaborati per questi ultimi e fondati sulle limitazioni logiche (supposto che simili argomenti abbiano ragion d'essere nel caso delle teorie assiomatiche ordinarie).

Uno degli aspetti più interessanti della questione che, avendolo solo marginalmente evidenziato nel corso del capitolo, merita, a nostro avviso, di essere ribadito, è la possibilità, con tutte le cautele del caso, di vedere nei sistemi dialettici una realizzazione di alcune delle idee di Turing.

Se infatti l'approfondimento di certe questioni logiche e metamatematiche presupposte dagli argomenti relativi alla 'superiorità' dell'uomo sulle macchine, possono essere proposti come il completamento formale della critica di Turing (e di Gödel) all' "obiezione matematica", il contributo di Magari, ci pare, permette di comprendere meglio l'ottica assunta dal matematico britannico e di valutarne la 'praticabilità' logica.

L'idea fondamentale, come abbiamo visto, consiste nel ritenere insoddisfacente, per vari motivi, il modello rappresentato dalle procedure effettive di computo. In particolare, la fonte di tale insoddisfazione risiede nella difficoltà di proporre un'analogia diretta con le modalità con cui pare operare, in certe circostanze, la comunità dei matematici. Per quanto da una prospettiva limitata che non tiene conto di molti degli aspetti più affascinanti del discorso di Turing (a cominciare dal progetto cruciale di una macchina che sia sottoposta ad un processo educativo) i quali, tuttavia, meno si prestano forse ad un'analisi in termini formali, con i sistemi dialettici siamo quantomeno in grado di esibire un esempio di come, quel 'superamento' del paradigma della nozione di algoritmo al quale abbiamo accennato sulla scorta degli scritti del matematico britannico possa avvenire, e a quale costruzione, da un punto di vista squisitamente logico, ciò possa dare vita.

È chiaro che, come nel caso dell'esperimento di Turing con il sistema di ricompense e punizioni, al quale potremmo accostare, volendo perseguire nel parallelo con il lavoro del matematico britannico, l'esperimento logico' dei sistemi dialettici, il risultato ottenuto appare piccola cosa se confrontato con i propositi che una simile interpretazione potrebbe prefigurare: si dovrebbe in particolare tenere presente il fatto che la costruzione e le sue possibili modifiche sono inevitabilmente legate alle procedure per tentativi ed errori o, più in generale, a strategie conoscitive di tipo probabilistico, e non è dunque chiaro se, e come, ad un'analisi simile potrebbero essere sottoposte quelle strategie che non abbiano, con le prime, alcuna relazione immediata.

Si tratta in ogni caso di un fatto da non sminuire che, già ad un simile livello, i risultati acquisiti consentano di rivelare i fraintendimenti che si celano dietro presunte confutazioni dell'ipotesi computazionale della mente, che vorrebbero oltretutto (è il caso dichiarato di Penrose) avere la pretesa di attaccare il fondamento concettuale di interi programmi di ricerca nel campo dell'Intelligenza Artificiale.

Ma questo non è l'unico aspetto, nei lavori di Magari, che ci pare rilevante ai fini della questione menti/macchine. Da non disdegnare, sono infatti le riflessioni che essi possono suscitare in relazione a quella problematica vista come la riproposizione, ad un livello diverso, della diatriba fondazionale.

Si è visto infatti come, soprattutto alla luce delle precisazioni concettuali di Gödel, sembrano esserci i presupposti per individuare, dietro l'analisi delle implicazioni 'filosofiche' dei teoremi di incompletezza, il quesito che il dibattito sui fondamenti della matematica ha posto al centro: quale è la 'vera' natura della conoscenza matematica?

Se si accetta quest'immagine della discussione fondazionale, che forse taluni vedrebbero come eccessivamente semplicistica e riduttiva, il paradigma che per lungo tempo ne ha dominato ed inevitabilmente condizionato gli esiti, il programma finitista di matrice hilbertiana, pare essere 'compromesso' con un certo razionalismo positivitico di fondo, che vede nell'intelletto una sorgente inesauribile di soluzioni ai quesiti che esso stesso formula (come quelli matematici), e che assume, implicitamente almeno, la possibilità di ridurne l'operare ad un sistema di regole ben definito. Un simile quadro, pur facendo indubbiamente violenza, per certi aspetti, allo sfondo concettuale sul quale Hilbert si muove, appare quantomeno compatibile con la rilettura, nella cornice del rapporto tra la mente e le macchine, degli esiti ai quali egli mira.

Alcuni dei risultati che si sono richiamati in questo capitolo, ben rifletterebbero il fallimento di quella prospettiva: c'è qualcosa di inconsistente nella miscela che si ottiene dall'ideale positivistico di un'onniscienza matematica, in combinazione con la convinzione nella riducibilità dei processi dai quali la prima scaturisce ad un sistema formale d'assiomi.

È pur vero che, come abbiamo visto, ciò continua a valere anche per proposte come i sistemi dialettici, seppur con sfumature diverse che, per problematiche più limitate come quelle poste dall' "obiezione matematica", non sono però totalmente irrilevanti.

Proprio la presa d'atto del fallimento della prospettiva originaria, tuttavia, potrebbe spingere ad accettare l'idea di una riformulazione proprio del quadro concettuale ad essa legato, modificando il positivismo hilbertiano in ragione delle procedure per tentativi ed errori, e vedendo così in costruzioni logiche come i sistemi dialettici uno sbocco naturale di un simile mutamento. Ma, considerazioni analoghe, potrebbero addirittura portare a vedere di buon occhio la ben più radicale 'trasformazione' dell'indagine metamatematica a cui sembrano puntare le considerazioni 'popperiane' di Magari, giungendo per questa strada, pur con prospettive profondamente mutate, ad individuare nozioni che, come la collezione degli enunciati 'sensatamente veri', non escludano la possibilità di essere trattate in termini molto vicini ai concetti che sono il frutto del primo cambiamento indicato (in fini dei conti, l'insieme V_0^S è un limite inferiore come lo sono i sistemi dialettici).

Se adesso, per riassumere, ritorniamo alla conclusione disgiuntiva di Gödel della *Gibbs Lecture* (o i principi evidenti della matematica non possono essere compresi in una regola finita - equivalentemente, la mente umana in campo matematico sorpassa infinitamente le capacità di ogni macchina finita - oppure ci sono problemi diofantei assolutamente indecidibili), e guardiamo ad essa come il limite dell'analisi rigorosa dei teoremi di incompletezza, ci si accorge che, vista come problema, una simile conclusione non sembra lasciare aperte, rispetto ai due termini che essa chiama in causa, la matematica e la mente umana, strade molto diverse da quelle che si sono fin qui incontrate (e variamente perseguite): o con un'analisi diretta dei processi razionali che sono all'origine della 'certezza matematica' si giunge a stabilire se la mente sia o meno equivalente ad una qualche macchina, oppure, mediante una disquisizione di tipo più filosofico sulla natura della disciplina, si giunge a disvelare l'irriducibilità di quest'ultima (anche nel senso di matematica 'in senso soggettivo', cioè di collezione di 'percezioni' con certezza matematica) ad un sistema finito di regole.

Bibliografia

[And64] A. R. Anderson, *Minds and machines*, Englewood Cliff, Prentice Hall, 1964.

[Ant96] Aldo Antonelli, *Gödel, Penrose e i fondamenti dell'intelligenza artificiale*, disponibile in rete all'indirizzo `http://kleene.ss.uci.edu/~ aldo`, 1996.

[Ant00] ———, *I teoremi di Gödel e la filosofia della mente*, disponibile in rete all'indirizzo `http://kleene.ss.uci.edu/~ aldo`, 2000.

[Art94] Sergei Artëmov, *Logic of proofs*, Annals of Pure and Applied Logic **67** (1994), 29–59.

[Art01] ———, *Explicit provability and constructive semantics*, Bulletin of Symbolic Logic **7** (2001), 1–36.

[Ber35] Paul Bernays, *Sur le platonisme dans les mathématiques*, L'Einseignement mathématique **34** (1935), 52–69.

[BP64] Paul Benaceraff e Hilary Putnam (cur.i), *Philosophy of mathematics: Selected readings*, Prentice-Hall, Englewood Cliffs, 1964.

[Car34] Rudolf Carnap, *Die logische Sintax der Sprache*, Springer, 1934.

[Chu32] Alonzo Church, *A set of postulates for the foundation of logic*, Annals of Mathematics **33** (1932), 346–366.

[Chu34] ———, *The Richard paradox*, American Mathematical Monthly **41** (1934), 356–361.

[Chu36] ———, *An unsolvable problem of elementary number theory*, American Journal of Mathematics **58** (1936), 345–363.

[CK36] Alonzo Church e Stephen Cole Kleene, *Formal definitions in the theory of ordinal numbers*, Fundamenta Mathematicae **28** (1936), 11–21.

[Col97] Susan Colvin, *Intelligent machinery: Turing's ideas and their relations to the work of Newell and Simon*, Tesi di master, Department of Philosophy - Carnegie Mellon University, 1997.

[Dav58] Martin Davis, *Computability & unsolvability*, McGraw-Hill, 1958.

[Dav65] Martin Davis (cur.), *The undecidable*, Raven Press, 1965.

[Dav82] ———, *Why Gödel didn't have Church's thesis*, Information and Control **54** (1982), 2–24.

[Daw85] John Dawson, Jr., *The reception of Gödel's incompleteness theorems*, Proceedings of the Biennial Meeting of the Philosophy of Science Association, vol. 2, 1985, ristampato in [Sha88], pp. 74–95, pp. 253–271.

[Daw96] ———, *Logical dilemmas: The life and work of Kurt Gödel*, A. K. Peters, 1996.

[Err52] F. Errera, *Le problème du continu*, Atti dell'Accademia ligure delle Scienze **9** (1952), 176–183.

[Fef60] Solomon Feferman, *Arithmetization of metamathematics in a general setting*, Fundamenta Mathematicae **49** (1960), 35–92.

[Fef62] ———, *Transfinite recursive progressions of axiomatic theories*, The Journal of Symbolic Logic **27** (1962), 259–316.

[Fef84] ———, *Kurt Gödel: Conviction and caution*, Philosophia Naturalis **21** (1984), 546–562.

[Fef87] ———, *Infinity in mathematics: Is Cantor necessary?*, L'Infinito nella Scienza (G. Toraldo di Francia, cur.), Istituto della Enciclopedia Italiana, 1987, pp. 151–209.

[Fef88] ———, *Turing in the land of O(z)*, The Unversal Turing Machine. A Half-Century Survey (R. Herken, cur.), Oxford University Press, 1988, pp. 103–134.

[Fef95] ———, *Penrose's gödelian argument. A review of shadows of the mind by Roger Penrose*, in [P$^+$95], 1995.

[Fin26] Paul Finsler, *Formale Beweis und die Entscheidbarkeit*, Mathematische Zeitschrifft **25** (1926), 676–682, ristampato e tradotto in [vH67], pp. 438–445.

[Gan96] Robin Gandy, *Human versus mechanical intelligence*, Machines and Thought: The Legacy of Alan Turing (P. Millican e A. Clark, cur.i), Mind Association Occasional Series, Clarendon Press, 1996, pp. 125–136.

[Göd29] Kurt Gödel, *Über die Vollständigkeit des Logikkalküls*, Habilitationschrifft, Università di Vienna, 1929, stampato e tradotto in inglese in [Göd86a], pp. 60–101.

[Göd30a] ———, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, Monatschefte für Mathematik und Physik **37** (1930), 349–360, ristampato in [Göd86a], pp. 102–123.

[Göd30b] ———, *Vortrag über Vollständigkeit des Funktionenkalküls*, stampato e tradotto in inglese in [Göd95], pp. 13–29, 1930.

[Göd31a] ———, *Diskussion zur Grundlegung der Mathematik*, Erkenntnis **2** (1931), 135–151, ristampato in [Göd86a], pp. 201–205.

[Göd31b] ———, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatschefte für Mathematik und Physik **38** (1931), 173–198, ristampato in [Göd86a], pp. 144–195.

[Göd32] ———, *Über Vollständigkeit und Wiederspruchsfreiheit*, Ergebnisse eines mathematischen Kolloquiums **3** (1932), 12–13, ristampato e tradotto in inglese in [Göd86a], pp. 234–237.

[Göd33] ———, *The present situations in the foundations of mathematics*, edito in [Göd95], pp. 36–53, 1933.

[Göd36] ______, *Über die Länge von Beweisen*, Ergebnisse eines mathematisches Kolloquiums **7** (1936), 6, ristampato e tradotto in inglese in [Göd86a], pp. 397–399.

[Göd38a] ______, *The consistency of the axiom of choice and of the generalized continuum hypotesis*, Proceedings of the National Academy of Sciences, U. S. A., vol. 24, 1938, ristampato in [Göd90a], pp. 26–27, pp. 556–557.

[Göd38b] ______, *Lecture at Göttingen*, edito in [Göd95], pp. 126–155, 1938.

[Göd38c] ______, *Undecidable diophantine propositions*, edito in [Göd95], pp. 157–175, 1938.

[Göd38d] ______, *Vortrag bei Zilsel*, edito in [Göd95], pp. 86–113, 1938.

[Göd40] ______, *Lecture [on the] consistency [of the] continuum hypotesis (Brown University)*, edito in [Göd95], pp. 175–185, 1940.

[Göd41] ______, *In what sense is intuitionistic logic constructive?*, edito in [Göd95], pp. 189–200, 1941.

[Göd44] ______, *Russell's mathematical logic*, The Philosophy of Bertrand Russell (Paul Arthur Schilpp, cur.), Northwestern University Press, 1944, ristampato in [Göd90a], pp. 119–141, pp. 123–153.

[Göd46a] ______, *Some observations about the relationship between theory of relativity and kantian philosophy*, versione preliminare di [Göd49c], stampata in [Göd95], pp. 230–246, 1946.

[Göd46b] ______, *Some observations about the relationship between theory of relativity and kantian philosophy*, versione preliminare di [Göd49c], stampata in [Göd95], pp. 247–259, 1946.

[Göd47] ______, *What is Cantor's continuum problem?*, American Mathematical Monthly **54** (1947), 515–525, ristampato in [Göd90a], pp. 230–259.

[Göd49a] ______, *An example of a new type of cosmological solutions of einstein's field equations of gravitation*, Reviews of Modern Physics **21** (1949), 447–450.

[Göd49b] ______, *Lecture on rotating universe*, stampato in [Göd95], pp. 269–287, 1949.

[Göd49c] ______, *A remark about the relationship between relativity theory and kantian philosophy*, Albert Einstein: Philosopher-Scientist (Paul Arthur Schilpp, cur.), Northwestern University Press, 1949, pp. 555–562.

[Göd51] ______, *Some basic theorems on the foundations of mathematics and their implications*, edito in [Göd95], pp. 304–323, 1951.

[Göd52] ______, *Rotating universes in general relativity theory*, Proocedings of the International Congress of Mathematicians; Cambridge, Massachussetts, U.S.A. August 30-September 6, 1950, I, Providence R.I.; American Mathematical Society, 1952, pp. 175–181.

[Göd58] ______, *Über ein bisher noch nicht benütze Erweiterung des finiten Standpunktes*, Dialectica **12** (1958), 280–287, ristampato e tradotto in inglese in [Göd90a], pp. 208–216.

[Göd61] ______, *The modern development of the foundations of mathematics in the light of philosophy*, edito in [Göd95], pp. 375–387, 1961.

[Göd64] ______, *What is Cantor's continuum problem?*, in Benaceraff e Putnam [BP64], ristampato in [Göd90a], pp. 150–153.

[Göd65a] ______, *On undecidable propositions of formal mathematical systems*, The Undecidable, Raven Press, 1965, ristampato in [Göd86a], pp. 346-371.

[Göd65b] ______, *Remarks before the Princeton bicentennial conference on problems of mathematics*, The Undecidable (Martin Davis, cur.), Raven Press, 1965, ristampato in [Göd90a], pp. 150–153, pp. 84–88.

[Göd70a] ______, *A proof of Cantor's continuum hypothesis from a highly plausible axiom about orders of growth*, stampato in [Göd95], pp. 422–423, 1970.

[Göd70b] ———, *Some considerations leading to the probable conclusion that the true power of the continuum is* $\aleph_2$, stampato in [Göd95], pp. 420–422, 1970.

[Göd86a] ———, *Collected works vol. I: Publications 1929-1936*, Oxford University Press, 1986.

[Göd86b] ———, *Recensione di [Hil31b]*, Collected Works vol. I: Unpublished Essays and Lectures, Oxford University Press, 1986, pp. 212–215.

[Göd90a] ———, *Collected works vol. II: Publications 1938-1974*, Oxford University Press, 1990.

[Göd90b] ———, *On an extension of finitary mathematics which has not yet been used*, in [Göd90a], pp. 271–280, 1990.

[Göd90c] ———, *Some remarks on the undecidability results*, in [Göd90a], pp. 305–306, 1990.

[Göd95] ———, *Collected works vol. III: Unpublished essays and lectures*, Oxford University Press, 1995.

[Göd03a] ———, *Collected works vol. IV: Correspondence A-G*, Oxford University Press, 2003.

[Göd03b] ———, *Collected works vol. V: Correspondence H-Z*, Oxford University Press, 2003.

[Gol65] E. M. Gold, *Limiting recursion*, The Journal of Symbolic Logic **18** (1965), 28–48.

[Háj77] Petr Hájek, *Experimental logics and* Π^0_3 *theories*, The Journal of Symbolic Logic **42** (1977), 515–522.

[Hal62] Paul Halmos, *Algebraic logic*, Chelsea, New York, 1962.

[HB39] David Hilbert e Paul Bernays (cur.i), *Grundlagen der Mathematik*, vol. 2, Springer-Verlag, 1939.

[Hil28] David Hilbert, *Die Grundlagen der Mathematik*, Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität **6** (1928), 65–85.

[Hil31a] ______, *Beweis der Tertium Non Datur*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse (1931), 120–125.

[Hil31b] ______, *Die Grundlegung der elementaren Zahlenlehre*, Mathematische Annalen **104** (1931), 485–494.

[Hil78] ______, *Problemi matematici*, Ricerche sui Fondamenti della Matematica. Scritti Fondazionali di David Hilbert (Vito Michele Abrusci, cur.), Bibliopolis, 1978, pp. 145–162.

[Hod91] Andrew Hodges, *Storia di un enigma. vita di Alan Turing*, Bollati Boringhieri, 1991, edizione italiana di (dello stesso autore) “Alan Turing: the enigma”, Burnett Books - Simon and Schuster, 1983.

[Hof79] Douglas Hofstadter, *Gödel, Escher, Bach: an eternal golden braid*, Basic Books, 1979.

[Inc92] D. C. Ince, *The collected works of A. M. Turing, vol. II*, North Holland, 1992.

[Jer75] R. G. Jeroslow, *Experimental logics and Δ^0_2-theories*, Journal of Philosophical Logic **4** (1975), 253–267.

[Kle36] Stephen Cole Kleene, *General recursive functions of natural numbers*, Mathematische Annalen **112** (1936), 727–742.

[Kle38] ______, *On notations for ordinal numbers*, The Journal of Symbolic Logic **3** (1938), 150–155.

[Kle43] ______, *Recursive predicates and quantifiers*, Transactions of the American Mathematical Society **53** (1943), 41–71.

[Kle52] ______, *Introduction to metamathematics*, North Holland, 1952.

[Kle55] ______, *On the forms of the predicates in the theory of constructive ordinals II*, American Journal of Mathematics **77** (1955), 405–428.

[KR35] Stephen Cole Kleene e John Barkley Rosser, *The inconsistency of certain formal logics*, Annales of Mathematics (2) **36** (1935), 630–636.

[Kre65a] Georg Kreisel, *Mathematical logic*, Lectures on Modern Mathematics (Thomas L. Saaty, cur.), Wyley, New York, 1965, pp. 95–195.

[Kre65b] Georg Kreisel, *Model-theoretic invariants: applications to recursive and hyperarithmetic operations*, The Theory of Models (J. W. Addison, L. Henkin, e A. Tarski, cur.i), North Holland, 1965, pp. 190–205.

[Kre87] Georg Kreisel, *Church's Thesis and the ideal of informal rigour*, Notre Dame Journal of Formal Logic **28** (1987), 449–519.

[Lin01] Per Lindstöm, *Penrose's new argument*, Journal of Philosophical Logic **30** (2001), 241–250.

[Löb50] M. H. Löb, *Solution of a problem of Leon Henkin*, The Journal of Symbolic Logic **20** (1950), 115–118.

[Luc61] J. R. Lucas, *Minds, machines and Gödel*, Philosophy **36** (1961), 112–127, ristampato in [And64], pp. 43–59.

[Luc96] ______, *Minds, machines and Gödel: A retrospect*, Machines and Thought: The Legacy of Alan Turing (P. Millican e A. Clark, cur.i), Mind Association Occasional Series, Clarendon Press, 1996, pp. 103–124.

[Mag74] Roberto Magari, *Su certe teorie non enumerabili (sulle limitazioni dei sistemi formali I)*, Annali di Matematica Pura ed Applicata **48** (1974), 119–152.

[Mag75] ______, *Significato e verità nell'aritmetica peaniana*, Annali di Matematica Pura ed Applicata **103** (1975), 343–368.

[Mar54] Werner Markwald, *Zur Theorie der konstruktiven Wohlordnungen*, Mathematische Annalen **127** (1954), 135–149.

[MS95] Daniele Mundici e Wilfried Sieg, *Paper machines*, Philosophia Mathematica **3** (1995), 5–30.

[MSS96] Franco Montagna, Giulia Simi, e Andrea Sorbi, *Logic and probabilistic systems*, Archive for Mathematical Logic **33** (1996), 225–259.

[P$^+$95] Roger Penrose et al., *Commentaries on 'Shadows of the mind'*, Psyche, 2, 1995, disponibile in rete all'indirizzo `http://www.psyche.cs.monash.edu.au`.

[Pen89] Roger Penrose, *The emperor's new mind*, Oxford University Press, 1989.

[Pen96] ———, *Ombre della mente. Alla ricerca della coscienza*, Rizzoli, 1996, edizione italiana di (dello stesso autore) "Shadows of The Mind", Oxford University Press, 1994.

[Pic03] Gualtiero Piccinini, *Alan Turing and the mathematical objection*, Mind and Machines **13** (2003), 23–48.

[Pos36] Emil Post, *Finite combinatory processes. Formulation I*, The Journal of Symbolic Logic **1** (1936), 103–105.

[Pos43] ———, *Formal reductions of the general combinatorial problem*, American Journal of Mathematics **65** (1943), 197–215.

[Pos65] ———, *Absolutely unsolvable problems and relatively undecidable propositions: Account of an anticipation*, in Davis [Dav65], pp. 248–291.

[Pud99] Pavel Pudlák, *A note on the applicability of the incompleteness theorems to human mind*, Annals of Pure and Applied Logic **96** (1999), 335–342.

[Put65] Hilary Putnam, *Trial and error predicates*, The Journal of Symbolic Logic **18** (1965), 49–57.

[Rei70] Constance Reid, *Hilbert*, Springer Verlag, 1970.

[Rog67] Hartley Rogers, *Theory of recursive functions and effective computability*, McGraw Hill Series, 1967.

[Ros36] John Barkley Rosser, *Extensions of some theorems of Gödel and Church*, The Journal of Symbolic Logic **1** (1936), 87–91.

[SB96] Wilfried Sieg e John Byrnes, *k graph machines: Generalizing Turing's machine and arguments*, Gödel '96 (Peter Hajek, cur.), Lecture Notes in Logic, Springer-Verlag, 1996, pp. 98–119.

[Sch60] Kurt Schütte, *Beweistheorie*, Springer Verlag, 1960.

[Sch77] H. Schwichtenberg, *Proof theory: Some applications of cut elimination*, Handbook of Mathematical Logic (J. Barwise, cur.), North Holland, 1977, pp. 867–895.

[Sha88] Stuart G. Shanker, *Gödel's theorem in sharper focus*, Croom Helm, 1988.

[Sho67] Joseph R. Shoenfield, *Mathematical logic*, Addison-Wesley, 1967.

[Sie91a] Wilfried Sieg, *Effectiveness and provability*, Rendiconti del Seminario Fisico e Matematico di Milano **61** (1991), 219–230.

[Sie91b] ———, *Herbrand analyses*, Archive for Mathematical Logic **30** (1991), 409–441.

[Sie94] ———, *Mechanical procedures and mathematical experience*, Mathematics and Mind (Alexander George, cur.), Oxford University Press, 1994, pp. 71–117.

[Sie97] ———, *Step by recursive step: Church's analysis of effective calculability*, The Bullettin of Symbolic Logic **3** (1997), 154–180.

[Sie01] ———, *Calculations by man and machine: Conceptual analysis*, Reflections on the Foundation of Mathematics. Essays in Honour of Solomon Feferman (Wilfried Sieg, Richard Sommer, e Carolyn Talcott, cur.i), Lecture Notes in Logic 15, A K Peters, 2001, pp. 387–406.

[Sin00] Hourya Sinaceur, *Address at the Princeton University bicentennial conference on problems of mathematics by Alfred Tarski*, The Bullettin of Symbolic Logic **6** (2000), 1–44.

[Sko70] Toralf Skolem, *Selected works in logic*, Universitetsforlaget, 1970.

[Sma61] J. J. C. Smart, *Gödel's theorem, Church's theorem and mechanism*, Synthèse **13** (1961), 105–110.

[Smo77] Craig Smorynski, *The incompleteness theorems*, Handbook of Mathematical Logic (J. Barwise, cur.), North Holland, 1977, pp. 821–865.

[Spe55] Clifford Spector, *Recursive well-orderings*, The Journal of Symbolic Logic **20** (1955), 151–163.

[Tai81] William W. Tait, *Finitism*, Journal of Philosophy **78** (1981), 524–546.

[Tak55] Gaisi Takeuti, *On the foundamental conjecture of GCLI*, Journal of the Mathematical Society of Japan **7** (1955), 249–275.

[Tie98] Richard Tieszen, *Gödel's path from the incompleteness theorems (1931) to phenomenology (1961)*, The Bulletin of Symbolic Logic **4** (1998), 181–203.

[Tur36] Alan Mathison Turing, *On computable numbers, with an application to the Entscheidungsproblem*, Proccedings of the London Mathematical Society (2) **42** (1936), 230–265, ristampato in [Dav65], pp. 116-154.

[Tur39] ———, *Systems of logic based on ordinals*, Proceedings of the London Mathematical Society **45** (1939), 161–228, ristampato in [Dav65], pp 155-222.

[Tur48] ———, *Intelligent machinery*, Machine Intelligence **5** (1948), 3–23, ristampato in [Inc92], pp. 107–127.

[Tur50] ———, *Computing machinery and intelligence*, Mind **59** (1950), 433–460.

[Tur86] ———, *Lecture to the London Mathematical Society on 20 February 1947*, A. M. Turing's ACE Report of 1946 and Other Papers (B. E Carpenter e R. W. Doran, cur.i), Cambridge University Press, 1986, ristampato in [Inc92], pp. 107–127., pp. 106–124.

[Urs77] Aldo Ursini, *A sequence of theories whose union is complete*, Rendiconti del Seminario Matematico dell'Università di Padova **57** (1977), 75–91.

[Urs78] ———, *On the set of 'meaningful' sentences of arithmetic*, Studia Logica **37** (1978), 237–241.

[vH67] Jean van Heijenoort (cur.), *From Frege to Gödel: A source book in mathematical logic*, Harward University Press, 1967.

[vHK03] Mark van Hatten e Juliette Kennedy, *On the philosophical development of Kurt Gödel*, The Bulletin of Symbolic Logic **9** (2003), 425–476.

[vN27] John von Neumann, *Zur hilbertischen Beweistheorie*, Mathematische Zeitschrifft **26** (1927), 1–46.

[Wan74] Hao Wang, *From mathematics to philosophy*, Routledge and Kegan Paul, 1974.

[Wan81] ———, *Some facts about Kurt Gödel*, The Journal of Symbolic Logic **46** (1981), 653–659.

[Wan96] ———, *A logical journey. From Gödel to philosophy*, MIT Press, 1996.