UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
CURSO DE MATEMÁTICA

Marcello Silveira Marochi

Incompletude concreta de sistemas aritméticos: um estudo sobre Teoremas de Gödel e suas consequências para o 'fazer matemático'

Florianópolis
2021

Marcello Silveira Marochi

Incompletude concreta de sistemas aritméticos: um estudo sobre Teoremas de Gödel e suas consequências para o 'fazer matemático'

Trabalho de Conclusão de Curso submetido ao Curso de Matemática da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Matemática.
Orientador: Prof. Cezar Augusto Mortari, Dr.

Florianópolis
2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Marochi, Marcello
Incompletude concreta de sistemas aritméticos : um estudo sobre os Teoremas de Gödel e suas consequências para o 'fazer matemático' / Marcello Marochi ; orientador, Cezar Augusto Mortari, 2021.
140 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro de Ciências Físicas e Matemáticas, Graduação em Matemática e Computação Científica, Florianópolis, 2021.

Inclui referências.

1. Matemática e Computação Científica. 2. metamatemática. 3. incompletude. 4. indecidibilidade. 5. funções recursivas. I. Mortari, Cezar Augusto. II. Universidade Federal de Santa Catarina. Graduação em Matemática e Computação Científica. III. Título.

Marcello Silveira Marochi

Incompletude concreta de sistemas aritméticos: um estudo sobre Teoremas de Gödel e suas consequências para o 'fazer matemático'

O presente trabalho em nível de Bacharel foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Cezar Augusto Mortari, Dr.
Universidade Federal de Santa Catarina

Prof. Décio Krause, Dr.
Universidade Federal de Santa Catarina

Prof. Fernando de Lacerda Mortari, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Bacharel em Matemática.

_____

Coordenação do Curso de Matemática

_____

Prof. Cezar Augusto Mortari, Dr.
Orientador

Florianópolis, 2021.

Este trabalho de conclusão de curso é dedicado a todos
os autores que não dedicam a si mesmos os seus
próprios trabalhos de conclusão de curso.

AGRADECIMENTOS

Antes de qualquer consideração, o presente trabalho não poderia ser elaborado sem o constante apoio de meus pais, Darcy Roberto e Cecília Silveira Marochi, durante todas as fases de minha vida. Agradeço, especialmente, pela compreensão quando desisti de minha antiga graduação e resolvi estudar Matemática.

Além deles, devo agradecer a todos os professores que, de alguma forma, fizeram-me despertar o interesse pelos assuntos que permeiam o tema deste trabalho. Meus profundos agradecimentos, em especial, a Fernando de Lacerda Mortari, com quem tive o prazer de aprender quatro matérias diferentes, e cujo rigor formal e capacidade oratória muito me inspiram a escrever e explicar as coisas de maneira apropriada; também dedico um agradecimento especial à Melissa Weber Mendonça, que, ao longo de meus quatro anos de graduação, foi a única professora a mencionar (e tentar explicar) os Teoremas de Incompletude Gödel — esse foi o pontapé inicial para minha pesquisa; ao professor Décio Krause, notável mestre do departamento de filosofia, que, apesar da extensão deste trabalho, aceitou participar da banca avaliadora; e, finalmente, agradeço ao meu orientador, Cezar Augusto Mortari, em primeiro lugar, pelas duas ótimas matérias de Lógica a mim lecionadas, que aumentaram significativamente meu interesse pelo assunto. Em segundo lugar, é claro, pela vontade, tempo e paciência dedicados a me orientar nesse projeto.

Embora, muito provavelmente, tais pessoas nunca venham a tomar conhecimento disso, agradeço imensamente às figuras de Alexandra Elbakyan (quem carrega nas costas boa parte da produção de conhecimento acadêmico neste mundo) e de Joel David Hamkins (professor da Universidade de Oxford, celebridade do fórum matemático MathOverflow, e autor da sentença original que usei como dedicatória, na página anterior; ela é uma versão do paradoxo do barbeiro).

Agradeço imensamente a todos os amigos que me acompanharam nessa caminhada. Principalmente, ao ouvido curioso de Danielly Sorato, querida amiga e colega, mestra cientista da computação. Nesses últimos meses, com certeza expliquei a ela sobre minha pesquisa mais do que para qualquer outra pessoa. Suas perguntas me fizeram escrever um trabalho muito melhor. Meu último agradecimento vai à minha gata de estimação, Siwé, que, também muito provavelmente, não saberá o quanto devo a ela em questão de afeto e saúde mental — mas fica aqui o registro, caso os felinos finalmente resolvam despertar consciência e dominar o mundo.

“Nós precisamos saber,
e nós vamos saber.”
(HILBERT, 1930)

RESUMO

Na virada da década de 1930, o lógico austríaco Kurt Gödel, no auge de seus 24 anos, abalou as bases fundacionais da Matemática, ao apresentar aquilo que viriam a chamar de seus teoremas de incompletude. Segundo eles, qualquer sistema formal capaz de descrever uma porção mínima de aritmética elementar deve conter sentenças impossíveis de ser demonstradas — em particular, sentenças que afirmam a *consistência* de um tal sistema são indemonstráveis. Essa descoberta pôs em xeque todos os esforços da comunidade matemática fundamentalista à época, que, na tentativa de escapar dos paradoxos e antinomias inerentes à teoria dos conjuntos de Cantor, procurava por uma axiomatização completa e garantidamente consistente para as mais diversas áreas da Matemática. Nesse sentido, o objetivo inicial do presente trabalho é apresentar uma prova detalhada dos teoremas de incompletude, desde a construção de um sistema formal específico (chamado de Aritmérica de Peano, ou simplesmente PA) até a construção de algumas sentenças indecidíveis. Para isso, montamos todo o arsenal necessário para atacar problemas de indecidibilidade — como os conceitos equivalentes de máquinas de Turing e funções recursivas; o fato de que toda função recursiva pode ser representada em PA; e que a linguagem utilizada está em correspondência com os números naturais (de modo que podemos construir sentenças aritméticas que falam *sobre* a aritmética). Com isso, pretendemos desmistificar os significados dos teoremas de incompletude, dirimindo algumas das (muitas) interpretações errôneas que estes receberam ao longo do século. Nosso segundo (e principal) objetivo concerne o aspecto metalinguístico dos exemplos originais de Gödel e Rosser: ambas estas sentenças, bem como as sentenças que afirmam a consistência de PA, dizem algo sobre elas mesmas, ou sobre o próprio sistema em que estão inseridas. Nesse sentido, mais comum do que matemáticos que desconhecem os teoremas de Gödel são os matemáticos que os negligenciam, pois eles também desconhecem o fato de que várias sentenças *não*-metalinguísticas (i.e., que ocorrem "naturalmente" nas áreas matemáticas usuais, i.e., fora da lógica-matemática e da teoria dos conjuntos) já tiveram sua indecidibilidade comprovada. Nossa tese, portanto, se dá ao mostrarmos exemplos de tais sentenças, ao mesmo tempo em que tentamos distinguir *o que são* sentenças que ocorrem "naturalmente" na Matemática.

Palavras-chave: Metamatemática. Indecidibilidade. Teoremas de Incompletude de Gödel.

ABSTRACT

In the turn of 1930s, the austrian logician Kurt Gödel, at the time a 24-year-old young man, shooked the foundational basis of Mathematics when he presented what would be named his incompleteness theorems. This results say that every formal system capable of describing a minimum portion of elementary arithmetics must contain sentences which are impossible to prove — in particular, sentences that affirm the *consistency* of such a system are unprovable. That discovery put in check all the effort of foundational mathematics community at the time, which, trying to escape from the paradoxes and antinomies inherent to Cantor's set theory, was searching for a complete and guaranteed consistent axiomatization for most of the Mathematics fields. In this sense, our initial goal with the present work is to present a detailed proof of the incompleteness theorems, from the construction of a specific formal system (namely, the Peano Arithmetic, or simply PA) to the construction of some undecidable sentences. For that, we set up the entire arsenal needed to attack problems of undecidability — like the equivalent concepts of Turing machines and recursive functions; the fact that every recursive function can be represented in PA; and that the language being used is in correspondence with the natural numbers (so that we can build arithmetical sentences that speak *about* arithmetic). Therewith, we pretend to demystify the meaning of the incompleteness theorems, resolving some of the (many) misinterpretations that such theorems received throughout the century. Our second (and main) goal concerns the metalinguistic aspect of the Gödel and Rosser original examples: both these sentences, as well as the sentences that affirm the consistency of PA, say something about theyself, or about the system itself where they occur. In this sense, more common than mathematicians that don't know the incompleteness theorems are the mathematicians that neglect them, because they are also unaware of the fact that many *non*-metalinguistic sentences (i.e., sentences that occur "naturally" in the usual mathematical branchs, i.e., outside mathematical-logic and set theory) have their undecidability already been verified. Our thesis, thus, is set when we give examples of such sentences, at the same time that we try to distinguish *what are* sentences that occur "naturally" in Mathematics.

Keywords: Metamathematics. Undecidability. Gödel's Incompleteness Theorems.

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

A História da Matemática, assim como a de qualquer área abstrata do conhecimento (como a filosofia, as ciências humanas e algumas ciências exatas puras), é marcada por diversos casos de conflito intelectual, sejam eles resultados dos embates de opiniões entre os pensadores e escolas de pensamento envolvidas, ou, simplesmente, por incongruências inerentes à própria teoria que está sendo construída. Na Matemática, de um modo geral, ocorre mais o segundo caso. Todavia, o que se viu durante o final do séc. XIX e início do séc. XX — período este em que a linguagem matemática passou por uma espécie de unificação, quando Frege e Cantor propuseram que todo constructo fosse baseado na teoria dos conjuntos e na lógica simbólica — foi um tremendo antagonismo de ideias, sobretudo quando essas mesmas ideias de Frege e de Cantor se provaram inconsistentes (o que, por sua vez, mostrava que a Matemática, como um todo, do modo como estava sendo unificada, estava sujeita a tais contradições). O presente trabalho tem nesse embate o seu ponto de partida.

Nesta introdução, vamos expor precisamente *quais* problemas surgiram com o advento da matemática moderna, de que *modo* os matemáticos da época tentaram remediá-los, e *quem* protagonizou o debate em questão. Nesse sentido, pelo caráter elementar desse tipo de discussão, definimos como nosso público-alvo qualquer pessoa bem versada em teoria dos conjuntos e lógica de primeira ordem. Em particular, acreditamos que um aluno de graduação em matemática, filosofia ou ciência da computação poderá entender a maioria dos conceitos aqui introduzidos (desde que esteja familiarizado com teorias axiomáticas e os princípios da lógica clássica).

Mas isso não significa, de modo algum, que as construções teóricas posteriores serão de fácil compreensão. Os passos para entender os Teoremas de Incompletude de Gödel — objeto central deste estudo, e que, de certo modo, veio para soterrar quase completamente as esperanças dos protagonistas do debate mencionado acima, ao apontar que qualquer sistema aritmético formal possui sentenças que não podem ser demonstradas, e que a própria consistência do sistema é uma dessas sentenças — incluem diversas abstrações um tanto quanto incomuns para a maioria dos matemáticos.

No capítulo 2, definiremos duas maneiras equivalentes de se abstrair um “processo efetivo de decisão” dentro da teoria de números. Em suma, tais processos modelam um método mecânico, automático, para se verificar a validade de uma sentença. Isso nos dará uma maneira bastante simples para verificar se determinada sentença é, ou não, um teorema matemático.

No capítulo 3, buscaremos dar sentido ao conteúdo dos Teoremas de Gödel. Para isso, precisaremos seguir vários passos, quase que de maneira algorítmica, como numa receita de bolo. Primeiro, na seção 3.1, especificaremos um sistema formal bas-

tante usual, chamado de Aritmética de Peano, determinando sua linguagem, seus axiomas e suas regras de inferência. Na seção 3.2. mostraremos que toda função utilizada para descrever os métodos efetivos de decisão do capítulo 2 pode ser representada na linguagem da Aritmética de Peano. Em particular, isso vai nos dizer que toda relação aritmética elementar, e, portanto, toda sentença imaginável sobre aritmética elementar, pode ser enunciada no sistema que escolhemos. A seção 3.3, provavelmente a mais técnica deste trabalho, servirá para mostrar uma correspondência entre a linguagem da aritmética e os próprios números naturais (que chamaremos de "codificação"). Deste modo, poderemos escrever qualquer sentença do sistema formal que escolhemos como uma sequência de números. Em particular, isso nos permitirá atribuir números a relações metalinguísticas: e.g., construiremos uma função numérica que vale para um número se e somente se ele codificar uma variável; e, mais importante para a prova dos Teoremas, construiremos uma relação numérica que se dá entre dois números se e somente se um deles é o código de uma demonstração para a sentença atribuída ao outro.

Na subseção 3.4.1, provaremos o primeiro dos teoremas de incompletude, partindo da relação numérica que representa a demonstrabilidade de uma sentença qualquer, até chegarmos numa sentença aritmética que diz, sobre si mesma, não ser demonstrável no sistema utilizado. Tal sentença, de fato, não será demonstrável na Aritmética de Peano, donde vamos concluir que este sistema é incompleto. Assumindo que o sistema é consistente, vamos descobrir que a sentença em questão é, além de indemonstrável, verdadeira. O segundo dos teoremas de incompletude será provado na subseção 3.4.2, quando estreitaremos um pouco mais nossas hipóteses, e construiremos uma sentença metateórica que diz, sobre o sistema, que o mesmo é consistente. Em suma, mostraremos que, sob a hipótese de a Aritmética de Peano ser consistente, ela mesma não é capaz de provar tal fato. Em seguida, na subseção 3.4.3, vamos apresentar alguma consequências lógicas imediatas do Teoremas. Em particular, uma anedota bastante interessante envolvendo a figura de Papai Noel será apresentada, e alguns dos mitos envolvendo incompletude serão extirpados.

Finalmente, no capítulo 4, daremos exemplos de algumas sentenças aritméticas indecidíveis (e que, portanto, também atestam a incompletude do sistema em questão) cujos conteúdos diferem totalmente daqueles que apresentamos no capítulo 3, pois não são metalinguísticos, e versam sobre entidades e problemas matemáticos usuais (i.e., problemas fora da lógica-matemática e da teoria dos conjuntos). Isso servirá para embasar nossa tese de que, realmente, há problemas de cunhos "puramente matemáticos" que, embora possam ser perfeitamente formulados em dada teoria, jamais conhecerão demonstrações que os verifiquem ou os refutem. Ou seja, mostraremos que os Teoremas de Incompletude de Gödel, de fato, impactam a atividade matemática cotidiana. Devido às técnicas necessárias para compreender tais resultados, desenvol-

veremos um breve apanhado de teoria dos números ordinais no Apêndice A.

1.1 A PROBLEMÁTICA DOS FUNDAMENTOS

Há algo com que todo aventureiro no mundo da Matemática se depara, geralmente logo no início de seus estudos, e que gera em si uma estranha e quase constante espécie de epifania (ao menos enquanto o aventureiro continuar desejando navegar por esses mares). Quase constante, porque tal ideia — e aqui justifica-se o uso do pronome indefinido no início do parágrafo — está estreitamente relacionada com o *fazer* matemático, ou seja, com o cotidiano de qualquer pessoa que se debruça sobre os conceitos abstratos de números e formas geométricas. E, embora a matemática seja reconhecida, pelo senso comum, como uma área *exata* do conhecimento, a ideia a que estamos nos referindo tem dois aspectos bastante subjetivos: liberdade e incerteza.

Matemáticos se sentem livres não só porque uma mesma teoria matemática pode ser construída de várias formas diferentes; nem apenas pelo fato de a matemática ser uma ciência dedutiva, em que os conceitos podem ser tão abstratos quanto se queira, desde que as sentenças acerca deles se sigam umas das outras por argumentação lógica adequada; mas também porque pequenas mudanças nas escolhas axiomáticas levam a descobertas totalmente distintas — tome como exemplo o axioma das retas paralelas na geometria euclidiana, cuja modificação faz surgir geometrias menos intuitivas, mas igualmente consistentes e úteis. Por outro lado, matemáticos passam boa parte de seu tempo incertos sobre as proposições que conjecturam (quer essa incerteza seja quanto ao modo como mostrá-las, ou simplesmente quanto à real validade ou não das mesmas); fazer matemática é, *grosso modo*, sugerir e (tentar) demonstrar teoremas.

Agora, o leitor deve notar que as sensações de liberdade e incerteza, embora não sejam contraditórias em essência, divergem no seguinte ponto: enquanto a primeira reflete o poder incomensurável da Matemática e nos dá um sem-fim de construções teóricas possíveis, a segunda atua exatamente como um *limitador* em nossa razão. Não há como ser totalmente livre quando há espaço para dúvidas. De certa forma, no entanto, o que se viu ao longo da história recente (ao menos a partir do séc. XIX, quando a matemática moderna nasceu) foi que questões levantadas por uma das sensações gerava respostas da outra, e vice-versa. Por um lado, foram justamente a dificuldade, por parte dos matemáticos, em se captar o axioma das paralelas, bem como as tentativas de mostrá-lo redundante, que levaram à prova de sua independência dos demais axiomas, donde surgiram posteriormente as geometrias riemanniana e hiperbólica. Por outro lado, dando mais um exemplo da geometria[1], foi graças ao

[1] Até o séc. XIX, a geometria era praticamente a única área da Matemática bem fundamentada como uma teoria axiomática.

desenvolvimento da álgebra abstrata — em particular, da teoria de equações algébricas — que extinguimos nossas dúvidas sobre os três clássicos problemas gregos, i.e., que soubemos ser *impossível*, via métodos construtivos com régua e compasso, (1) trissectar um ângulo; (2) dobrar um cubo, i.e., encontrar a raiz cúbica do número dois; e (3) construir um quadrado com mesma área que um dado círculo. Em suma: ao nos questionarmos sobre a força de um axioma específico, fizemos surgir classes completamente diferentes de geometrias; e usando de ferramentas totalmente abstratas da álgebra moderna, fruto da liberdade de criação matemática, pudemos responder questões sobre os limites dessa tal liberdade.

A insolubilidade dos três problemas gregos, no entanto, foi apenas a primeira de muitas respostas obtidas sobre questões desse tipo. O séc. XIX, em geral, foi marcado por outras tantas ideias revolucionárias. Talvez a mais impactante tenha sido a prova de Georg Cantor (1845-1918) de que há diversos tamanhos de infinitos — o que, por sua vez, fez cair por terra uma das verdades absolutas mais antigas do pensamento filosófico: de que o todo é maior do que as partes.[2] A ideia, fruto de pesquisas iniciadas por J. W. Richard Dedekind (1831-1916) e Karl Weierstrass (1815-1897) no campo da Análise Matemática (área esta que veio para tentar explicar os fundamentos sobre os quais o cálculo diferencial se sustenta), foi recebida com assombro por grande parte da comunidade matemática na época, mas posteriormente seus métodos se mostraram perfeitamente práticos e úteis, desencadeando um desenvolvimento teórico nunca antes visto. Várias novas áreas de pesquisa surgiram, com níveis de abstração e rigor cada vez maiores; a matemática não mais servia com o único propósito de representar o mundo que vemos ou sentimos, como a geometria euclidiana e a aritmética dos números inteiros — e até mesmo teorias menos elementares, como o cálculo e a análise combinatória — fizeram por séculos.

De maneira natural, portanto, os matemáticos passaram a se preocupar cada vez mais com questões concernentes à consistência das teorias que estavam construindo. Usando os termos informais da ideia que introduzimos anteriormente, pode-se dizer que tamanha liberdade criativa virou uma pulga atrás da orelha da comunidade matemática quanto às incertezas sobre o que se estava criando.

Uma das primeiras tentativas de estabelecer consistência foi a criação de modelos, i.e., via métodos de interpretação. Por exemplo, se traduzirmos *esfera* como sendo *plano euclidiano*, e *círculos máximos da esfera* como *retas no plano*, estaremos interpretando a geometria riemanniana nos moldes da geometria clássica. Supondo que a última é consistente — e, nesse caso, é natural fazer tal hipótese, pois é a geometria

[2] Cantor é o responsável pelo lugar-comum, na matemática moderna, de que toda entidade matemática pode ser reduzida a um conjunto; grande parte de sua carreira foi devotada à construção de uma teoria dos conjuntos em que se pudesse formalizar o tratamento com o infinito, e, dessa forma, abarcar toda a Matemática. No teorema referenciado, ele mostrou, de maneira mais geral, que qualquer conjunto (finito ou não) é menor que a coleção de todos os seus subconjuntos.

que exprime o mundo real, observável pelos nossos olhos —, estabelecemos também a consistência da primeira. Outro resultado conhecido é devido a David Hilbert (1862-1943), que demonstrou a consistência da geometria euclidiana via geometria analítica — ou seja, sob a hipótese de a teoria dos números reais ser consistente. Perceba, no entanto, que tais resultados não são absolutos — para provar a consistência de uma teoria, assumimos que a teoria modelo é consistente. Do ponto de vista epistemológico, tais demonstrações são insuficientes. (Uma forma suficiente de se determinar consistência, sem assumir a consistência da teoria modelo, seria encontrando modelos *finitos* para essas teorias, já que interpretações finitas podem ser inspecionadas em um número finito de passos. Infelizmente, são poucas as teorias matemáticas que admitem tais modelos.)

Respostas mais concretas, entretanto, vieram somente depois de algumas descobertas, nada agradáveis, fazerem as fundações da Matemática implodirem. Primeiro, o próprio Cantor descobriu uma antinomia dentro de sua teoria dos conjuntos, ao considerar a existência do conjunto de todos os conjuntos (e também encontrou contradições ao definir o conjunto de todos os ordinais). Felizmente, tal inconsistência foi contornada ao se distinguir os conceitos de *classe* e *conjunto* (em algumas áreas mais complexas da Matemática, como topologia e teoria de módulos, tal distinção é crucial). Mas foi o paradoxo de Bertrand Russell (1872-1970), descoberto pelo mesmo em 1901 enquanto estudava os trabalhos de Gottlob Frege (1848-1925) sobre a formalização da lógica simbólica, que realmente colocou a comunidade matemática em alerta. O que Russell fez foi construir um conjunto, M, cujos elementos seriam conjuntos que não pertencem a si mesmos, i.e., o conjunto

$$M = \{X : X \notin X\}.$$

A construção de M, propriamente, não possui incongruências: na verdade, a maioria dos conjuntos que podemos pensar pertencem a M (e.g., o conjunto de todos os números inteiros positivos não é, ele mesmo, um número inteiro positivo). O absurdo acontece quando nos perguntamos se $M \in M$. Caso seja verdade que $M \in M$, tal conjunto deve satisfazer a relação que define os elementos de M, i.e., deve ocorrer $M \notin M$, o que contradiz nossa suposição. Via *reductio ad absurdum*, descartamos essa hipótese, o que nos dá $M \notin M$, ou seja, que M satisfaz a dada condição para pertencer a M, logo $M \in M$. Assim, concluímos que $M \in M$ e $M \notin M$, o que é absurdo. Essa antinomia se popularizou sob a forma do *paradoxo do barbeiro*, cuja anedota fala de uma certa vila onde um barbeiro faz a barba de todas (e somente) aquelas pessoas que *não* fazem sua própria barba. A contradição surge quando nos perguntamos: “Esse barbeiro faz sua própria barba?”

Algo de suma importância no paradoxo de Russell é que, embora não haja contradições em sua concepção (como já apontamos), ele surge justamente pelo fato

de as bases axiomáticas da teoria dos conjuntos e da lógica simbólica (de Frege) permitirem tal construção. Descobriu-se que o problema estava, *grosso modo*, na ideia de que um predicado com uma única variável determina um conjunto, e vice-versa — uma ideia perfeitamente correta, segundo nossa intuição. De fato, os matemáticos se acostumaram a representar conjuntos dessa maneira, e ela sempre pareceu bastante óbvia e conveniente — é muito mais simples e eficaz falar do conjunto P dos números pares como sendo a coleção de todos os números que são o dobro de algum outro, ou seja, escrevendo $P = \{p \in \mathbb{N} : \exists n \in \mathbb{N}(p = 2n)\}$, do que fazer uma listagem, passível de ambiguidades, como $P = \{0, 2, 4, ...\}$. Ficou evidente, portanto, após a descoberta desse paradoxo, que os problemas estavam impregnados até mesmo nas práticas mais usuais do cotidiano matemático[3], o que pôs em xeque tudo que vinha sendo construído nos últimos anos. A necessidade de revisar nossas crenças — que, desde as descobertas de Cantor sobre o infinito, vinham sofrendo forte abalo — havia se tornado premente.

1.2 O PROGRAMA DE HILBERT

Naturalmente, tais problemas, capazes de fazer desabar toda a edificação da Matemática, geraram uma discussão bastante plural (e acalorada) sobre o que fazer para corrigi-los. Embora diversas ideias tenham surgido, prevaleceram três correntes filosóficas, que, inclusive, estavam demarcadas geopoliticamente: o logicismo (inglês) de Russell e Alfred Whitehead (1861-1947); o intuicionismo (franco-holandês) de Luitzen E. J. Brouwer (1881-1966); e o formalismo (alemão) de Hilbert.[4]

A corrente logicista se baseava na ideia de que a Matemática é (apenas) um ramo da Lógica, algo que desde Frege já vinha sendo discutido — ou muito antes, uma vez que G. W. Leibniz (1646-1716) defendia tal tese (cf. (KLEENE, 1971), p. 43) —, e que culminou num trabalho monumental escrito por Russell e Whitehead: o famoso *Principia Mathematica*, uma tentativa de mostrar cada teorema da aritmética como um teorema lógico. O problema é que a maioria das definições lógicas presumia um conceito matemático (e.g., se definirmos *número natural* como um número cardinal para o qual vale o princípio de indução); ao definirmos, portanto, uma ideia matemática por meio de lógica simbólica pura, muitas vezes entramos num círculo vicioso — o que, no paradoxo do barbeiro, parece ser justamente a fonte de contradição.

A solução encontrada pelos autores foi construir uma nova teoria, chamada de

[3] Na verdade, posteriormente Russell conseguiu provar que o problema era inerente à própria lógica: chamando de K a propriedade de *ser uma propriedade que não se aplica a si mesma*, o paradoxo surge ao nos perguntamos se K se aplica a si mesma.

[4] Lembre-se que o final do séc. XIX e início do séc. XX foram marcados por disputas imperialistas na Europa, que culminaram nas duas grandes guerras mundiais. Os congressos e as discussões matemáticas da época, sendo a problemática dos fundamentos uma delas, refletiam essa disputa, como aponta (SMORYŃSKI, 1988).

teoria dos tipos, segundo a qual se classificariam as entidades lógicas necessárias para as bases matemáticas. No entanto, tal construção se baseava em dois axiomas que eram, senão contra-intuitivos, ao menos bastante convenientes. (A saber, o *axioma da reducibilidade*, que impedia definições circulares de *tipos*, e o *axioma do infinito*, que postulava a existência de ao menos um conjunto infinito; digamos, aquele que contém todos os números naturais.) Por conta desse tipo de conveniência, e também por causa da dificuldade em se traduzir todo conceito e teorema matemático em suas versões lógicas (complicando demasiadamente noções bastante simples), pode-se dizer que os logicistas permaneceram à margem das discussões, e que estas foram protagonizadas por intuicionistas e formalistas. De fato: embora a proposta de redução da Matemática à Lógica soasse bastante ousada, as restrições intuicionistas tinham um caráter muito mais geral.

Como o nome da escola de Brouwer sugere, as únicas fontes de inspiração para as construções matemáticas deveriam ser as abstrações mais ordinárias de nosso intelecto — como o ato de contar, por exemplo. Segundo (HEYTING, 1934), "De acordo com Brouwer, a Matemática é idêntica à parte *exata* do nosso pensamento. (...) Nenhuma ciência, em especial nem a lógica nem a filosofia, podem ser um pressuposto para a Matemática". Há de se destacar, no entanto, os pensamentos de alguns que vieram antes de Brouwer, como Leopold Kronecker (1823-1891) e Henri Poincaré (1854-1912). Ambos apontaram os métodos envolvendo quantidades infinitas como sendo a raiz de todo o mal — Poincaré, inclusive, sempre foi um crítico ferrenho à teoria dos números cardinais de Cantor, muitas vezes usando de deboche para tecer tais críticas — cf. (ROGERS, 2015). Na visão de Kronecker, os números inteiros eram uma espécie de obra divina, e o processo de indução, em $\mathbb{Z}$ ou $\mathbb{N}$, era visto como a única forma perfeitamente intuitiva de se tratar o infinito; todo o resto da Matemática, portanto, deveria ser construído na base de tal estrutura — cf. (MURAWSKI, 1999). Poincaré, também, apontava que a fonte de contradição em todos os paradoxos e antinomias estava nas definições e condições impredicativas[5] que nossa lógica e teoria conjuntista permitiam. Brouwer, no entanto, juntou tudo isso e mais um pouco, e propôs um conjunto profundamente drástico de soluções.

A doutrina encabeçada pelo matemático holandês rejeitava, sobretudo, métodos não construtivos — em particular, demonstrações via *reductio ad absurdum* de senten-

[5] Por *definição (condição) impredicativa* entende-se toda definição (condição) autorreferente, i.e., que faz referência a uma totalidade que pressupõe o próprio conceito que está sendo definido. No paradoxo de Russell, o conjunto *M* é especificado por uma tal condição. Deve-se notar, entretanto, que nem toda definição impredicativa leva a uma contradição, como Poincaré argumentava; na axiomática de ZFC, há esquemas de axiomas em que a substituição de um predicado **A** por um impredicativo não gera paradoxos (para um detalhamento de como a axiomática conjuntista solucionou problemas de impredicatividade, alterando as regras de como definir conjuntos por especificação, veja (SUPPES, 1972), ou (ROGERS, 2015)). Também, a definição de $\mathbb{N}$ como 'o *menor* conjunto contendo o zero e fechado pela operação sucessor' é um conceito impredicativo — pois alude à totalidade de conjuntos que satisfazem tais condições —, mas não é paradoxal, cf. (MURAWSKI, 1999)

ças existenciais, algo usado em praticamente qualquer teoria matemática.[6] Novamente citando Heyting, "Na matemática intuicionista, não se faz inferências de acordo com regras fixadas, que podem ser compiladas numa certa lógica, mas, sim, cada inferência é imediatamente testada com base em suas evidências". Além disso, somente o conceito de infinito enumerável poderia ser aceito, e toda teoria deveria ser edificada única e exclusivamente sobre a base nos números naturais (pois esta é a construção mais primitiva e intuitiva do pensamento matemático). Assim, de maneira geral, a principal objeção dos intuicionistas se dava quanto à intangibilidade de conjuntos infinitos não-enumeráveis. Especificamente, entretanto, o foco das críticas de Brouwer sempre foi a lei do terceiro excluído — i.e., dada qualquer sentença A, então ou é verdade que A, ou é verdade que $\neg A$; no caso de domínios infinitos (enumeráveis ou não), seria impossível percorrer tal caminho, infindável, a fim de determinar o valor de verdade de um sentença B, já que os métodos de verificação deveriam ter caráter construtivo, i.e., que deveríamos mostrar uma forma concreta de estabelecer a verdade de B.

Para um intuicionista, portanto, a Matemática Moderna se divide em duas partes completamente distintas: a Matemática *Clássica*, que aceita todo conceito de infinito e demonstrações por contradição; e a Matemática *Intuicionista*, que aceita somente os princípios de indução finita e métodos construtivos de prova. A pergunta que naturalmente surge é: até que ponto vai a capacidade da última, imposta a tamanhas restrições? Ou seja, em que contextos podemos nos sentir confortáveis por termos construído uma teoria palpável, em que cada verdade pode ser efetivamente verificada, de modo que podemos, de fato, rejeitar uma construção clássica equivalente?

Essa é uma questão crucial no embate entre intuicionistas e formalistas. Caso a matemática clássica pudesse ser completamente substituída por métodos construtivos, a problemática dos fundamentos estaria resolvida. Por um lado, Hilbert reconhecia o papel do infinito e da autorreferência no surgimento dos paradoxos; a tarefa primordial de seu Programa, como veremos, era encontrar um processo não-infinito de se estabelecer a consistência da aritmética; no mesmo sentido, viu-se que a maioria das provas não-construtivas da teoria de números podia ser substituída por versões construtivas — cf. (KLEENE, 1971). Mas, por outro lado, a Análise Matemática, i.e., a teoria dos números reais, jamais conheceu uma construção tão completa quando a de Weierstrass e Dedekind. Na abordagem proposta por Hermann Weyl (1885-1955), em que que se substitui a definição de cortes (impredicativa por natureza) por um método construtivo e não-circular, embora se tenha alcançado em partes o objetivo, jamais se conseguiu provar um dos resultados mais básicos de tal teoria: o de que todo conjunto limitado

[6] Outra ideia matemática supostamente imprescindível mas permeada de dúvidas, não só da parte dos intuicionistas, como, também, por teóricos da axiomática conjuntista, é o Axioma da Escolha, segundo o qual dada qualquer coleção de conjuntos, existe uma maneira (à qual se dá o nome de *função escolha*) de se obter exatamente um elemento de cada um desses tais conjuntos. Dentre os resultados mais fortes (e elementares) que se sustentam sobre esse axioma, destaca-se o fato de todo espaço vetorial possuir base.

de números reais possui supremo (i.e., uma menor cota superior).

Para compreender melhor a batalha entre as escolas de Hilbert e Brouwer, é importante notar que uma das principais discordâncias entre os dois se dava quanto a seus interesses acadêmicos: enquanto o segundo sempre se interessou por questões filosófico-matemáticas, a carreira do primeiro, conquanto uma das mais prolíficas e universais dos últimos tempos, ficou marcada pelo rigor formal do método axiomático; em seu trabalho mais famoso, *Grundlagen der Geometrie*, ou *Fundamentos da Geometria*, ele literalmente termina a obra de Euclides (sob a ótica da matemática moderna). Segundo (SMORYŃSKI, 1988), "Hilbert (...) tinha pouca paciência com filosofia, podendo sua própria Filosofia Matemática ser descrita como um otimismo ingênuo — uma espécie de *fé* nas habilidades de um matemático de resolver qualquer problema com que se compromete". Na problemática dos fundamentos, tal oposição de interesses se materializou da seguinte forma: Brouwer buscava *revisar* o modo como tratávamos algumas questões peculiares, como o infinito e a real existência de outras entidades pouco intuitivas; ao passo que o objetivo de Hilbert era *salvar* a matemática clássica — uma vez que determinássemos a consistência da mesma, não precisaríamos abdicar de coisa alguma.[7] Dessa forma, podemos perceber uma grande diferença do formalismo para as duas demais correntes filosóficas: logicistas e intuicionistas buscavam novas maneiras de descrever as teorias matemáticas, trabalhando *dentro* delas, propondo novas bases e delimitando conceitos problemáticos; a consistência que Hilbert procurava, no entanto, fazia parte de um universo *exterior* a tais teorias, em que se provam sentenças *sobre* as mesmas.

Assim, num processo que germinou a partir de casos como a insolubilidade dos três problemas gregos e a criação de modelos, os formalistas agora buscavam um resultado que englobasse todas as teorias possíveis, procurando a consistência matemática *em si*. Ao tratar a Matemática como linguagem-objeto (ou *teoria-objeto*), o formalismo criou uma nova área de conhecimento, à qual Hilbert deu o nome de Teoria da Prova (*Beweistheorie*), e que posteriormente ficou conhecida como *Metamatemática*.

Aqui, poderíamos entrar numa longa discussão sobre *de que maneira* os formalistas achavam ser possível tratar as teorias matemáticas como um todo, a ponto de se conseguir inferir respostas tão gerais sobre elas. No entanto, além de essa ser, em parte, uma das finalidades dos próximos capítulos — por exemplo, dedicaremos um bom tempo mostrando de que forma pode-se definir matematicamente *o que é uma prova* —, acreditamos que a maioria dos detalhes técnico-filosóficos do Programa de Hilbert foge de nosso escopo. Cabe elencar, todavia, duas particularidades do método formalista.

[7] Citando o próprio personagem em questão, "Ninguém deveria ser capaz de nos tirar do paraíso que Cantor criou", cf. (HILBERT, 1926).

A primeira se assemelha à ideia reducionista presente no logicismo; em sua tentativa de dar um sentido claro ao conceito intuitivo de *demonstração*, Hilbert procurou descrever formalmente toda a Matemática (inclusive sua parte infinita), baseando-se no método axiomático. Como consequência, além, de se poder falar de quantidades infinitas, tornava-se possível despir de semântica todas as entidades matemáticas, prevalecendo somente a *forma*. (Como num jogo de tabuleiro qualquer, reduzimos nossa teoria-objeto a um conjunto de peças sem significado, sujeitas às regras que bem escolhermos.)

O segundo aspecto, em contrapartida, diz respeito à metateoria — mais especificamente, ao modo como abordar o infinito. Como já frisamos anteriormente, Hilbert reconhecia as objeções intuicionistas de Brouwer, mas insistia em manter ilesa toda teoria matemática (ou seja, toda teoria-objeto) que apelasse a tais conceitos. Ele propôs, então, uma distinção entre métodos *finísticos* e *infinísticos*. Dentro da metateoria, apenas provas e construções finísticas poderiam ser aceitas.[8] A princípio, uma vez formalizada a teoria-objeto, seria totalmente viável trabalhar apenas com métodos finísticos — se pensarmos em fórmulas como listas finitas de símbolos, e provas como listas finitas de fórmulas, fica claro que não precisaríamos apelar a nenhum conceito infinito (ao menos na metateoria).

O Programa de Hilbert, propriamente dito, consistiu em usar os métodos que acabamos de sumarizar (e tantos outros pormenores omitidos), na esperança de se obter várias respostas metateóricas, dentre as quais a consistência era apenas uma — embora sempre tenha sido considerada a questão primordial. Em 1900, no Congresso Internacional de Matemática, Hilbert já colocava tal problema (aqui considerando somente a consistência dos axiomas da aritmética) como o número 2 de uma lista de 23 desafios matemáticos que precisavam ser superados.[9] Durante as duas décadas que se seguiram, tais ideias foram sendo aprimoradas, resultados positivos foram (presumidamente) obtidos[10], e cada vez mais o matemático alemão se convencia de que a

[8] Distinção essa que nunca ficou totalmente clara (e esse é mais um dos motivos para não alongar demais tal discussão). Em linhas gerais, objetos finísticos da Matemática são todos aqueles não-problemáticos, concretos, que não aludem aos conceitos rejeitados pela matemática intuicionista; de maneira ilustrativa, não há dúvidas que o conjunto $\{\mathbb{N}, \mathbb{Z}, \mathbb{R}\}$ é *finito*, pois tem três elementos. Mas ele não é *finístico*.

[9] Dentre eles, figuravam não apenas questões fundacionais. O de número 6, por exemplo, diz respeito à axiomatização da Física (o que não deixa de ser, na visão de Hilbert, uma extensão dos problemas que aparecem nas geometrias, já que modelos físicos são essencialmente geométricos). Um exemplo menos ligado aos fundamentos da matemática é o de número 8, que versa sobre a hipótese de Riemann e suas consequências. Todavia, o topo da lista tem caráter naturalmente fundamentalista: o primeiro e mais importante de todos é a solução da hipótese do *continuum* (que, hoje, já sabemos ser algo impossível em ZFC).

[10] Ao menos em partes. Wilhelm Ackermann (1896-1962), pupilo de Hilbert, conseguiu provar a consistência de um fragmento da teoria de números, em 1924-25.

vitória sobre Brouwer estava próxima.[11]

Ninguém esperava, no entanto, que um jovem lógico, antes de completar seus 25 anos, respondesse a dois dos principais problemas enumerados por Hilbert, e muito menos que uma dessas respostas fosse no sentido contrário ao que o Programa visava. Em sua tese de doutorado, Kurt Gödel (1906-1978) provou que a lógica de primeira ordem é completa, i.e., que nela é possível demonstrar todas suas sentenças válidas — algo que, dentre os principais objetivos metamatemáticos apontados por Hilbert no congresso de 1928, era considerado o resultado mínimo esperado, já que grande parte das teorias matemáticas tem tal lógica como camada mais inferior. No sentido totalmente contrário, entretanto, Gödel também provou a existência de sentenças indecidíveis dentro da aritmética dos números naturais, bem como em *qualquer* extensão da mesma (supondo consistentes tais aritméticas); ou seja, na mais elementar das teorias matemáticas, há proposições que nunca conseguiremos demonstrar nem refutar, mesmo elas sendo verdadeiras. Ademais, como corolário, ele concluiu que nenhum tal sistema é capaz de provar sua própria consistência (e provar tal coisa era a principal finalidade ao se tratar a Matemática como teoria-objeto!).

Resultados extraordinários como esses colocavam em xeque não só a ideia formalista de que a consistência de um sistema implica a existência de uma interpretação para o mesmo — cf. (MURAWSKI, 1999). Diante de tal descoberta, se tornava real a possibilidade de qualquer conjectura, em qualquer campo da Matemática, *não poder ser decidida*. Dessa forma, muitos dos problemas em aberto poderiam (ironicamente) permanecer, *para sempre*, em aberto — e isso invalidaria áreas de pesquisa inteiras, como as decorrentes da assunção da hipótese de Riemann, por exemplo. De certa maneira, portanto, o próprio fazer matemático foi impactado por tal descoberta.

Acontece que a sentença indecidível de Gödel, embora construída *dentro* da aritmética, tem significado *meta*matemático, diferente da maioria das conjecturas que ocorrem nas diversas teorias matemáticas. Mais especificamente, embora ainda de uma maneira bastante informal, tal sentença dizia que ela mesma não é demonstrável dentro do sistema. Deste modo, a princípio, o que seus teoremas provavam era apenas a *possibilidade* de um problema em aberto puramente matemático ser insolúvel. Nosso objetivo principal no decorrer dos próximos capítulos, além de formalizar a metateoria e de mostrar como construir algumas sentenças aritméticas indecidíveis, é esclarecer até que ponto os teoremas de incompletude realmente afetam a atividade rotineira

[11] A bem da verdade, no final dos anos 1920 já não havia mais uma *disputa* entre formalismo e intuicionismo: a maioria da comunidade matemática concordava com o método axiomático não-construtivista, e ansiava pela prova absoluta da consistência. Tal batalha, todavia, tinha um cunho pessoal para Hilbert — cf. (SMORYŃSKI, 1988)

do matemático comum[12] — i.e., se de fato alguma espécie de "receio", quanto à insolubilidade de um problema, é justificável quando conjecturamos proposições dentro das mais diversas áreas em que atuamos, e não só no campo metateórico.

[12] Por "matemático comum", estamos designando a maioria da comunidade matemática, que em suas pesquisas geralmente não está preocupada com as questões fundacionais e filosóficas dessa área do conhecimento. Em particular, o matemático comum, quando se depara com a sentença que Gödel usou para provar seus teoremas, não a reconhece como um objeto de estudo relevante, e tende a ignorar o impacto da incompletude.

2 ELEMENTOS DE TEORIA DA RECURSÃO

No caminho para se formalizar tanto a metamatemática quanto sua teoria-objeto, de modo que possamos construir sentenças de cunho metateórico dentro da linguagem comum da aritmética, primeiramente será necessário dar uma definição precisa para algumas ideias que permeiam o conceito central do Programa de Hilbert, que é a noção intuitiva de *prova*. O objetivo principal deste capítulo é apresentar, em específico, duas maneiras equivalentes de se abstrair aquilo que chamamos, informalmente, de *processo efetivo* (ou *mecânico*) de decisão. Para isso, nos basearemos principalmente nas obras de (BOOLOS, G. *et al.*, 2007), (ROGERS JR., 1987), e (MURAWSKI, 1999). Tais abstrações deverão obedecer às imposições *finísticas* do método formalista, i.e., nosso campo de trabalho deve se restringir àquilo de mais intuitivo no pensamento matemático: os números naturais e o processo de indução finita.

2.1 MÁQUINAS DE TURING

Ainda no campo da informalidade, comecemos nossa discussão pelo básico: quando dizemos que uma função nos (ou relação entre) números naturais pode ser *computada*[1] via um processo efetivo — ou, como também o chamaremos daqui em diante, via um *algoritmo* —, quais atributos um tal processo deve ter? Talvez a melhor maneira de responder a essa questão seja analisando algum algoritmo que conhecemos. É possível que o leitor se lembre, prontamente, da divisão euclidiana (já que tal processo geralmente carrega o nome *algoritmo* junto a ele), mas a aritmética elementar está recheada deles — e.g., o *crivo de Erastótenes* lista uma quantidade finita de números primos tão grande quanto se queira, e as próprias operações de adição e multiplicação são processos mecânicos, que aprendemos desde a escola e se tornam praticamente intrínsecos ao nosso saber. O algoritmo de Euclides tem algumas características que podemos elencar como universais nos processos de decisão: (i) tal processo pode ser caracterizado como uma lista *finita* de instruções, que algum agente externo (digamos, um jovem matemático estudando álgebra dos inteiros) deve seguir; (ii) há maneiras de guardarmos e reutilizarmos informação sobre cada etapa executada (de fato: é isso que se faz quando dividimos um número por outro e analisamos seu resto, sucessivamente); e (iii) o ato de passar de uma instrução à próxima tem caráter determinístico e discreto, i.e., não temos de apelar para a sorte nem para métodos contínuos.

É importante notar que tais atributos têm seus paralelos com qualquer máquina computacional conhecida: o primeiro item reflete a noção de *programa* e suas bases

[1] Perceba como a informalidade nos leva a conceitos cada vez mais cíclicos: aqui, introduzimos a ideia de se computar uma função, i.e., *encontrar* seu valor para uma entrada qualquer, mas tal ideia carece, por ora, como o resto dos conceitos citados, de qualquer formalização.

lógicas e físicas que permitem processá-lo; o segundo item é análogo ao conceito de *memória* de armazenamento; e o determinismo discreto presente no terceiro item é um paralelo com as propriedades *digitais* e *mecânicas* dos computadores. O paralelo existe porque o conceito de *computador* é o mesmo em ambas as situações (um humano estudando álgebra ou uma máquina digital rodando um programa que implementa o algoritmo da divisão). A bem da verdade, a ciência da computação teve seu advento nas ideias de Alan Turing (1912-1954), ao se abstrair processos efetivos de decisão para funções nos inteiros positivos. O que hoje chamamos de uma *máquina de Turing* nada mais é do que a idealização de um aparato capaz de computar, de maneira mecânica, tais funções.

Podemos imaginar um tal mecanismo como sendo uma fita, infinita em ambas as direções, dividida em quadrados, e por cima da qual um agente (computador) pode se mover à direta ou à esquerda (o que representaremos por R e L, respectivamente), mas um quadrado de cada vez. De maneira ilustrativa, temos o seguinte:

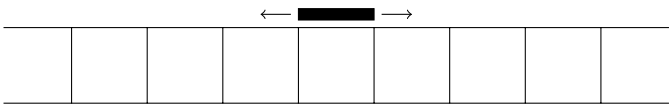


Figura 1 – Computador se movendo sobre uma fita.

Tal agente inspeciona o conteúdo presente no quadrado sobre o qual está no momento, podendo inserir (ou apagar) um símbolo nele. Cada quadrado pode estar em branco ou marcado com um traço; representaremos o primeiro por S_0, ou B, e o segundo, por S_1, ou 1.

Além disso, em cada estágio da computação, o operador da máquina precisa receber uma instrução do que fazer seguida, baseado no que ele fez antes e no símbolo que está observando, de modo que sejam estabelecidas formalmente as características determinística e discreta de um algoritmo. Podemos fazer isso definindo uma quantidade finita de *estados*, digamos, $q_1, q_2, ..., q_m$, em que a máquina pode se encontrar — e aqui, *estado* pode significar várias coisas que digam respeito aos aspectos da estrutura interna e do funcionamento da máquina em dado momento (digamos, algum parâmetro eletrônico ou mecânico), mas isso não importa para a nossa discussão. O que importa é que consigamos especificar cada instrução tão-somente pelo estado atual da máquina e o símbolo avaliado em tal estado, i.e., que *no estado q_i, siga-se a instrução i*.

Em suma, portanto, há cinco coisas que, diante de uma instrução, o agente pode fazer: (1) inserir S_0 no lugar de qualquer símbolo analisado (ou seja, apagar um traço ou nada fazer, caso o quadrado já esteja em branco); (2) inserir S_1 no lugar de qualquer símbolo analisado (ou seja, escrever um traço ou nada fazer, caso já haja um traço no quadrado); (3) mover-se um quadrado à direita; (4) mover-se um quadrado à esquerda; e (5) parar a computação. Assim, dados quaisquer estado e análise de conteúdo do

quadrado, o agente realizará uma dessas cinco ações explícitas; no entanto, para que a computação prossiga, também é necessário haver, na instrução, uma ação *implícita*, a saber, para qual estado a máquina deve ir depois de executar a ação explícita. Ou seja, de modo geral, o computador entra num estado, q_r, observa um dos símbolos (S_0 ou S_1), executa uma das ações (R, L, inserir S_0 ou S_1), e vai para um outro estado, q_s. [Perceba que dizer que a computação *parou* significa que a máquina entrou num estado para o qual não há instrução do que se fazer em seguida, independetemente do símbolo analisado.]

Há várias formas de representar o *programa completo de instruções* para uma máquina de Turing. Por exemplo, podemos fazê-lo via diagramas de fluxo; a Figura 2 representa um programa capaz de marcar três traços na fita:

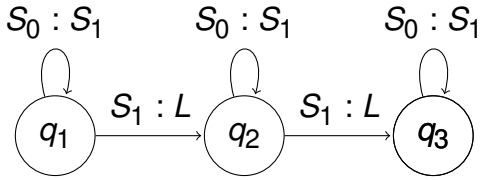


Figura 2 – Máquina de Turing representada por diagrama de fluxo

De fato: como a fita está inicialmente toda em branco, o computador vai reconhecer um quadrado em branco no estado q_1, depois irá marcá-lo com um traço, e voltar para o estado q_1. Daí, lendo um traço no quadrado que anteriormente estava em branco, segue a instrução de se mover à esquerda, onde entra no estado q_2. O mesmo processo se repete para este segundo quadrado, e chegando no estado q_3, a máquina simplesmente insere um traço no mesmo (sem receber uma instrução do que fazer quando ler um traço em tal quadrado; ou seja, tal máquina deve parar aí sua computação).

Esse tipo de representação é bastante intuitivo e carrega o caráter mecânico do que estamos descrevendo, mas a representação por *quádruplas*, que sintetiza o procedimento geral descrito no parágrafo anterior, é uma forma bem mais simples. Quando formos especificar uma máquina de Turing, o faremos por esse método. Assim, o mesmo programa poderia ser descrito pela seguinte lista de instruções:

$$q_1 S_0 S_1 q_1$$
$$q_1 S_1 L q_2$$
$$q_2 S_0 S_1 q_2$$
$$q_2 S_1 L q_3$$
$$q_3 S_0 S_1 q_3$$

Fica subentendido, pela forma que listamos as quádruplas, que a máquina inicia o processo pelo estado de menor número, e essa é a prática comum em teoria da recursão. Mas nada impede que a computação comece em outro estado (lembre-se:

estado é somente uma descrição da atividade interna dos mecanismos que efetuam a computação). Vejamos, agora, um exemplo um pouco menos trivial; até o momento, fomos capazes de descrever uma máquina que abstraísse o processo mais intuitivo da Matemática, i.e., o ato de contar (embora, no exemplo acima, tenhamos contado somente até 3; mas é relativamente simples de ver como podemos estender a ideia para qualquer $n \in \mathbb{N}$).

Afirmamos, dessa vez, que existe uma máquina de Turing que computa a função adição, $f(x, y) = x + y$, da seguinte maneira: dado $(r, s) \in \mathbb{N} \times \mathbb{N}$, inicialmente a fita está completamente em branco, exceto por dois blocos de traços, um à esquerda, com r traços, e outro à direita, com s traços, ambos separados por um espaço em branco. A máquina começa analisando o traço mais à esquerda do bloco com r traços, e ela para analisando o traço mais à esquerda de um bloco com $r + s$ traços. Com efeito: tudo que a máquina precisa fazer é apagar o traço mais à esquerda, se mover à direita na fita, procurando pelo espaço em branco, preenchê-lo, e voltar para o (novo) traço mais à esquerda (que antes era o segundo traço mais à esquerda). Uma maneira de fazer isso é listando as seguintes instruções: $q_1 S_1 S_0 q_1$, $q_1 S_0 R q_2$, $q_2 S_1 R q_2$, $q_2 S_0 S_1 q_3$, $q_3 S_1 L q_3$, $q_3 S_0 R q_4$. Ilustrativamente, tome $(r, s) = (2, 3)$, ou seja, computemos $2 + 3 = 5$. Num primeiro momento, temos a seguinte configuração:

$$11B111.$$

Seguindo a primeira instrução, ficamos com $B1B111$, ao que passo as próximas três nos dão $B11111$, e as duas últimas apenas nos fazem encontrar aquele B supérfluo, determinando o novo traço mais à esquerda, quando enfim obtemos

$$11111.$$

Aqui, há de se fazer uma observação: nem nossa listagem de instruções, nem o exemplo prático de somar 2 com 3 constituem, a rigor, provas de que f é computável por tal máquina. A princípio, o leitor pode facilmente se convencer, apenas *imaginando* dois blocos arbitrários de traços, separados por um espaço em branco, que tal lista é suficiente (e realmente é!), porque o procedimento, em si, é muito simples (apagar um traço e preencher o espaço entre os dois blocos); para casos mais complexos, no entanto — como o da multiplicação, ou até mesmo de $f(x) = 2x$ —, com certeza não seria fácil convencer alguém unicamente por meio de uma lista de instruções ou um exemplo, como foi feito acima. Muito provavelmente, o leitor não conseguirá se convencer, sem fazer alguns exemplos ou apelar para indução, que uma máquina capaz de dobrar uma quantidade qualquer de traços pode ser representada pelo seguinte conjunto (extenso) de (vinte e uma) quádruplas:

$$
\begin{array}{ll}
q_1 S_1 L q_2 & q_7 S_1 S_0 q_7 \\
q_2 S_1 L q_3 & q_7 S_0 L q_8 \\
q_2 S_0 L q_3 & q_8 S_1 L q_9 \\
q_3 S_0 S_1 q_3 & q_9 S_1 L q_9 \\
q_3 S_1 L q_4 & q_9 S_0 L q_{10} \\
q_4 S_0 S_1 q_4 & q_{10} S_1 L q_{10} \\
q_4 S_1 R q_5 & q_{10} S_0 R q_2 \\
q_5 S_1 R q_5 & \\
q_5 S_0 R q_6 & q_8 S_0 L S_{11} \\
q_6 S_1 R q_6 & q_{11} S_1 L S_{11} \\
q_6 S_0 L q_7 & q_{11} S_0 R q_{12}
\end{array}
$$

[Na prática, o que tal máquina faz é bastante simples: ela simplesmente escreve dois traços para cada traço que apaga, e é isso que as instruções que começam com $q_1, ..., q_8$ descrevem. No entanto, há uma encruzilhada no oitavo estado, dependendo de qual símbolo é analisado: caso seja S_1, ainda há traços para apagar e dobrar, de modo que as instruções que começam com q_9 e q_{10} fazem o processo se repetir; quando não há mais traços do bloco inicial que necessitam ser apagados, usamos as últimas três quádruplas para finalizar o processo, no traço mais à esquerda de um bloco com o dobro do número de traços que havia no começo.]

De maneira geral, todas as demonstrações mencionadas só seriam efetivas via indução finita (o que seria totalmente viável), mas optamos pela omissão, uma vez que nosso foco, nessa seção, é apenas apresentar de que modo Alan Turing abstraiu procedimentos efetivos de decisão. [Por fim, um último comentário quanto aos nossos métodos: em teoria da computação, podemos definir máquina de Turing de maneira muito mais rigorosa — perceba que nossa abordagem levou algumas páginas, apelando às nossas percepções visuais, i.e., a um modelo físico, diferentemente de uma definição matemática usual. Para os fins do presente trabalho, no entanto, tal formalização é suficiente; caso o leitor tenha interesse por um olhar mais apurado, sugerimos (SIPSER, 2006), e (DAVIS, 1985). Ainda assim, podemos resumir a caracterização de uma máquina de Turing, dentro do formalismo necessário para nossos propósitos, como a lista de quádruplas que define seu comportamento, estabelecendo-se a condição de que toda computação tem tanto seu início quanto seu fim no traço mais à esquerda da fita — se não for o caso de ela estar totalmente em branco em algum desses momentos —; além disso, todo eventual espaço em branco presente no início (que ocorrem na representação das entradas de uma função de múltiplas variáveis) deve estar preenchido quando a máquina parar — caso ela realmente pare.]

Outra questão importante de se notar nas abstrações que estamos sugerindo é sobre a limitação das capacidades dos componentes de tais máquinas. Sendo mais específico: durante nossa caracterização, deixamos implícita a suposição de que, a

princípio, máquinas de Turing podem computar valores de funções para qualquer quantidade finita de entradas, bem como seguir listas de instruções tão grandes quanto se queira — afinal, tais máquinas devem operar sobre uma fita de comprimento infinito; logo, entradas de tamanhos arbitrários podem ser inseridas e manipuladas nela, desde que sejam finitos. Ou seja, quando se investiga a computabilidade de uma determinada função, os paralelos com dispositivos mecânicos e eletrônicos são meramente ilustrativos, pois toda máquina real *possui* limitação, seja quanto à memória ou à capacidade de processamento.

Nesse sentido, chamemos uma função nos inteiros positivos de *Turing-computável* caso haja ao menos uma máquina de Turing capaz de computá-la. Então, é verdade que dada uma função como essa, *existe* um processo efetivo para se determinar todos os seus valores, ainda que seja impraticável, até mesmo para o mais potente dos supercomputadores, realizar tal computação — e mesmo que tal processo leve centenas de milhões de anos para ser finalizado; o que importa é que o número de passos seja finito, e que mostremos, construtiva e indutivamente, como fazê-lo (as discussões sobre o quão mal *otimizado* tal algoritmo seria, nós, lógico-matemáticos, deixamos para os colegas cientistas da computação e da matemática aplicada[2]. De qualquer maneira, obviamente concordamos que Turing-computabilidade *é uma* forma de computabilidade; mas será que é a única?

Coloquemos a questão de forma mais clara. Com certeza, o modo como definimos o processo mecânico de inspecionar compartimentos, e de inserir e apagar informações neles, não é a única forma de se abstrair tal conceito; poderíamos tê-lo feito imaginando, em vez de uma fita, uma malha — bidimensional, portanto —, também infindável, em que o agente poderia se mover com muito mais liberdade, em quatro direções, e eventualmente marcar mais símbolos distintos. Também, seria possível abstrair de maneira mais eficiente o conceito de *memória interna*, algo bastante falho nas máquinas de Turing unidimensionais. Máquinas de Lambek, ou máquinas de *ábaco*, são idealizações que alcançam esse objetivo; ao se inserir entidades chamadas *registradores*, em que se pode armazenar qualquer número e acessá-lo livremente — sem ter de percorrer um quadrado por vez —, se torna muito mais simples especificar algoritmos. Acontece que *qualquer* função computável por ábaco ou por "máquina de Turing em malha" é, também, computável por uma máquina de Turing unidimensional. (Tente se convencer disso pensando no seguinte: ambas as abstrações citadas nada mais são do que "evoluções" da máquina de Turing em fita, mais simples e intuitiva.[3])

[2] A bem da verdade, questões sobre o tempo necessário para a resolução de um problema computacional envolvem muito mais que o conceito de *otimização* de um algoritmo. O principal problema em aberto da ciência da computação é verificar a validade (ou não) da igualdade $P = NP$.) Em suma, isso quer dizer que os problemas computacionais cujas soluções podem ser *verificadas* em tempo que cresce *polinomialmente* também podem ser *resolvidos* em tempo polinomial.

[3] Para uma caracterização detalhada de ábacos e uma prova desses fatos, cf. (BOOLOS, G. *et al.*, 2007), p. 51-55.

Nossa pergunta, então, pode ser reformulada: será que, *essencialmente*, máquina de Turing é o único tipo de processo efetivo de decisão?

O leitor já deve pressentir o anticlímax, mas é *impossível* responder a tal questão, ao menos de acordo com o conhecimento filosófico-matemático produzido até hoje, e ao menos no sentido ortodoxo de uma *prova* matemática. Isso se dá justamente porque máquinas de Turing — ou qualquer outro espécime de abstração de algoritmos — são *tentativas de dar sentido formal* à noção *completamente intuitiva* de procedimento de decisão. Ou seja, qualquer abstração *formal* de um algoritmo é, de fato, um algoritmo no sentido brouweriano, que Hilbert optou por seguir em seu Programa. Há, todavia, bons motivos para nos convencermos de uma resposta afirmativa à questão que temos elaborado. Primeiramente, é possível demonstrar que toda função aritmética, i.e., definível a partir de adição ou multiplicação, é computável por ábacos, do que segue que são Turing-computáveis — e isso dá conta de uma *enorme* quantidade de funções. Por outro lado, todas as funções (que conhecemos) para as quais *não há* uma máquina de Turing que a opere, i.e., todas as funções Turing-*in*computáveis, desconhecem, também, *qualquer* outro método efetivo de decisão. [Para darmos um exemplo de uma tal função, é preciso considerar um detalhe que, facilmente, passa despercebido. Para início de conversa, o fato de *existirem* funções Turing-incomputáveis é, puramente, uma questão de cardinalidade de conjuntos. Com efeito: o conjunto $\mathbb{N}^{*\mathbb{N}^*}$, de todas as funções nos inteiros positivos, é um conjunto não-enumerável (algo que conseguimos provar via diagonalização); em contrapartida, como máquinas de Turing podem ser identificadas com suas listas de quádruplas, que são descritas por uma quantidade finita de símbolos, é possível *listar* todas essas máquinas — i.e., há apenas uma quantidade enumerável de máquinas de Turing. Por conseguinte, há mais funções nos inteiros positivos do que máquinas que possam descrevê-las. Em suma, portanto, é possível montar a sequência $(f_n)_n = (f_1, f_2, ..., f_n, ...)$ de todas a funções unárias Turing-computáveis, em correspondência com as máquinas que as computam. Definimos, então, a *função diagonal d*, da seguinte maneira:

$$d(n) = \begin{cases} 2, & \text{se } f_n(n) \text{ está definida e é } = 1; \\ 1, & \text{caso contrário.} \end{cases}$$

De modo que derivemos uma contradição, suponha que d está listada em $(f_n)_n$; dessa forma, existe $m \in \mathbb{N}^*$ tal que $d = f_m$, i.e., ou é o caso de ambos os valores $d(n)$ e $f_m(n)$ estarem definidos para um mesmo n inteiro positivo qualquer, ou ambos não estão definidos. Mas considere o caso em que $n = m$; teremos

$$d(m) = f_m(m) = \begin{cases} 2, & \text{se } f_m(m) \text{ está definida e é } = 1; \\ 1, & \text{caso contrário.} \end{cases}$$

Se $f_m(m)$ está definida e seu valor é 1, então a definição acima nos dá $f_m(m) = 2$, donde vem que $2 = 1$; por outro lado, se $f_m(m)$ está definida mas é $\neq 1$, ficamos com $f_m(m) = 1$, logo $1 \neq 1$; e caso $f_m(m)$ não esteja definida, concluímos que $f_m(m) = 1$, do que segue que ela está, na verdade, muito bem definida. Em qualquer um dos três casos, obtemos um absurdo. Portanto, d não é um termo da sequência $(f_n)_n$, i.e., não há máquina de Turing que compute tal função.[4]]

À conjectura de que Turing-computabilidade é, em essência, a única forma de se decidir os valores de uma função em $\mathbb{N}^*$, dá-se o nome de *tese de Turing*. Pelos motivos elencados antes do parêntese, tal tese é, cada vez mais, dentro do campo da teoria da recursão, vista como perfeitamente aceitável. No entanto, deixemos nossa conclusão sobre este assunto para depois. Nas próximas páginas, buscaremos uma abordagem mais formal para processos efetivos de decisão — i.e., uma maneira mais ortodoxamente matemática e menos apelativa aos nossos sentidos —; mas lembre-se: nossa definição de máquina de Turing seria suficiente para o conteúdo do presente trabalho, e ainda faremos alusões aos conceitos introduzidos até aqui.

2.2 FUNÇÕES RECURSIVAS

Antes de sairmos listando definições matemáticas precisas e provando teoremas sobre funções e relações efetivamente computáveis, façamos algumas preliminares sobre as bases intuitivas da teoria que pretendemos construir. Como discutido na seção anterior, parece haver funções mais simples de se decidir do que outras — em particular, é mais fácil nos convencermos sobre a computabilidade da adição do que de $f(x) = 2x$. Mas há funções que aparentemente são mais simples do que a adição, embora suas máquinas de Turing, bem como seu tempo de computação, possam ser mais extensos do que os da referida função.

Por exemplo, para computarmos, numa fita unidimensional, a *função zero*, ou seja, $\mathrm{z} : \mathbb{N}^k \longrightarrow \mathbb{N}$, com $\mathrm{z}(x_1, ..., x_k) = 0$ para todo $(x_1, ..., x_k) \in \mathbb{N}^k$, precisaríamos percorrer todos os blocos de traços, eliminando cada um deles, e identificar o momento de parar.[5] No entanto, não há dúvidas de que, quanto à decidibilidade, a função zero é muito mais elementar: não importa o argumento, o valor de saída será sempre nulo. Algo semelhante acontece com as funções identidade, também chamadas de *projeções*;[6] para o caso unário, uma máquina que computa $\mathrm{id}^1_1(x) = x$ simplesmente

[4] Perceba que o argumento usado é análogo ao de Cantor na prova de existência de conjuntos não-enumeráveis. De maneira geral, o método de diagonalização representa um papel bastante importante nas demonstrações metamatemáticas, sobretudo em questões de indecidibilidade.

[5] Note que tal exemplo é uma função nos naturais, um domínio diferente do que trabalhamos na seção anterior, que era os inteiros positivos. De qualquer forma, podemos modificar a definição de máquina de Turing de modo que trabalhemos em $\mathbb{N}$, convencionando que o número $m \in \mathbb{N}$ seja representado por $m + 1$ traços.

[6] A notação para tais funções é da forma id^m_n, em que o índice sobrescrito, m, representa o tamanho do argumento (i.e., do domínio), e o subscrito nos diz qual entrada projetar — neste caso, a n-ésima.

deixaria intacto o bloco de $x + 1$ traços, mas para $\mathrm{id}_i^k(x_1, ..., x_i, ..., x_k) = x_i$, seria necessário apagar todos os outros blocos de traços que não fossem o i-ésimo. De qualquer forma, decidir o valor de uma projeção i-ésima de qualquer k-upla é um processo bastante intuitivo — podemos imaginar tal processo como o ato de *escolher* a i-ésima entrada do argumento dado. Assim, ambas as funções mencionadas são, de certa maneira, computáveis em apenas um passo (ao menos em *algum* sentido intuitivo de passo — e.g., considerando abstrações mais eficientes do que máquinas de Turing unidimensionais). Como último exemplo de função decidível via um único ato, temos a *função sucessor*, i.e., $\mathrm{s}(x) = x + 1$ para todo $x \in \mathbb{N}$, unária por natureza (com efeito: uma máquina de Turing para s apenas adicionaria um traço no final do bloco de traços representando o argumento).

Nosso objetivo, ao estabelecermos o fato de que existem funções mais intuitivamente computáveis do que outras, é que possamos definir uma classe de funções computáveis *a partir* dessas funções mais simples.[7] A essa coleção daremos o nome de *classe das funções recursivas*, e a representaremos por $\mathcal{R}$. Para isso precisaremos, também, escolher procedimentos matemáticos que sejam capazes de "fundir" essas funções básicas, i.e., combiná-las de modo que se obtenha novas funções computáveis. Provavelmente o processo mais usual de se definir funções a partir de outras seja a *composição*. E parece intuitivo o suficiente pensar que, dadas $f : \mathbb{N}^k \longrightarrow \mathbb{N}$ e $g_1, ..., g_k : \mathbb{N}^j \longrightarrow \mathbb{N}$, todas computáveis, o valor $f(g_1(x_1, ..., x_j), ..., g_k(x_1, ..., x_j))$ seja, também, passível de decisão, para qualquer $(x_1, ..., x_j) \in \mathbb{N}^j$. De fato: como $g_1, ..., g_k$ são computáveis, podemos determinar, para cada $i \in \{1, ..., k\}$, o valor $g_i(x_1, ..., x_j) = y_i$ em um número finito de passos, e depois calcular $f(g_1(x_1, ..., x_j), ..., g_k(x_1, ..., x_j)) = f(y_1, ..., y_k)$, também num número finito de etapas. O número total de passos nessa computação será a soma das quantidades de passos para computar cada uma das $f, g_1, ..., g_k$. Ou seja, a efetividade desse processo é tão intuitivamente clara quanto o ato de realizar uma ação seguida de outra, numa ordem específica.

Outro processo bastante presente nas mais diversas áreas da Matemática é o de *definição por indução*, ou por *recursão primitiva*. Como o nome sugere, podemos definir uma função ao mostrar como fazê-lo para um caso inicial, seguindo uma regra para se obter, sucessivamente, os demais valores. Por exemplo, a seguinte lista define a operação fatorial:

$$f(0) = 1;$$
$$f(x + 1) = (x + 1) * f(x).$$

O número de passos para computar $n!$ será a soma dos números de passos para calcular 0!, 1!, e assim por diante, até o último, que é multiplicar n por $(n - 1)!$. Dessa

[7] Assim como em qualquer teoria matemática, em que devemos determinar os conceitos básicos e indefiníveis, a partir dos quais definem-se as demais entidades, de modo que não entremos num círculo vicioso.

forma, fica intuitivamente claro como a recursão primitiva preserva computabilidade.

Vejamos, no entanto, como formalizar tal processo. Na tentativa de generalizar o que foi exemplificado, acima, pela operação fatorial, queremos definir uma função através de outras duas funções (computáveis) pré-estabelecidas, de modo que todos os seus valores sejam computados a partir dos valores anteriores (o que caracteriza o processo como indutivo), e usando as funções que temos de antemão. Assim, dadas h e g funções computáveis de, respectivamente, $k+1$ e $k-1$ variáveis, definimos a *função obtida de h e g por recursão primitiva*, $f : \mathbb{N}^k \longrightarrow \mathbb{N}$, como sendo uma função de k argumentos que satisfaz as seguintes regras:

$$f(0, x_2, ..., x_k) = g(x_2, ..., x_k);$$
$$f(y+1, x_2, ..., x_k) = h(y, f(y, x_2, ...x_k), x_2, ..., x_k).$$

Caso seja $k = 1$, convencionaremos que g ser uma função de zero variáveis significa que g é igual a um número natural específico. Vemos esse fenômeno no exemplo acima, em que g é igual a 1. Já a outra função é $h(x_1, x_2) = \mathrm{s}(\mathrm{id}^2_1(x_1, x_2)) * \mathrm{id}^2_2(x_1, x_2)$, pois

$$\begin{aligned} f(x+1) &= h(x, f(x)) \\ &= \mathrm{s}(\mathrm{id}^2_1(x, f(x)) * \mathrm{id}^2_2(x, f(x)) \\ &= \mathrm{s}(x) * f(x). \end{aligned}$$

Como exemplo de uma função obtida através das chamadas funções *básicas* (aquelas mais simples de se computar), temos a própria adição. A maneira usual de se definir adição é via função sucessor, colocando-se $x + 0 = x$ e $x + \mathrm{s}(y) = \mathrm{s}(x + y)$. Para enfatizar as características binária e funcional da adição, e colocar as condições no formato oficial de uma definição via recursão primitiva, podemos escrever:

$$+(0, x) = x;$$
$$+(\mathrm{s}(y), x) = \mathrm{s}(+(y, x)).$$

De qualquer forma, fica evidente que g é a função unária id^1_1; por outro lado, h é a função ternária $h(x_1, x_2, x_3) = \mathrm{s}(\mathrm{id}^3_2(x_1, x_2, x_3))$, pois devemos ter

$$\begin{aligned} +(\mathrm{s}(y), x) &= h(y, +(y, x), x) \\ &= \mathrm{s}(\mathrm{id}^3_2(y, +(y, x), x)) \\ &= \mathrm{s}(+(y, x)). \end{aligned}$$

Usando essa definição e a função sucessor, podemos também definir, via recursão primitiva, a multiplicação — mas omitiremos tal prova; o leitor interessado pode verificar que $*$ é obtida quando $g = z$ (a função zero) e h é a composição da adição com a composta $\mathrm{id}^3_3 \circ \mathrm{id}^3_3$ (i.e., $+(\mathrm{id}^3_3, \mathrm{id}^3_3)$).

O último método para combinar funções que precisamos considerar tem a ver com a existência de um elemento do domínio que satisfaz uma determinada condição. Mais especificamente, o processo de *minimização* gera uma nova função da seguinte maneira: suponha que g seja uma função de $k+1$ argumentos, e que φ seja qualquer fórmula aberta na variável x. Então, por $\mu x\varphi(x)$ entenderemos *o menor x tal que* $\varphi(x)$, e daremos o nome de μx à *operação* de minimização (não confundir com o *processo*, que ainda estamos definindo). [Por exemplo, para a fórmula aberta $H(x)$: $x^2 - 3x + 2 = 0$, temos $\mu x H(x) = 1$.] Agora, assuma que g é computável, e tome $(y_1, ..., y_k) \in \mathbb{N}^k$. Suponha, também, que exista algum $x \in \mathbb{N}$ tal que

$$g(y_1, ..., y_k, x) = 0. \tag{1}$$

Assim, podemos definir uma nova função, f, de apenas k variáveis, a partir de g e via minimização:

$$f(y_1, ..., y_k) = \mu x[g(y_1, ..., y_k, x) = 0].$$

Perceba que o fato de g ser computável, junto da condição (1), garantem que f pode ser computada para qualquer valor em $\mathbb{N}^k$. Com efeito: dado $\boldsymbol{x} = (x_1, ..., x_k)$, para calcular o valor de $f(\boldsymbol{x})$ é necessário, apenas, que computemos sucessivamente os valores $g(\boldsymbol{x}, 0)$, $g(\boldsymbol{x}, 1)$, ..., até que cheguemos a $g(\boldsymbol{x}, x_0) = 0$. Sabemos que haverá um tal x_0, pois (1) nos assegura isso. Por meio da minimização, encontramos o valor de $f(\boldsymbol{x})$.

Para finalizar essa discussão introdutória sobre a classe das funções recursivas, será interessante estabelecermos algumas notações. Primeiro, combinemos que os símbolos s, z e id_i^k devem ser exclusivos das funções sucessor, zero e projeção, respectivamente. (Lembre-se: essas são as funções *básicas*, que darão fruto a todas as outras funções recursivamente computáveis.) Além delas, chamaremos a adição por $+$ e a multiplicação por $*$, que são suas notações mais usuais. Quanto aos métodos de geração de novas funções, é prática comum na literatura abreviá-los pelos símbolos Cn, Rp e Mn (composição, recursão primitiva e minimização, respectivamente), seguidos de colchetes, e das funções envolvidas no processo. Por exemplo, a adição poderia ser escrita como $+ = \mathrm{Rp}[\mathrm{id}_1^1, \mathrm{Cn}[\mathrm{s}, \mathrm{id}_2^3]]$, o que é uma forma claramente mais compacta do que especificar g e h, como fizemos anteriormente. A forma geral da minimização, como no exemplo que demos, seria $f = \mathrm{Mn}[g](x_1, ..., x_k)$, omitindo a operação μx. Isto posto, finalmente podemos dizer que as definições indicadas desde o início da seção motivam a seguinte caracterização:

Definição 2.2.1. A *classe $\mathcal{R}$ das funções recursivas* é a menor classe $\mathcal{C}$ (i.e., a intersecção de todas as classes $\mathcal{C}$) de funções que satisfaz as seguintes condições:

1. Todas as funções zero estão em $\mathcal{C}$, i.e., dado qualquer $k \in \mathbb{N}^*$, tem-se que z : $\mathbb{N}^k \longrightarrow \mathbb{N}$ está em $\mathcal{C}$;

2. Todas as funções projeção estão em $\mathcal{C}$, i.e., dado qualquer $k \in \mathbb{N}^*$, tem-se que para qualquer $i \in \{1, ..., k\}$, $\mathrm{id}_i^k : \mathbb{N}^k \longrightarrow \mathbb{N}$ está em $\mathcal{C}$;

3. A função sucessor está em $\mathcal{C}$, i.e., s : $\mathbb{N} \longrightarrow \mathbb{N}$ está em $\mathcal{C}$;

4. Se g é uma função de m argumentos e está em $\mathcal{C}$, e se $h_1, ..., h_m$ são, todas elas, funções de n argumentos, estando todas em $\mathcal{C}$, então a função $f = \mathrm{Cn}[g, h_1, ..., h_m]$ também está em $\mathcal{C}$;

5. Se h e g são, respectivamente, funções de $k + 1$ e $k - 1$ argumentos e estão ambas em $\mathcal{C}$, então a função $f = \mathrm{Rp}[g, h]$ também está em $\mathcal{C}$;

6. Se uma função g de $k + 1$ argumentos está em $\mathcal{C}$, e é verdade que

$$\forall \boldsymbol{x} \in \mathbb{N}^k \exists y \in \mathbb{N}[g(\boldsymbol{x}, y) = 0],$$

   então tem-se que $f = \mathrm{Mn}[g](\boldsymbol{x})$ também está em $\mathcal{C}$.

Não existe outra função em $\mathcal{R}$ além das presentes nos itens *1-3*, ou geradas a partir delas através dos métodos mencionados em *4-6*.

Em suma, portanto, $\mathcal{R}$ é a menor classe de funções contendo zero, sucessor e as identidades, e fechada pelas técnicas de composição, recursão primitiva e minimização. Dessa maneira, diremos que toda $f \in \mathcal{R}$ é uma *função recursiva*, ou, ainda, *recursivamente computável*. Isso encerra, por ora, a formalização de tais funções, que será nossa maneira de abstrair computabilidade daqui em diante. Mas quanto às *relações* entre números naturais? Como fazer para decidir, por exemplo, se um inteiro não nulo é menor do que outro? Ora, o leitor deve perceber que, como toda função é uma relação (de um tipo bastante especial), há de ser possível definir computabilidade de relações por meio de computabilidade de funções. De fato: podemos fazê-lo usando a *função característica* de uma relação; se esta for computável, então a relação também será. Dada uma relação n-ária qualquer, sua função característica atribui o valor 1 caso se verifique a relação entre os membros da n-upla $(x_1, ..., x_n)$, e atribui 0 se tais membros não se relacionam.

Voltando nossa atenção ao exemplo da relação "menor que", à qual comumente se designa o símbolo <, vejamos como computá-la. Primeiramente, precisamos definir três novas funções computáveis, via recursão primitiva. Chamaremos de *antecessor* a função dada por

$$\mathrm{a}(0) = 0;$$
$$\mathrm{a}(\mathrm{s}(x)) = x.$$

Usando essa função, podemos definir a *subtração modificada*, designada por $\ominus$, a qual opera $x \ominus y = x - y$ se for verdade que $x - y \in \mathbb{N}$, e $x \ominus y = 0$, caso contrário:

$$x \ominus 0 = x;$$
$$x \ominus \mathrm{s}(y) = \mathrm{a}(x \ominus y).$$

Por fim, temos a função *sinal*, cujo efeito é apenas nos dizer se determinado número é nulo ou não, i.e., chamando-a por sg, devemos ter $\mathrm{sg}(0) = 0$ e $\mathrm{sg}(x) = 1$, caso seja $x \in \mathbb{N}^*$. Assim, por recursão primitiva, temos:

$$\mathrm{sg}(x) = 1 \ominus (1 \ominus x).$$

É importante notar que, nessas definições, fizemos uso de fatos que não demonstramos, como a computabilidade da subtração. No entanto, imaginamos ser bastante crível, ao menos nesses casos, que tais funções são recursivas (podendo ser definidas por meio da adição e dos métodos que elencamos).[8] De toda forma, agora nos é possível determinar a função característica da relação $<$. Com efeito: chamando-a por $K_<$, e fazendo $K_<(x, y) = 1$ se e somente se $x < y$, tal função deve ter a seguinte forma:

$$K_<(x, y) = \mathrm{sg}(y \ominus x).$$

A relação de *igualdade* também pode ser definida usando as funções sinal e subtração modificada (além da adição e da subtração): estabelecendo-se que $K_=(x, y) = 1$ se e somente se $x = y$, temos:

$$K_=(x, y) = 1 - (\mathrm{sg}(x \ominus y) + \mathrm{sg}(y \ominus x)).$$

Mas chega de exemplos, por enquanto. Deixemos apenas registrado que

Definição 2.2.2. Uma *relação R* nos números naturais é dita *recursiva* se e somente se sua função característica, K_R, for recursiva.

Estando, agora, em posse de definições matematicamente precisas para o conceito abstrato de "processo de decisão" de problemas matemáticos (diferentemente do que fizemos na seção anterior), enunciemos e provemos algumas proposições gerais de teoria da recursão. Elas serão muito úteis para o desenvolvimento futuro do trabalho, principalmente na concepção de novas funções e relações computáveis. A primeira delas nos garante que ao relacionarmos funções recursivas por meio de uma relação recursiva, teremos uma nova relação recursiva:

[8] Além do mais, e aqui vale o alerta de *spoiler* (que já foi dado nas primeiras linhas deste capítulo), é possível mostrar que toda função recursiva é computável por alguma máquina de Turing, e vice-versa. Como mencionado na seção anterior, toda função aritmética é computável por ábacos, logo também o é por alguma máquina de Turing, donde concluímos que é recursiva.

Teorema 2.2.1. *Sejam P uma relação n-ária e $f_1, ..., f_n$ funções quaisquer. Suponha que todas são recursivas, sendo que cada uma das funções possui k argumentos. Então, dado $\boldsymbol{x} = (x_1, ..., x_k)$ arbitrário, é recursiva a relação*

$$R(\boldsymbol{x}) = P(f_1(\boldsymbol{x}), ..., f_n(\boldsymbol{x})).$$

Demonstração. Temos que $K_R(\boldsymbol{x}) = K_P(f_1(\boldsymbol{x}), ..., f_n(\boldsymbol{x}))$, i.e., que $K_R = \mathrm{Cn}[K_P, f_1, ..., f_n]$. Da definição 2.2.1, vem que K_R é recursiva. □

Teorema 2.2.2. *Toda função constante é recursiva, i.e., para quaisquer $m \in \mathbb{N}$, $k \in \mathbb{N}^*$ e função $f : \mathbb{N}^k \longrightarrow \mathbb{N}$, caso seja $f(x_1, ..., x_k) = m$ para todo $(x_1, ..., x_k) \in \mathbb{N}^k$, então f é recursiva.*

Demonstração. De fato, resolvemos isso via indução em m: denotando por $f^{(=m)}$ a função k-ária com valor constante igual m, temos, para $m = 0$, e tomando $\boldsymbol{x} = (x_1, ..., x_k)$, que

$$f^{(=0)}(\boldsymbol{x}) = \mu y[\mathrm{id}^{k+1}_{k+1}(\boldsymbol{x}, y) = 0].$$

Supondo que para um $m \in \mathbb{N}$ arbitrário seja verdade que $f^{(=m)}$ é recursiva, provamos agora que $f^{(=m+1)}$ também o é. Com efeito: podemos definir $f^{(=m+1)}$ como sendo

$$f^{=(m+1)}(\boldsymbol{x}) = \mu y[f^{(=m)}(\boldsymbol{x}) < y],$$

e como $f^{(=m)}$ e a relação $<$ são recursivas, segue-se do teorema anterior que $f^{(=m+1)}$ é recursiva. □

O próximo resultado determina que os conectivos lógicos, quando interpretados como operações relacionais[9], preservam recursividade nas novas relações (e funções) geradas por eles:

Teorema 2.2.3. *Sejam P e Q relações k-árias quaisquer, e suponha que elas são recursivas. Então, as relações $\neg P$, $P \vee Q$, $P \wedge Q$, $P \to Q$ e $P \leftrightarrow Q$ também o são.*

Demonstração. Dado $\boldsymbol{x} = (x_1, ..., x_k)$ arbitrário, temos que as funções características de $\neg P$ e $P \vee Q$ são dadas por

$$K_{(\neg P)}(\boldsymbol{x}) = K_{=}(0, K_P(\boldsymbol{x}));$$
$$K_{(P \vee Q)}(\boldsymbol{x}) = \mathrm{sg}(K_P(\boldsymbol{x}) + K_Q(\boldsymbol{x})).$$

Para os demais casos, usamos o fato de que os conectivos lógicos respectivos podem ser definidos tão-somente por disjunção e negação, i.e., que

$$P \wedge Q \coloneqq \neg(\neg P \vee \neg Q);$$
$$P \to Q \coloneqq \neg P \vee Q;$$
$$P \leftrightarrow Q \coloneqq (P \to Q) \wedge (Q \to P).$$

[9] Exemplificando o que isso quer dizer: dadas quaisquer relações P e Q, ambas k-árias, e $\boldsymbol{x} = (x_1, ..., x_k)$, podemos definir a equivalência $(P \wedge Q)(\boldsymbol{x}) \leftrightarrow P(\boldsymbol{x}) \wedge Q(\boldsymbol{x})$.

□

Dessa forma, concluímos que podemos estender a noção *lógica* de operador para o campo da *aritmética*. Mais precisamente, provamos que a classe das relações (e funções) recursivas é *fechada* por tais conectivos, quando estes são interpretados como *operações sobre relações entre números naturais*. A pergunta que prontamente surge é se podemos, também, estender essa ideia para os quantificadores, uma vez que precisamos de lógica quantificacional para construir, praticamente, qualquer teoria matemática relevante. Infelizmente, a resposta para tal pergunta é negativa. A explicação para esse fato[10] foge um pouco de nosso escopo, mas o que importa é que podemos contornar tal empecilho, sem comprometer nossos objetivos principais. Introduzindo o conceito de *quantificadores limitados*, provaremos um resultado semelhante ao teorema 2.2.3, donde concluiremos que, de certa maneira, a classe das relações recursivas é fechada por quantificação. A notação para tais quantificadores será a mais intuitiva possível: fixando $v \in \mathbb{N}$ e φ uma fórmula qualquer (livre em v e, possivelmente, em outras variáveis), colocamos $\forall v < u\varphi(v)$ e $\exists v < u\varphi(v)$. Explicitamente, a primeira significa que $\exists v(v < u \wedge \varphi(v))$, enquanto a segunda diz que $\forall v(v < u \rightarrow \varphi(v))$. [Para provar o resultado citado, será útil considerar quantificações com respeito à ordem não-estrita, i.e., poderemos, também, escrever $\forall v \leq u\varphi(v)$ e $\exists v \leq u\varphi(v)$.][11]

Definição 2.2.3. Dada R uma relação $(k+1)$-ária qualquer, pela relação obtida de R via *quantificação universal limitada*, entende-se a relação S que se dá entre $x_1, ..., x_k, u \in \mathbb{N}$ se e somente se para todo $v < u$ é válido que $R(x_1, ..., x_k, v)$; ou seja, devemos ter que

$$S(x_1, ..., x_k, u) \leftrightarrow \forall v < uR(x_1, ..., x_k, v).$$

Analogamente, define-se a relação obtida de R via *quantificação existencial limitada*:

$$T(x_1, ..., x_k, u) \leftrightarrow \exists v < uR(x_1, ..., x_k, v).$$

[10] O problema, *grosso modo*, é que cada vez que adicionamos um quantificador a uma fórmula, aumentamos seu grau de complexidade. Admitindo modelos infinitos (como é o caso da aritmética dos naturais), e como não há restrição quanto ao número de quantificadores que podemos inserir, pode-se entender, ao menos de maneira intuitiva, de que modo as coisas podem fugir do controle (ou seja, é relativamente fácil ver como quantificadores ilimitados *não devem* preservar computabilidade, já que, nesse caso, precisaríamos de uma quantidade infindável de etapas para verificar a validade de uma sentença). Para estudar essas nuances, define-se uma *hierarquia lógica* para as fórmulas de uma teoria, cujos níveis de complexidade são medidos pelo número de quantificadores que elas possuem (e disso segue que tal forma de categorizar classes de relações possui infinitamente muitos patamares). Os andares mais baixos dessa hierarquia são as hierarquias aritmética e analítica, e o nível mais baixo da hierarquia aritmética é a classe das funções recursivas. No sentido dado por tal abstração, pode-se provar que toda classe *acima* de $\mathcal{R}$ é fechada por quantificação — mas uma prova abrangendo $\mathcal{R}$ é impossível.

[11] Perceba que podemos definir a relação de ordem $\leq$ como sendo $x \leq y \leftrightarrow \neg(y < x)$. Logo, sua recursividade está garantida.

O seguinte lema será necessário para a demonstração do resultado que estamos aludindo:

Lema 2.2.1. *Seja f uma função recursiva primitiva qualquer (i.e., gerada por recursão primitiva), e suponha que ela tem $k+1$ argumentos. Tomando $\boldsymbol{x} = (x_1, ..., x_k)$, temos que as seguintes funções são recursivas primitivas:*

$$g(\boldsymbol{x}, y) = f(\boldsymbol{x}, 0) + f(\boldsymbol{x}, 1) + ... + f(\boldsymbol{x}, y) = \sum_{i=0}^{y} f(\boldsymbol{x}, i);$$

$$h(\boldsymbol{x}, y) = f(\boldsymbol{x}, 0) * f(\boldsymbol{x}, 1) * ... * f(\boldsymbol{x}, y) = \prod_{i=0}^{y} f(\boldsymbol{x}, i).$$

Demonstração. De fato, podemos definir g pelas equações

$$g(\boldsymbol{x}, 0) = f(\boldsymbol{x}, 0);$$
$$g(\boldsymbol{x}, \mathrm{s}(y)) = g(\boldsymbol{x}, y) + f(\boldsymbol{x}, \mathrm{s}(y)),$$

e h de maneira análoga (usando produto no lugar de soma). □

Teorema 2.2.4. *Seja R uma relação $k+1$-ária qualquer, supondo-a recursiva. Então, as relações S e T, obtidas a partir de R por quantificação limitada (universal e existencial, respectivamente), também são recursivas.*

Demonstração. Tome $\boldsymbol{x} = (x_1, ..., x_k) \in \mathbb{N}^k$ arbitrário, e considere a função característica $K_R(\boldsymbol{x}, y)$ da relação R, recursiva por hipótese. Então, as funções características das relações $S(\boldsymbol{x}, y) \leftrightarrow \forall v \leq y R(\boldsymbol{x}, v)$ e $T(\boldsymbol{x}, y) \leftrightarrow \exists v \leq y R(\boldsymbol{x}, v)$ podem ser definidas, respectivamente, por

$$K_S(\boldsymbol{x}, y) = \prod_{i=0}^{y} K_R(\boldsymbol{x}, i);$$

$$K_T(\boldsymbol{x}, y) = \mathrm{sg}\left(\sum_{i=0}^{y} K_R(\boldsymbol{x}, i)\right),$$

pois, para K_S, o produtório será igual a 1 se e somente se *todas* as parcelas também o forem, i.e., se $R(\boldsymbol{x}, i)$ for válida para todo $i \in \{0, ..., y\}$; e se dentro do parênteses da fórmula que descreve K_T houver *alguma* parcela igual a 1, então $K_T(\boldsymbol{x}, y) = 1$. Para o caso de quantificadores estritos, apenas precisamos substituir y por $y \ominus 1$. □

A proposição acima, juntamente do teorema 2.2.3, formam as *propriedades de fecho* da classe das funções recursivas. Com elas, garantimos que, até certo ponto, podemos definir novas funções e relações recursivamente computáveis do mesmo modo que os conectivos e quantificadores lógicos são capazes de construir infinitamente muitas novas sentenças de primeira ordem. Indiquemos, agora, um último método de

definição de funções que preserva recursividade; tal técnica é comumente chamada de *definição por casos*.

Teorema 2.2.5. *Sejam $g_1, ..., g_n$ funções recursivas e $R_1, ..., R_n$ relações recursivas quaisquer, todas de m argumentos. Suponha que, dado $\boldsymbol{x} = (x_1, ..., x_m)$ arbitrário, exatamente uma das $R_1(\boldsymbol{x}), ..., R_n(\boldsymbol{x})$ se verifique. Então, é recursiva a função $f : \mathbb{N}^m \longrightarrow \mathbb{N}$, dada por*

$$f(\boldsymbol{x}) = \begin{cases} g_1(\boldsymbol{x}), & \text{se } R_1(\boldsymbol{x}); \\ \vdots & \vdots \\ g_n(\boldsymbol{x}), & \text{se } R_n(\boldsymbol{x}). \end{cases}$$

Demonstração. Chamando por K_i a função característica de R_i, definimos, via recursão primitiva, a função

$$h_i(0, \boldsymbol{x}) = 0;\ h_i(\mathrm{s}(z), \boldsymbol{x}) = g_i(\boldsymbol{x}).$$

Colocando $c_i(\boldsymbol{x}) = h_i(K_i(\boldsymbol{x}), \boldsymbol{x})$, teremos sempre que $c_i(\boldsymbol{x}) = 0$, a menos que $R_i(\boldsymbol{x})$ valha — caso em que $c_i(\boldsymbol{x}) = g_i(\boldsymbol{x})$, pois daí $K_i(\boldsymbol{x}) = 1$. Assim, temos que

$$f(\boldsymbol{x}) = c_1(\boldsymbol{x}) + c_2(\boldsymbol{x}) + ... + c_n(\boldsymbol{x}),$$

o que prova que f é recursiva. □

Com isso, concluímos a tarefa de especificar técnicas (secundárias) para a geração de funções recursivas, que tornam esse tipo de trabalho mais fácil do que quando havíamos somente os métodos mencionados na definição 2.2.1. Listemos, agora, uma série de funções e relações recursivas que podem ser obtidas através delas; algumas dessas relações terão papel fundamental na definição de *prova* que pretendemos estabelecer.

1. As funções *mínimo* e *máximo*, que nos dão, respectivamente, o menor e o maior valor dentre dois números, são dadas por

   $$\min(x, y) = x \ominus (x \ominus y);$$
   $$\max(x, y) = y + (x \ominus y).$$

   Podemos estender essas funções para uma quantidade (finita) qualquer de variáveis. Para a função mínimo, por exemplo, perceba que

   $$\min(x_1, ..., x_n) = \min(\min(x_1, ..., x_{n-1}), x_n).$$

   Assim, para provar que, independentemente da quantidade de variáveis, a função mínimo é recursiva, basta proceder via indução finita em n, usando a equação acima.

2. A relação de *divisibilidade*, chamada $\text{Div}(x, y)$, que ocorre se e somente se y divide x, é dada por

$$\text{Div}(x, y) \leftrightarrow \exists z \leq x(x = y * z).$$

3. A relação que se dá entre x, y, z quando $x = y * z$ pode ser vista como o *gráfico* da função produto (que já sabemos ser recursiva);[12]

4. A relação de *primalidade*, que nos diz se determinado número é primo ou não, pode ser definida por

$$\text{Pr}(x) \leftrightarrow x \neq 0 \wedge x \neq 1 \wedge \forall u < x \forall v < x(u * v \neq x),$$

sendo que $x \neq y \leftrightarrow \neg(x = y)$. Usando essa relação, podemos definir a função *próximo primo*, que para cada $x \in \mathbb{N}$, nos dá o menor $y > x$ tal que y é primo. Isso se dá pela minimização da relação $R(x, y) \leftrightarrow (x < y \wedge \text{Pr}(y))$;

5. Fixada R uma relação qualquer, e assumindo que $C = \{y_1, ..., y_k = w\}$ é o conjunto de todos os números $y_i \leq w$ tais que $R(\boldsymbol{x}, y_i)$, temos as funções de *minimização* e *maximização limitadas*, dadas por

$$\text{Min}[R](\boldsymbol{x}, w) = \begin{cases} \min(y_1, ..., y_k), & \text{caso } C \neq \emptyset; \\ w + 1, & \text{caso contrário;} \end{cases}$$

$$\text{Max}[R](\boldsymbol{x}, w) = \begin{cases} \max(y_1, ..., y_k), & \text{caso } C \neq \emptyset; \\ 0, & \text{caso contrário;} \end{cases}$$

6. As funções exponenciais, chamadas $f(x) = b^x$ e definidas por recursão primitiva a partir de 1 e da multiplicação, para cada $b \in \mathbb{N}$.

7. Se colocarmos $R(x, y, z) \leftrightarrow y * z \leq x$, então as funções *quociente* e *resto da divisão euclidiana de x por y* podem ser definidas, respectivamente, por $\text{q}(x, y) = \text{Max}[R](x, y, z)$ e $\text{r}(x, y) = x \ominus (\text{q}(x, y) * y)$;

8. Pelo *logaritmo de x na base y*, entendemos o maior $z \in \mathbb{N}$, $z \leq x$, tal que y^z divide x. Assumindo que $x, y > 1$ e que $R(x, y, z) \leftrightarrow \text{Div}(x, y^z)$, i.e., que R se dá entre x, y, z quando y^z divide x sem deixar resto, temos que

$$\log(x, y) = \text{Max}[R](x, y, z).$$

Caso x ou y sejam ≤ 1, então $\log(x, y) = 0$.

[12] Mais geralmente, pode-se provar que se f é uma função recursiva, e considerando que o gráfico de f é a relação definida por $\text{Gr}(\boldsymbol{x}, y) \leftrightarrow f(\boldsymbol{x}) = y$, então $\text{Gr}(\boldsymbol{x}, y)$ é recursiva. Tal resultado provém da ideia introduzida no teorema 2.2.1.

As próximas três funções recursivas envolvem o conceito de *sequência finita de números naturais*, que designamos, de forma geral, por $s_n = (a_0, a_1, ..., a_{n-1})$. Mais especificamente, envolvem a *codificação* destas por um único $y \in \mathbb{N}$. Em virtude do Teorema Fundamental da Aritmética, que enuncia a decomposição (única) de qualquer número natural em fatores primos, podemos estabelecer uma correspondência entre sequências em $\mathbb{N}$ e o próprio $\mathbb{N}$.[13] Combinemos que s_n será codificada por

$$\langle s_n \rangle = 2^n * 3^{a_0} * ... * \pi(n)^{a_{n-1}}, \tag{2}$$

em que $\pi(n)$ é o n-ésimo número primo (considerando 2 como o 0-ésimo) — função que definimos agora, dando continuidade à lista:

8. Considere f como sendo a função *próximo primo*, mencionada no quarto item. Então, temos que $\pi(0) = 2$ e $\pi(s(x)) = f(x)$;

9. A função *comprimento* de s_n, que, quando aplicada ao código em (2), deve nos dar n, pode ser definida por

$$\text{lh}(\langle s_n \rangle) = \log(\langle s_n \rangle, 2).$$

   De fato: assim sendo, $\text{lh}(\langle s_n \rangle)$ será o maior $z \in \mathbb{N}$ tal que 2^z divide, sem deixar resto, o número $2^n * 3^{a_0} * ... * \pi(n)^{a_{n-1}}$. De maneira similar, definimos a função *i-ésima entrada* de s_n, pondo $\text{entr}(\langle s_n \rangle, i) = \log(\langle s_n \rangle, \pi(i+1))$;

10. Dadas duas sequências $s_n = (a_0, a_1, ..., a_{n-1})$ e $s_m = (b_0, b_1, ..., b_{m-1})$, e escrevendo $a = \langle s_n \rangle$ e $b = \langle s_m \rangle$, podemos definir a *concatenação de s_n e s_m*, se decodificarmos o que segue:

$$\begin{aligned}\langle s_n \star s_m \rangle = \mu x[&\text{lh}(x) = \text{lh}(a) + \text{lh}(b) \wedge \\ &\wedge \forall i < \text{lh}(a)(\text{entr}(x, i) = \text{entr}(a, i)) \wedge \\ &\wedge \forall i < \text{lh}(b)(\text{entr}(x, \text{lh}(a) + i) = \text{entr}(b, i))].\end{aligned}$$

   Como resultado, teremos a sequência $s_n \star s_m = (a_0, ..., a_{n-1}, b_0, ..., b_{m-1})$;

11. Por fim, temos a função *truncagem* de s_n, que, dado $i < n$ arbitrário, nos dá o código da sequência $(a_0, a_1, ..., a_{i-1})$:

$$\text{tr}(a, i) = \mu x[\text{lh}(x) = i \wedge \forall j < i(\text{entr}(x, j) = \text{entr}(a, j))].$$

[13] A bem da verdade, existem muitas formas de se codificar sequências de números. No próximo capítulo abordaremos outros métodos possíveis, e um deles se mostrará como uma generalização para esse conceito.

Finalizada a listagem, podemos dizer que a maior parte do trabalho do presente capítulo está concluída. Ainda cabem, no entanto, alguns comentários interessantes. Primeiro, quanto à etimologia, é importante notar que o motivo para os lógicos-matemáticos terem escolhido dar o nome de "funções *recursivas*" aos elementos de $\mathcal{R}$ é que, inicialmente, pensava-se que toda função computável — no sentido dado nesta seção — poderia ser descrita via um processo indutivo, i.e., um processo baseado no ato de *recorrer* a valores iniciais e avaliar o que acontece com seus sucessores. Ou seja: nos primórdios da teoria da recursão, imaginava-se que o método de recursão primitiva era o único capaz de abstrair processos efetivos de decisão. O contra-exemplo mais conhecido para tal afirmação é a *função de Ackermann*, devido ao pupilo de Hilbert mencionado anteriormente. Sua caracterização, melhorada por R. M. Robinson (1911-1995) e Rózsa Péter (1905-1977), utiliza o que chamamos de *dupla recursão*:

$$\begin{aligned} &f(0, y) = y + 1; \\ &f(\mathrm{s}(x), 0) = f(x, 1); \\ &f(\mathrm{s}(x), \mathrm{s}(y)) = f(x, f(\mathrm{s}(x), y)). \end{aligned}$$

Pode-se provar que f é recursiva, e que para qualquer função g que seja recursiva primitiva, existem $m, n \in \mathbb{N}$ tais que $f(m, y) > g(y)$, sempre que $y \geq n$. [Efetivamente, embora de maneira muito grosseira, uma prova para isso começaria mostrando que é necessário aplicar recursão primitiva (1) ao menos uma vez para se obter uma função que cresce tão rapidamente quanto a adição; (2) ao menos duas vezes para se obter uma função que cresce tão rapidamente quanto a multiplicação; e assim por diante, indefinidamente. Em suma, não haveria um número finito de aplicações de recursão primitiva que fosse capaz de gerar uma função que crescesse de maneira tão veloz quanto f. Daí, seguiria que a função de Ackermann não pode ser recursiva primitiva.]

O segundo ponto de interesse diz respeito a *conjuntos* recursivos. Tal ideia é bastante simples, tanto intuitiva quanto formalmente: podemos pensar que um conjunto é recursivo, ou *recursivamente decidível*, caso exista um modo efetivo, mecânico, por meio do qual possamos decidir quando determinado elemento pertence, ou não, a ele. Pensando um pouco melhor, no entanto, vemos que esse conceito já foi generalizado quando definimos relações recursivas, pois todo conjunto pode ser visto como uma relação de um só argumento. De qualquer forma, vale o registro de um resultado exclusivo para conjuntos:

Teorema 2.2.6. *São verdadeiras as seguintes proposições:*

(i) O complemento de todo conjunto recursivo também é recursivo;

(ii) União e intersecção de quaisquer dois conjuntos recursivos são ambas recursivas;

(iii) Todo conjunto finito é recursivo.

Demonstração. Para os itens (*i*) e (*ii*), basta ver que as funções características de tais conjuntos podem ser definidas segundo as operações especificadas no teorema 2.2.3 — respectivamente, via negação, disjunção e conjunção. Para o item (*iii*), no caso em que o conjunto é $A = \emptyset$, garantimos a recursividade observando que $x \in A \leftrightarrow x < x$. Seja, então, $A = \{k_1, k_2, ..., k_n\}$. Assim,

$$x \in A \leftrightarrow x = k_1 \vee x = k_2 \vee ... \vee x = k_n.$$

Pelo mesmo teorema citado, temos que A é recursivo. □

Em conclusão, voltemos agora à questão sobre a "unicidade" de nossos objetos de estudo. Na seção precedente, fizemos uma boa discussão sobre como certas máquinas abstratas, imaginadas como compartimentos onde podemos fazer operações básicas (nos mover de um lado ao outro, inserir, reconhecer e apagar símbolos), parecem ser, em essência, a única maneira pela qual, ao menos nós, humanos, somos capazes de construir mentalmente um processo efetivo de decisão matemática que transmita a característica mecânica, determinística, de um maquinário ideal. A extensão desse raciocínio parece nos indicar que, na verdade, qualquer função *decidível* deveria ser computável por uma máquina de Turing — e, contrariamente, que toda função que desconhecesse uma máquina capaz de computá-la mecanicamente deveria, também, desconhecer qualquer outro método de decisão. Como o leitor atento já deve prever, o mesmo acontece quando formalizamos processos por funções recursivas.

Recapitulando o que fizemos nas últimas páginas, fica evidente como os dois processos descritos — um mais formalmente do que o outro — devem ser, essencialmente, o mesmo. Para começar, introduzimos as funções básicas imaginando como elas deveriam ser computadas numa máquina de Turing (e o mesmo foi feito para os métodos geradores de novas funções). Essa escolha não foi somente didática, muito menos necessidade da continuidade narrativa — pois poderíamos, perfeitamente, ter trocado a ordem de abordagem dos conceitos (primeiro recursão, depois máquinas) —, mas também porque, de fato, ambas são a mesma coisa. Ou mais precisamente: funções recursivas e máquinas de Turing computam *exatamente* as mesmas relações em $\mathbb{N}$.

Com a bagagem de resultados que temos em mão, poderíamos, muito bem, provar a afirmação acima; mas, além de extensa, uma tal demonstração não faz parte do objetivos principais deste trabalho.[14] Deixemos apenas enunciado que

Teorema 2.2.7. *Uma função é recursiva se e somente se é Turing-computável.*

[14] Nossa referência principal nesta seção — *Computabililty and Logic*, de George Boolos, 2007 — reserva seu oitavo capítulo inteiro a essa prova.

Perceba que, mesmo equivalentes, essas não devem ser, tampouco, as *duas* únicas formas de se obter decidibilidade. De fato: podemos definir tantas outras, como computabilidade por ábacos, ou por *algoritmos de Markov*, ou, ainda, por outro métodos mais complexos, como as funções computáveis no sentido de Herbrand-Gödel-Kleene, a noção de λ-definibilidade, devido a Alonzo Church (1903-1995), e a teoria de sistemas normais, de Emil Post (1897-1954). Acontece que *todas* essas ideias são equivalentes, cada qual tendo suas vantagens — e.g., máquinas de Turing e funções recursivas parecem ser as mais intuitivas e fáceis de se trabalhar, ao menos quando alguém inicia seus estudos sobre computabilidade.

Tantas formas equivalentes de se definir processos de decisão sugerem que, de fato, toda função efetivamente computável deve ser decidível por algum desses métodos (e, consequentemente, por todos eles). Mas, como já dissemos, uma prova formal para isso é, em princípio, intangível — podemos apenas nos basear na quantidade de evidências em seu favor. Nesse sentido, paralela à formulação de Turing, existe a *tese de Church*, que considera função recursiva e função efetivamente computável como sendo a mesma coisa. [A bem da verdade, Church foi pioneiro nessa ideia; mas como, à época, não se sabia que ambas as abstrações eram equivalentes, a tese de Turing ficou igualmente conhecida. Além disso, o próprio Church, assim como Gödel, reconheceram a simplicidade e a beleza das ideias de Turing, chegando o primeiro a dizer que a abstração por máquinas era superior às demais, principalmente porque ia direto ao ponto — i.e., caracterizava o conceito de processo efetivo de decisão sem precisar provar teoremas iniciais, ou assumir a computabilidade de funções básicas *a priori*. Segundo Gödel, as máquinas de Turing se apresentavam como uma ótima noção *epistemológica* da ideia de decisão, sem precisar apelar para o formalismo ortodoxo latente das teorias matemáticas em geral.[15] E, assim como a tese de Turing não foi a primeira do tipo, tampouco foi a última; praticamente toda abstração equivalente teve sua respectiva tese — no caso do sistema de Markov, por exemplo, aventou-se o que chamamos de *princípio de normalização*.]

Isto posto, dizemos que todas essas teses são formuladas em *linguagem coloquial*, ou seja, uma mistura de intuição com formalismo matemático. Assim, qualquer tentativa de se argumentar, contra ou a favor delas, embora não seja uma tarefa inútil nem sem sentido (pois é natural pensarmos que as abstrações que conhecemos podem não ser, de fato, as únicas), deve ser feita no campo da filosofia: não há maneira de se argumentar sobre o tema utilizando conceitos puramente matemáticos. Não fosse isso, a tese de Church poderia ser vista como uma mera definição nominal, i.e., dentro de uma teoria, poderíamos definir função efetivamente computável como sendo função recursiva. Evidentemente, tal definição não pode ser encontrada na literatura,

[15] Os textos contendo tais comentários podem ser encontrados em outra de nossas principais referências: *Recursive Functions and Metamathematics*, de Roman Murawski, 1999, p. 93.

pois o consenso de que essa é uma questão multidisciplinar já está estabelecido. Também é (quase)[16] consenso que, de fato, a tese deve valer; e, neste momento, cremos não ser mais necessário elencar os motivos pelos quais concordaremos em aceitá-la, como a maior parte dos autores faz.

[16] Alguns pares, como Kalmár (1959) e Bowie (1973), se esforçaram bastante em argumentações que negam a tese de Church. Nenhum dos dois, obviamente, foi capaz de mostrar uma função computável não-recursiva (pois se tivessem-no feito, não teríamos discutido esse assunto, para início de conversa), mas alguns resultados soam bastante peculiares. Por exemplo, Kalmár conseguiu mostrar como, assumindo a tese de Church, pode-se construir uma sentença falsa que é absolutamente indecidível (por mais contraditório que isso pareça).

3 INCOMPLETUDE DOS SISTEMAS ARITMÉTICOS

No capítulo anterior, conseguimos abstrair o conceito-chave, o alicerce da teoria da recursão, quando construímos uma classe de funções que podem ter seus valores decididos mecanicamente. Nossa tarefa, contudo, ainda tem um longo caminho adiante. Primeiramente, a fim de que o trabalho da seção anterior faça realmente sentido, precisamos formalizar a base que o sustenta — i.e., há de se definir precisamente *o que é* um sistema capaz de descrever relações entre números naturais, bem como deduzir os *teoremas* que versam sobre tais relações. Somente depois disso é que seremos capazes de mostrar como, de fato, as funções recursivas que apresentamos anteriormente podem ser representadas dentro de um sistema formal, o que será visto na seção 3.2. A partir disso, teremos bagagem o suficiente para tratar conceitos metamatemáticos *dentro* do sistema baseado em $\mathbb{N}$ — em particular, a ideia de "prova" terá sua caracterização puramente matemática. Tal conquista se dará na seção 3.3, onde conseguiremos *aritmetizar* (i.e., codificar em $\mathbb{N}$) tanto sentenças lógico-matemáticas usuais quanto sentenças de cunho metateórico, usando a linguagem especificada em 3.1. Em suma, finalmente teremos condições de falar *sobre* Matemática *usando* Matemática.

Na seção final do capítulo, mostraremos como produzir, a partir da aritmetização, sentenças indecidíveis no sistema formal em questão. Mais especificamente, na subseção 3.4.1, apresentaremos duas sentenças *autorreferentes*, i.e., que versam sobre si próprias, e que atestam o Primeiro Teorema: o sistema aritmético que construímos é incompleto. O Segundo Teorema é demonstrado na subseção 3.4.2, quando discutimos maneiras de se representar uma sentença que afirme a consistência de nosso sistema formal, e mostramos que qualquer delas é indecidível. Consequências lógicas imediatas, desmistificações sobre teoremas de incompletude, dentre outros comentários são apresentados na subseção 3.4.3.

3.1 ARITMÉTICA DOS NÚMEROS NATURAIS

A caracterização do sistema aritmético formal que usaremos se deve, em grande parte, ao ao matemático italiano Giuseppe Peano (1858-1932), um dos desbravadores da lógica-matemática, ao lado de Frege.[1] Assim sendo, chamemos tal sistema pela sigla PA (do inglês *Peano Arithmetic*). Formalmente, no nosso entendimento, a Aritmética de Peano é uma teoria axiomática baseada em lógica de primeira ordem, cujos indivíduos (e interações entre indivíduos) devem ser interpretados como números naturais (e relações entre números naturais). Abaixo, especificamos os diferentes tipos de símbolos presentes em sua linguagem. Temos:

[1] "Em grande parte" porque a formulação de Peano, datada de 1889, não embutia teoria dos conjuntos, como será o nosso caso.

- as *constantes lógicas*: $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$, $\exists$ e $\forall$; sua semântica é a usual;
- as *variáveis individuais*: uma quantidade infinita enumerável de variações subscritas dos símbolos x, y, z (e.g., $x_1, x_2, ..., x_n, ...$);
- uma *constante individual*: 0, que lemos "zero", em português;
- um símbolo de *predicado binário*: =, que lemos "é igual a";
- um símbolo de *função unária*: s, chamada "sucessor de";
- símbolos para as *funções binárias* de adição e multiplicação: + e $*$, respectivamente;
- dois símbolos *técnicos*, responsáveis por "pontuar" as expressões artiméticas: os parênteses esquerdo, (, e direito,).

Isso configura uma espécie de "alfabeto" de PA, junto de seus "sinais de pontuação"; mas, assim, como a gramática da língua portuguesa nos impede de, por exemplo, dar qualquer interpretação à sequência de símbolos

;;hUTs...TCalççíbb—C??ASloló!!ã,

dizendo que isso não constitui nem uma palavra, nem uma frase, nem uma oração (enfim, nada que um ser humano versado em português conseguiria ler), precisamos dizer quais sequências de símbolos de PA são *expressões bem formadas*.[2] Vamos distingui-las entre o que chamamos de *termos* e *fórmulas*:

Definição 3.1.1. (i) Toda variável individual x_i, bem como a constante 0, são *termos* da linguagem de PA; (ii) se α é um termo, então s(α) também o é; (iii) se α e β são termos, então as expressões (α) + (β) e (α) $*$ (β) também o são; (iv) não há outros termos na linguagem de PA além dos listados em (i) e dos obtidos, através destes, pelas regras expostas em (ii) e (iii).

Definição 3.1.2. (i) Dados quaisquer termos α e β em PA, temos que a expressão $\alpha = \beta$ é uma *fórmula* da linguagem de PA; (ii) se φ é uma fórmula, então as expressões $\neg(\varphi)$, $\exists x_i(\varphi)$ e $\forall x_i(\varphi)$ também o são; (iii) se φ e ψ são fórmulas, então as expressões

[2] A rigor, uma expressão bem formada dentro de um sistema formal não precisa ter uma "boa aparência", ou ser intuitiva e reconhecível aos olhos de quem lê. Em particular, a bizarra sequência de símbolos da língua portuguesa que apresentamos acima poderia, de fato, ser considerada uma expressão bem formada, em algum sistema formal que use o abecedário comum ao português e os mesmos sinais de pontuação (por exemplo, nada impede que, neste momento, essa mesma expressão bizarra seja o conteúdo de uma mensagem ultra-confidencial criptografada, a ser decodificada em alguma língua — que pode nem ser a língua portuguesa). Ou seja, *boa formação* é, *grosso modo*, uma questão de convenção. Entretanto, quando estamos construindo um sistema, inerentemente estamos levando em conta a interpretação que daremos às sequências de símbolos da linguagem, e é por isso que escolhemos determinadas regras para boa formação de expressões em detrimento de outras.

$(\varphi) \vee (\psi)$, $(\varphi) \wedge (\psi)$, $(\varphi) \rightarrow (\psi)$ e $(\varphi) \leftrightarrow (\psi)$ também o são; (iv) não há outras fórmulas da linguagem de PA além das que podem ser obtidas através das regras presentes nos itens anteriores.

Para que nossa notação não fique demasiadamente carregada, combinemos que dada uma fórmula qualquer, são desnecessários os parênteses quando esta é uma negação ou uma fórmula *simples* — i.e., não pode ser subdividida em subfórmulas, não tendo conectivos internos a ela. Além disso, convenciona-se que os símbolos $\vee$ e $\wedge$ operam mais fortemente do que $\rightarrow$ e $\leftrightarrow$, assim como $*$ tem prioridade sobre $+$. Também, será conveniente fazer uso de outros sinais técnicos, como colchetes, [,], e chaves, {, }, para melhor compreensão de expressões com muitas subfórmulas. Dessa maneira, tomando φ, ψ e θ fórmulas quaisquer de PA, e entendendo que coisas como $\varphi(x)$, $\psi(y)$ e $\theta(z)$ são fórmulas aplicadas a tais variáveis — i.e., expressões como $x = x$, $y = y$, $y = z$, etc., ou expressões obtidas dessas adicionando conectivos lógicos —, será útil convencionar que, em vez de escrevermos

$$\exists x(\forall y(((\varphi(x)) \wedge (\psi(y))) \rightarrow \exists z((\psi((x * y) + z)) \leftrightarrow (\theta(z))))),$$

façamos o seguinte, deixando a leitura muito mais clara:

$$\exists x\{\forall y[\varphi(x) \wedge \psi(y) \rightarrow \exists z(\psi(x * y + z) \leftrightarrow \theta(z))]\}.$$

Neste momento, estamos em posse, apenas, dos elementos linguísticos de PA. Precisamos, então, escolher um conjunto de axiomas que seja completo o suficiente para caracterizar os conceitos aritméticos presentes na teoria de números. Podemos fazer uma distinção entre três conjuntos de axiomas de PA: os axiomas *lógicos* (referentes aos conectivos e quantificadores), os de *identidade* (referentes ao predicado binário $=$) e os *não-lógicos*, i.e., os que fazem referência exclusiva aos números naturais e suas interações entre si. Considerando p, q, r como variáveis proposicionais que podem ser substituídas por fórmulas de PA, o primeiro grupo de axiomas pode ser especificado da seguinte maneira:

$$\begin{aligned}
&(L1) \quad p \rightarrow (q \rightarrow p);\\
&(L2) \quad \neg\neg p \rightarrow p;\\
&(L3) \quad p \rightarrow \neg\neg p;\\
&(L4) \quad p \wedge q \rightarrow p;\\
&(L5) \quad p \wedge q \rightarrow q;\\
&(L6) \quad p \rightarrow p \vee q;\\
&(L7) \quad q \rightarrow p \vee q;
\end{aligned}$$

$$(L8) \quad (p \leftrightarrow q) \rightarrow (p \rightarrow q);$$
$$(L9) \quad (p \leftrightarrow q) \rightarrow (q \rightarrow p);$$
$$(L10) \quad (p \rightarrow q) \rightarrow (\neg q \rightarrow \neg p);$$
$$(L11) \quad (p \rightarrow q) \rightarrow [(q \rightarrow p) \rightarrow (p \leftrightarrow q)];$$
$$(L12) \quad (p \rightarrow q) \rightarrow [(p \rightarrow r) \rightarrow (p \rightarrow q \wedge r)];$$
$$(L13) \quad (p \rightarrow r) \rightarrow [(q \rightarrow r) \rightarrow (p \vee q \rightarrow r)];$$
$$(L14) \quad [p \rightarrow (q \rightarrow r)] \rightarrow [(p \rightarrow q) \rightarrow (p \rightarrow r)].$$

Esse é um conjunto completo de axiomas para o cálculo proposicional (i.e., lógica de ordem zero), pois todas as tautologias da mesma podem ser derivadas a partir deles, por meio das regras de substituição proposicional[3] e *modus ponens* (que especificaremos mais adiante). A seguir, apresentamos os axiomas de identidade:

$$(I1) \quad \forall x(x = x);$$
$$(I2) \quad \forall x \forall y(x = y \rightarrow y = x);$$
$$(I3) \quad \forall x \forall y(x = y \rightarrow \mathrm{s}(x) = \mathrm{s}(y));$$
$$(I4) \quad \forall x \forall y(x = y \wedge y = z \rightarrow x = z);$$
$$(I5) \quad \forall x \forall y \forall z(x = y \rightarrow x + z = y + z);$$
$$(I6) \quad \forall x \forall y \forall z(x = y \rightarrow z + x = z + y);$$
$$(I7) \quad \forall x \forall y \forall z(x = y \rightarrow x * z = y * z);$$
$$(I8) \quad \forall x \forall y \forall z(x = y \rightarrow z * x = z * y).$$

Finalmente, temos os axiomas que concernem especificamente à álgebra dos números naturais — os chamados *axiomas de Peano*:

$$(P1) \quad \neg \exists x[0 = \mathrm{s}(x)];$$
$$(P2) \quad \forall x \forall y[\mathrm{s}(x) = \mathrm{s}(y) \rightarrow x = y];$$
$$(P3) \quad \forall x(x + 0 = x);$$
$$(P4) \quad \forall x \forall y[x + \mathrm{s}(y) = \mathrm{s}(x + y)];$$
$$(P5) \quad \forall x(x * 0 = 0);$$
$$(P6) \quad \forall x \forall y[x * \mathrm{s}(y) = x * y + x];$$
$$(P7) \quad \varphi(0) \wedge \forall x[\varphi(x) \rightarrow \varphi(\mathrm{s}(x))] \rightarrow \forall x \varphi(x).$$

Perceba que ($P7$), que traduz a indução finita, enuncia *uma infinidade* de axiomas (um para cada fórmula envolvendo números naturais), diferentemente dos demais;

[3] Também chamada de *substituição atômica*, ou simplesmente *substituição*, essa regra diz que dada uma fórmula q qualquer, escrita na linguagem do cálculo proposicional e composta de $n \in \mathbb{N}$ fórmulas atômicas $p_1, ..., p_n$, se q for válida, então a fórmula q', resultante de q pela substituição simultânea de $p_1, ..., p_n$ por, respectivamente, fórmulas quaisquer $r_1, ..., r_n$, também deve ser válida. Ilustrativamente, como $p \rightarrow p$ é válida, $q \wedge r \rightarrow q \wedge r$ também deve ser. Nesse trabalho, no entanto, preferimos reservar o nome "substituição" à regra de inferência de PA, especificada mais adiante.

assim, chamamos (*P*7) de um *esquema de axioma*.[4]. Mas o fato de haver uma quantidade enumerável de axiomas não-lógicos não chega a ser um empecilho para o nosso trabalho: desde que haja um processo mecânico para decidir se determinada sentença é uma instância deste esquema (o que, de fato, sempre acontece, tendo em vista o caráter recursivo da indução finita), nosso conjunto de axiomas continua sendo um conjunto recursivo — i.e., decidível. Veja, ainda, que (*P*1) e (*P*2) mostram que a função sucessor é injetiva, e que sua imagem é $\mathbb{N} - \{0\}$. Os axiomas (*P*3) a (*P*6) nada mais são do que as definições recursivas de adição e multiplicação. [Aqui, também é importante notar o seguinte: diferentemente do modo como definimos a classe $\mathcal{R}$, no capítulo anterior, nossa caracterização de PA pressupõe, além da função sucessor, as funções adição e multiplicação, e não as funções zero e as projeções. Entretanto, não há nada que nos impossibilite de mudar a construção de PA, de modo que os símbolos destas últimas funções sejam incluídos de antemão; apenas não o fazemos porque o sistema PA, como definido acima, é suficiente para provar tudo que precisamos, e também é a maneira mais usual de se descrever a álgebra dos inteiros não-negativos.]

Com a estrutura axiomática em mãos, resta-nos especificar a estrutura argumentativa de PA, por meio da qual será possível derivar seus teoremas. Fazemos isso listando uma série de *regras de inferência*. Como mencionamos acima, faremos uso de *modus ponens*, uma regra que exprime a peculiaridade do conectivo $\rightarrow$: dada uma implicação, o único caso em que podemos obter uma falsidade é quando o antecedente é verdadeiro e o consequente é falso; sendo assim, tendo $\varphi \rightarrow \psi$ e φ como teoremas, devemos poder derivar ψ como um teorema também. Outra regra muito importante é a de *substituição* — que, aqui, entenderemos como a substituição de *variáveis* por *termos* da linguagem. Respectivamente, temos:

$$\frac{\varphi \rightarrow \psi, \quad \varphi}{\psi} \qquad \frac{\varphi}{\varphi(x/\alpha)}$$

[A notação $\varphi(x/\alpha)$ significa que, em φ, podemos substituir a variável x por qualquer termo α; isso, no entanto, pode acontecer somente se α for *substituível*, i.e., caso satisfaça a seguinte condição: dada uma variável y qualquer ocorrendo em α, não existe nenhuma parte de φ nas formas $\forall y(\psi)$ ou $\exists y(\psi)$ tais que x ocorre livremente em φ.] Outras duas regras são as *eliminações de quantificadores* — respectivamente, universal e existencial:

$$\frac{\varphi \rightarrow \forall x(\psi)}{\varphi \rightarrow \psi} \qquad \frac{\exists x(\varphi) \rightarrow \psi}{\varphi \rightarrow \psi}$$

[4] Nesse sentido, prova-se que nenhum conjunto finito de axiomas para PA pode ser construído; em particular, (*P*7) não pode ser substituído por uma quantidade finita de axiomas. O leitor pode se perguntar se não seria possível tratar PA como uma teoria de segunda ordem — pois, assim, (*P*7) deixaria de ser um *esquema*, já que poderíamos estipular variáveis sobre conjuntos de tamanhos quaisquer. Ocorre que, de fato, podemos fazê-lo, como vários campos da Matemática o fazem, e um tal sistema seria capaz de descrever coisas que PA não consegue. Nossa escolha por trabalhar com uma lógica de primeira ordem, no entanto, se dá justamente pelo caráter elementar de PA. Além disso, a lógica de segunda ordem é incompleta, diferentemente da lógica de primeira ordem.

E, finalmente, supondo que x não ocorre livre em, respectivamente, φ e ψ, temos as *introduções de quantificadores universal e existencial*:

$$\frac{\varphi \to \psi}{\varphi \to \forall x(\psi)} \quad \Bigg| \quad \frac{\varphi \to \psi}{\exists x(\varphi) \to \psi}$$

A partir dessas "leis" e das tautologias da lógica proposicional, ainda podemos definir a regra auxiliar de *generalização*, que obtém $\forall x(\varphi)$ de φ. Com isso, é possível definir o que, formalmente, é um *teorema* de PA:

Definição 3.1.3. Todas as sentenças (i.e., fórmulas fechadas) dos axiomas de PA (i.e., todas as instâncias dos axiomas lógicos, os axiomas não-lógicos e os axiomas de identidade) são *teoremas* de PA. Toda sentença obtida de um teorema por alguma regra de inferência especificada acima é, também, um teorema de PA. Nada, além disso, é um teorema de PA.

Isto posto, se φ for um teorema de PA, escreveremos $\text{PA} \vdash \varphi$ (caso contrário, faremos $\text{PA} \nvdash \varphi$).

Assim como podemos definir novas regras de inferência, é possível designar símbolos predicativos e funcionais a novas relações, definíveis a partir dos que já possuímos. Por exemplo, será útil atribuir $<$ à relação *menor que*, definida por

$$\forall x \forall y[x < y \leftrightarrow \exists z(z \neq 0 \wedge y = x + z)].$$

Disso, segue diretamente que

Lema 3.1.1. *São verdadeiras as proposições:*

1. $\text{PA} \vdash \neg(x < 0)$;
2. $\text{PA} \vdash x < s(y) \leftrightarrow x < y \vee x = y$;
3. $\text{PA} \vdash x < y \vee x = y \vee y < x$.

Também, será importante enunciar, para uso posterior, os seguintes teoremas sobre equivalência e igualdade entre fórmulas e termos. Aqui, omitimos as demonstrações:

Teorema 3.1.1. *Dada φ uma fórmula qualquer na linguagem de* PA*, considere φ' como sendo a fórmula obtida de φ ao substituirmos as ocorrências de suas sub-fórmulas $\psi_1, \psi_2, ..., \psi_n$ (não necessariamente todas) por $\psi'_1, \psi'_2, ..., \psi'_n$, respectivamente. Se para todo $i \leq n$,* $\text{PA} \vdash \psi_i \leftrightarrow \psi'_i$*, então,*

$$\text{PA} \vdash \varphi \leftrightarrow \varphi'.$$

Teorema 3.1.2. *Dado α um termo qualquer na linguagem de* PA*, considere α' como sendo o termo obtido de α ao substituirmos as ocorrências de seus sub-termos $\beta_1, \beta_2, ..., \beta_n$, nenhum no escopo de algum quantificador, e não necessariamente todos eles, por $\beta'_1, \beta'_2, ..., \beta'_n$, respectivamente. Além disso, para uma fórmula φ arbitrária, seja φ' obtida pelo mesmo tipo de substituição. Se para todo $i \leq n$,* $\mathsf{PA} \vdash \beta_i = \beta'_i$*, então,*

$$\mathsf{PA} \vdash \alpha = \alpha' \quad e \quad \mathsf{PA} \vdash \varphi \leftrightarrow \varphi'.$$

Na última nota de rodapé, evidenciamos como PA, sendo uma teoria baseada em lógica de primeira ordem, não deixa margem para dúvidas quanto à interpretação de suas variáveis — pois, já de início, *estabelecemos* que estas devem se referir ao zero e aos números obtidos dele indutivamente, e porque há somente um tipo de variável nessa teoria. No entanto, a interpretação apresentada aqui, que pode ser abreviada por $\mathfrak{N}_0 = \langle \mathbb{N}, 0, \mathsf{s}, +, * \rangle$ e que chamamos de *interpretação padrão*, ou *modelo standard*, não é a única possível para PA — há, inclusive, modelos não isomorfos a $\mathfrak{N}_0$, conhecidos pelo nome de *não-standard*. Nesse sentido, pode-se dizer que PA foi construída com o objetivo de dar forma à nossa visão intuitiva dos números naturais, modelada por $\mathfrak{N}_0$. Uma discussão mais detalhada sobre teoria de modelos, todavia, foge de nosso escopo,[5] embora ainda façamos referência a alguns conceitos no próximo capítulo. A dúvida que surge é: será que nossa construção de PA, bem como qualquer outra formalização da aritmética, é *adequada* para descrever as teorias matemáticas que versam sobre $\mathbb{N}$? É claro que ainda precisamos definir que tipo de "adequação" seria essa, mas veremos que, segundo os teoremas de Gödel, a resposta deve ser negativa, não importando qual caracterização seja dada para a aritmética dos naturais.

3.2 REPRESENTABILIDADE NA ARITMÉTICA DE PEANO

Em oposição ao que foi dito logo acima, existe ao menos um contexto no qual PA *é* adequada: dentro dela, podemos *representar* todas as funções recursivas. Ou seja, mesmo que haja incerteza quanto à veracidade de uma sentença aritmética na linguagem de PA, é possível, por meio desta, "falar sobre" qualquer relação efetivamente computável entre números naturais (assumindo a tese de Church). Ora, isso parece bastante claro para alguns casos, como adição e multiplicação, mas tal obviedade ocorre justamente porque designamos símbolos especiais para exprimir tais funções.[6] Para formalizar essa ideia, comecemos especificando notação e terminologia. Podemos distinguir os termos de PA, subdividindo-os em *termos variáveis* e *termos*

[5] Para se aprofundar em tais conceitos, sugerimos os capítulos 12 e 13 de *Computability and Logic*, 2007; ou, ainda, cf. (BUTTON; WALSH, 2018)

[6] Além disso, em decorrência da prática comum de se padronizar notação, alguns símbolos, como + e $*$, se entranham em nosso intelecto de tal forma que poucas dúvidas surgem quanto ao fato de, e.g., a fórmula $z = x * y$ representar a multiplicação.

constantes (ou *numerais*). Indutivamente, diremos que 0 é um *numeral*, e, caso α seja um numeral, então s(α) também o será. Convencionando que

$$\underbrace{\mathrm{s}(\mathrm{s}(...\mathrm{s}}_{n \text{ vezes}}(0)...))) = \underbrace{\mathrm{ss...s}}_{n \text{ vezes}}(0) = \overline{n},$$

de modo que, por exemplo, $\overline{5}$ = sssss(0), $\overline{n}$ pode ser visto como um outro nome, mais compacto, para o numeral ss...s(0).

Definição 3.2.1. Dadas φ uma fórmula com n variáveis livres, escrita na linguagem de PA, e $R \subset \mathbb{N}^n$ uma relação, dizemos que

(i) φ *representa fracamente* R se e somente se para quaisquer $k_1, ..., k_n \in \mathbb{N}$, verifica-se que

$$R(k_1, ..., k_n) \text{ se e somente se PA} \vdash \varphi(\overline{k_1}, ..., \overline{k_n}).$$

(ii) φ *representa fortemente* R se e somente se para quaisquer $k_1, ..., k_n \in \mathbb{N}$, verifica-se que

$$\text{se } R(k_1, ..., k_n), \text{ então PA} \vdash \varphi(\overline{k_1}, ..., \overline{k_n});$$
$$\text{se } \neg R(k_1, ..., k_n), \text{ então PA} \vdash \neg\varphi(\overline{k_1}, ..., \overline{k_n}).$$

Assim, tomando uma relação R arbitrária, e supondo que ela seja n-ária, dizemos que ela é (i) *fracamente representável* caso haja alguma fórmula φ, com n variáveis livres, que a represente fracamente; e (ii) *fortemente representável* caso haja alguma fórmula φ, com n variáveis livres, que a represente fortemente.

Perceba que, se PA for consistente (i.e., caso não aconteça PA $\vdash \psi \wedge \neg\psi$, qualquer que seja a fórmula ψ), então as implicações citadas na definição 3.2.1.(ii) se tornam equivalências. Assim, concluímos que as relações *fortemente* representáveis são aquelas representadas *fracamente* por alguma fórmula φ, e cujos complementos, i.e., as uplas de elementos que *não se relacionam*, são fracamente representados por $\neg\varphi$, a negação de φ. De maneira similar, podemos definir representabilidade de funções:

Definição 3.2.2. Dadas φ uma fórmula com $n+1$ variáveis livres, escrita na linguagem de PA, e $f : \mathbb{N}^n \longrightarrow \mathbb{N}$ uma função, dizemos que φ *representa* f se e somente se para $k_1, ..., k_n \in \mathbb{N}$ arbitrários, verifica-se que

$$\text{PA} \vdash \forall y[\varphi(\overline{k_1}, ..., \overline{k_n}, y) \leftrightarrow (y = \overline{f(k_1, ..., k_n)})],$$

Assim, $f : \mathbb{N}^n \longrightarrow \mathbb{N}$ é dita *representável* em PA caso haja alguma fórmula φ com $n+1$ variáveis livres que a represente em PA.

Nosso trabalho, de agora em diante, será provar que todos os elementos de $\mathcal{R}$ — i.e., que todas as relações recursivas — são representáveis (no sentido dado pelas definições 3.2.1 e 3.2.2). Comecemos mostrando, rigorosamente, como a relação de identidade é representável na linguagem de PA:

Teorema 3.2.1. *A fórmula $x = y$ representa fortemente a relação de identidade.*

Demonstração. Tomando $m, n \in \mathbb{N}$ arbitrários, devemos mostrar que

$$\text{se } m = n, \text{ então PA} \vdash \overline{m} = \overline{n};$$
$$\text{se } m \neq n, \text{ então PA} \vdash \neg(\overline{m} = \overline{n}).$$

É claro que se os números m e n são iguais, então seus numerais devem ser o mesmo; do axioma ($I1$) de identidade, vem que PA $\vdash \overline{m} = \overline{n}$. Quanto à segunda implicação, usemos indução em n para provar que

$$\forall m[m \neq n \rightarrow \text{PA} \vdash \neg(\overline{m} = \overline{n})].^{7}$$

Fazendo $n = 0$, devemos ter $m \neq 0$. Como o axioma ($P1$) enuncia que $\forall x \neg(0 = \mathrm{s}(x))$, concluímos que PA $\vdash \neg(\overline{m} = 0)$. Tomando $n > 0$, temos como hipótese de indução que

$$\forall m[m \neq n - 1 \rightarrow \text{PA} \vdash \neg(\overline{m} = \overline{n-1})].$$

[Aqui, $n - 1$ é apenas uma notação para o antecessor de $n \in \mathbb{N}$. Adiante, indicamos um modo como representá-la em PA.] Sendo m um número natural qualquer, suponha que $m \neq n$. Assim, tem-se também que $m - 1 \neq n - 1$, e da hipótese acima segue que

$$m - 1 \neq n - 1 \rightarrow \text{PA} \vdash \neg(\overline{m-1} = \overline{n-1}),$$

ou seja, por *modus ponens* temos que PA $\vdash \neg(\overline{m-1} = \overline{n-1})$. Da injetividade da função sucessor, vem que PA $\vdash (\overline{m} = \overline{n}) \rightarrow (\overline{m-1} = \overline{n-1})$; assim, pela contrapositiva, obtemos:

$$\text{PA} \vdash \neg(\overline{m-1} = \overline{n-1}) \rightarrow \neg(\overline{m} = \overline{n}),$$

Novamente por *modus ponens*, ficamos com PA $\vdash \neg(\overline{m} = \overline{n})$, e temos o resultado. □

Tendo em vista os dois próximos lemas, provemos que PA $\vdash \overline{m} + \overline{n} = \overline{m+n}$, e que PA $\vdash \overline{m} * \overline{n} = \overline{m*n}$. [No primeiro caso, isso significa que a representação da soma (i.e., adição em $\mathbb{N}$) de dois *números* é a mesma que a da soma (i.e., adição em PA) de dois *numerais*.] Com efeito: por indução em n, seja $n = 0$. Então, via ($P3$) e substituição, temos que

$$\text{PA} \vdash \overline{m} + 0 = \overline{m} = \overline{m+0}.$$

Assuma, então, que a tese vale para n. Pelo axioma ($P4$), vem que

$$\text{PA} \vdash \overline{m} + (\overline{n+1}) = \overline{m} + \mathrm{s}(\overline{n}) = \mathrm{s}(\overline{m} + \overline{n}).$$

[7] Perceba que tal processo de indução acontece na metateoria.

Além disso, pela hipótese de indução, $\mathrm{s}(\overline{m}+\overline{n}) = \mathrm{s}(\overline{m+n}) = (\overline{m+n})+\overline{1}$, logo $\overline{m}+(\overline{n+1}) = (\overline{m+n})+\overline{1}$. Da definição de numeral, temos que $\mathrm{s}(\overline{m+n})$ pode ser visto como tal, de modo que escrevemos $\mathrm{s}(\overline{m+n}) = \overline{\mathrm{s}(m+n)} = \overline{m+n+1}$. Finalmente, usando *modus ponens* e a transitividade da identidade, temos:

$$\mathrm{PA} \vdash \overline{m} + (\overline{n+1}) = \overline{m+n+1}.$$

Para a multiplicação, temos que, no caso base, $\mathrm{PA} \vdash \overline{m} * 0 = 0 = 0$, em decorrência de ($P5$) e da regra de substituição. Assumindo que o resultado vale para um $n \in \mathbb{N}$ qualquer, usamos o axioma ($P6$) para derivar

$$\mathrm{PA} \vdash \overline{m} * \mathrm{s}(\overline{n}) = (\overline{m} * \overline{n}) + \overline{m} = (\overline{m*n}) + \overline{m}.$$

Pelo caso da adição, vem que $\mathrm{PA} \vdash (\overline{m*n}) + \overline{m} = \overline{m*n+m} = \overline{m*\mathrm{s}(n)}$. Com isso, chegamos ao resultado:

$$\mathrm{PA} \vdash \overline{m} * (\overline{n+1}) = \overline{m*(n+1)}.$$

Teorema 3.2.2. *A fórmula $x + y = z$ representa a adição de números naturais.*

Demonstração. Tome $m, n, k \in \mathbb{N}$ tais que $m + n = k$. Precisamos provar que

$$\mathrm{PA} \vdash (\overline{m} + \overline{n} = y) \leftrightarrow (y = \overline{k}).$$

Usando os axiomas ($I2$) e ($I4$) de identidade, substituindo as variáveis x, y, z apropriadamente, e pela regra de *exportação*[8], temos que

$$\mathrm{PA} \vdash (x = z) \to (x = y \to y = z).$$

Colocando $\overline{m} + \overline{n}$ e $\overline{k}$ no lugar de, respectivamente, x e z, obtemos:

$$\mathrm{PA} \vdash (\overline{m} + \overline{n} = \overline{k}) \to (\overline{m} + \overline{n} = y \to y = \overline{k}).$$

Como $m + n = k$, temos que $\mathrm{PA} \vdash \overline{m+n} = \overline{k}$; mas também é verdade que $\mathrm{PA} \vdash \overline{m+n} = \overline{m} + \overline{n}$, logo $\mathrm{PA} \vdash \overline{m} + \overline{n} = \overline{k}$. Por *modus ponens*, segue que

$$\mathrm{PA} \vdash (\overline{m} + \overline{n} = y) \to (y = \overline{k}).$$

De maneira similar, pode-se provar a recíproca. O resultado segue. □

Teorema 3.2.3. *A fórmula $x * y = z$ representa a multiplicação de números naturais.*

Demonstração. Basta provar que $\mathrm{PA} \vdash (\overline{m} * \overline{n} = y) \leftrightarrow (y = \overline{k})$, em que $k = m * n$, e y é uma variável qualquer na linguagem de PA. O mesmo raciocínio do teorema 3.2.2 nos dá o que precisamos. □

[8] Dadas variáveis proposicionais p, q, r quaisquer, tem-se que, numa prova, $(p \wedge q) \to r$ pode ser substituída por $p \to (q \to r)$.

Com esses três resultados, vemos que duas das funções — e uma das relações — mais elementares da aritmética podem ser representadas dentro do sistema formal que escolhemos. Mas isso já era mais do que esperado, uma vez que desde o início, na definição de PA, estipulamos signos para tais operações.[9] Convém, portanto, elencar alguns exemplos menos corriqueiros. A função *antecessor*, que provamos ser recursiva no capítulo passado, pode ser representada pela seguinte fórmula, considerando que y é imagem de x por tal função:

$$(x = 0 \wedge y = 0) \vee x = \mathrm{s}(y).$$

Também, as funções *quociente* e *resto da divisão euclidiana* são representadas, respectivamente, por

$$(y = 0 \wedge z = 0) \vee \exists u < y(x = y * z + u);$$
$$(y = 0 \wedge z = x) \vee [z < y \wedge \exists u \leq x(x = u * y + z)],$$

em que, também de maneira respectiva, z é o quociente e o resto da divisão de x por y. Já o próximo exemplo merece uma boa demonstração:

Teorema 3.2.4. *A relação de ordem estrita (ou "menor que") é fortemente representada pela fórmula $x < y$.*

Demonstração. Provemos, por indução metateórica, as duas implicações da definição 3.2.1.(ii). Colocando $n = 0$ e supondo $m < n$, ficamos com $m < 0$, o que é impossível em $\mathbb{N}$. Por vacuidade, *"se $m < 0$, então* $\mathrm{PA} \vdash \overline{m} < 0$*"* é verdadeira. Agora, sendo $\neg(m < 0)$, o lema 3.1.1(1) e a regra de substituição devem nos dar

$$\mathrm{PA} \vdash \neg(\overline{m} < 0).$$

Isso prova que *"se* $\neg(m < 0)$ *então* $\mathrm{PA} \vdash \neg(\overline{m} < 0)$*"* é verdadeira, e finalizamos o caso base. Suponha, para a hipótese de indução, que as duas implicações metateóricas valem para $n \in \mathbb{N}$. Pelo lema 3.1.1(2), obtemos:

$$\mathrm{PA} \vdash (\overline{m} < \overline{n+1}) \leftrightarrow (\overline{m} < \overline{n} \vee \overline{m} = \overline{n}). \tag{3}$$

Para a primeira implicação, faça $m < n+1$. Então, ou $m < n$ ou $m = n$; no primeiro caso, temos, pela hipótese de indução, que $\mathrm{PA} \vdash \overline{m} < \overline{n}$, e da tautologia $p \rightarrow p \vee q$ vem que $\mathrm{PA} \vdash \overline{m} < \overline{n} \vee \overline{m} = \overline{n}$. Usando (3), obtemos:

$$\mathrm{PA} \vdash \overline{m} < \overline{n+1}.$$

Se $m = n$, então $\mathrm{PA} \vdash \overline{m} = \overline{n}$, pois a identidade é representável (vide teorema 3.2.1). Pela mesma tautologia acima, $\mathrm{PA} \vdash \overline{m} < \overline{n} \vee \overline{m} = \overline{n}$, e daí $\mathrm{PA} \vdash \overline{m} < \overline{n+1}$. Isso esgota os casos da primeira implicação.

[9] Mesmo assim, as demonstrações rigorosas que demos para tais fatos não se tornam menos interessantes ou menos necessárias.

Para a segunda parte, suponha que $\neg(m < n+1)$, de modo que $n+1 \leq m$, i.e., $n < m$. Da hipótese de indução para esse caso, vem que PA $\vdash \neg(\overline{m} < \overline{n})$. Também, tem-se que $\neg(m = n)$, do que segue que PA $\vdash \neg(\overline{m} = \overline{n})$. Usando a tautologia $p \to (q \to p \wedge q)$, chegamos a

$$\text{PA} \vdash \neg(\overline{m} < \overline{n}) \wedge \neg(\overline{m} = \overline{n}).$$

Por fim, perceba que a lei de De Morgan aplicada a (3) nos dá

$$\text{PA} \vdash \neg(\overline{m} < \overline{n+1}) \leftrightarrow \neg(\overline{m} < \overline{n}) \wedge \neg(\overline{m} = \overline{n}),$$

donde concluímos que PA $\vdash \neg(\overline{m} < \overline{n+1})$, e a prova está completa. □

Mas voltemos nossa atenção às *funções básicas* de $\mathcal{R}$. Como vimos, a relação de identidade se traduz, em PA, por $y = x$; desse modo, podemos encarar a função id^1_1 como sua equivalente — i.e., id^1_1 também é representada por $y = x$. Indo além, cada projeção id^n_k pode ser representada pela fórmula $y = x_k$, ou, mais geralmente, por

$$x_1 = x_1 \wedge ... \wedge x_n = x_n \wedge y = x_k.$$

Perceba também que, a rigor, não estabelecemos a representabilidade da função sucessor, mesmo fazendo uso dela em todas as demonstrações anteriores. No entanto, tendo em vista seu caráter unário e considerando que ela exprime o processo mais corriqueiro da teoria de números (i.e., o ato de contar), digamos simplesmente que $y = \mathrm{s}(x) = x + \overline{1}$ representa o sucessor de x (embora uma prova formal faça uso apenas da definição de numeral e, como de costume, de indução metateórica). Para completar o elenco das funções que definem $\mathcal{R}$, temos que *zero*, i.e., $\mathrm{z}(x) = 0$ para todo x, é representada por $x = x \wedge y = 0$. De fato: neste caso, o que precisa ser provado é que

$$\text{PA} \vdash \forall y[(\overline{n} = \overline{n} \wedge y = 0) \leftrightarrow (y = 0)].$$

Ora, sendo $x = x$ uma fórmula válida — como também o é $p \wedge \top \leftrightarrow p$, em que $\top$ é uma tautologia qualquer —, tem-se o resultado. Com isso, obtemos o seguinte lema:

Lema 3.2.1. *As funções zero, sucessor e todas as projeções são representáveis na linguagem de* PA*.*

A fim de tornar a demonstração do próximo teorema mais simples, precisaremos de mais alguns resultados auxiliares:

Lema 3.2.2. *Dada $R \subset \mathbb{N}^p$ uma relação qualquer, temos que R é fortemente representável em* PA *se e somente se sua função característica, K_R, também o é.*

Demonstração. Suponha que $\varphi(x_1, ..., x_k)$ representa fortemente a relação R. Nosso objetivo é mostrar que sua função característica pode ser representada por uma fórmula $(k+1)$-ária, digamos, $\psi(x_1, ..., x_k, y)$, tal que

$$(\varphi(x_1, ..., x_k) \wedge y = \overline{1}) \vee (\neg\varphi(x_1, ..., x_k) \wedge y = 0).$$

Tomando $n_1, ..., n_k \in \mathbb{N}$ arbitrários, e assumindo que $K_R(n_1, ..., n_k) = 1$, tem-se que a relação vale para tais números, i.e., $R(n_1, ..., n_k)$. Daí,

$$\text{PA} \vdash \varphi(\overline{n_1}, ..., \overline{n_k}). \tag{4}$$

Agora, devemos provar que neste caso — em que os números $n_1, ..., n_k$ estão relacionados por R —, ambas $\varphi(\overline{n_1}, ..., \overline{n_k})$ e $y = \overline{1}$ devem ocorrer simultaneamente, i.e., deve haver a equivalência:

$$\text{PA} \vdash \psi(\overline{n_1}, ..., \overline{n_k}, y) \leftrightarrow (y = \overline{1}). \tag{5}$$

Primeiramente, note que $p \rightarrow [q \rightarrow (p \wedge q)]$ é uma tautologia (chamada de *lei da conjunção*). Assim, fazendo as substituições apropriadas, temos que

$$\text{PA} \vdash \varphi(\overline{n_1}, ..., \overline{n_k}) \rightarrow [y = \overline{1} \rightarrow (\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = \overline{1})].$$

Usando isso, (4) e *modus ponens*, obtemos:

$$\text{PA} \vdash y = \overline{1} \rightarrow (\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = \overline{1}). \tag{6}$$

Novamente por uma tautologia (nesse caso, o axioma lógico $p \rightarrow p \vee q$), temos que

$$\begin{aligned}\text{PA} \vdash (\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = \overline{1}) \rightarrow (&\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = \overline{1}) \\ &\vee (\neg\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = 0),\end{aligned}$$

e, juntando com (6), concluímos a primeira implicação necessária para (5):

$$\begin{aligned}\text{PA} &\vdash y = \overline{1} \rightarrow [(\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = \overline{1}) \vee (\varphi(\overline{n_1}, ..., \overline{n_k}) \wedge y = 0)] \\ &\vdash y = \overline{1} \rightarrow \psi(\overline{n_1}, ..., \overline{n_k}, y).\end{aligned}$$

Mostremos, então, que a implicação no sentido contrário também pode ser provada em PA. Inicialmente, veja que

$$p \rightarrow [(p \wedge r) \vee (\neg p \wedge q) \rightarrow r]$$

é uma tautologia. Portanto, substituindo p, q, r respectivamente pelas fórmulas $\varphi(\overline{n_1}, ..., \overline{n_k})$, $y = 0$ e $y = \overline{1}$, devemos ter

$$\text{PA} \vdash \varphi(\overline{n_1}, ..., \overline{n_k}) \rightarrow [\psi(\overline{n_1}, ..., \overline{n_k}, y) \rightarrow y = \overline{1}].$$

Novamente via *modus ponens* e o que obtivemos em (4), chegamos a

$$\text{PA} \vdash \psi(\overline{n_1}, ..., \overline{n_k}, y) \rightarrow y = \overline{1}.$$

Dessa forma, finalmente terminamos a prova de (5). O outro caso, decorrente da hipótese que $K_R(n_1, ..., n_k) = 0$, pode ser provado de maneira totalmente análoga, de modo que tenhamos $\text{PA} \vdash \psi(\overline{n_1}, ..., \overline{n_k}, y) \leftrightarrow y = 0$. Com isso, concluímos que se uma relação é representável, então sua função característica também deve ser.

No sentido contrário, suponha que K_R seja representada, em PA, pela fórmula $\theta(x_1, ..., x_k, y)$, e tome $n_1, ..., n_k \in \mathbb{N}$ quaisquer. Como anteriormente, deve haver dois casos: primeiro, se $R(n_1, ..., n_k)$ vale, então $K_R(n_1, ..., n_k) = 1$. Daí,

$$\text{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_k}, y) \leftrightarrow (y = \overline{1}).$$

Usando a substituição $y = \overline{1}$, obtemos $\text{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_k}, \overline{1}) \leftrightarrow (\overline{1} = \overline{1})$, e como é verdade que $\text{PA} \vdash \overline{1} = \overline{1}$, basta usar *modus ponens* para ver que

$$\text{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_k}, \overline{1}),$$

i.e., que $\theta(x_1, ..., x_k, 1)$ representa fortemente $R(x_1, ..., x_k)$.

Agora, assumindo que $\neg R(n_1, ..., n_k)$, temos que $K_R(n_1, ..., n_k) = 0$. Então, usando novamente a substituição $y = \overline{1}$, e pela contrapositiva, temos sucessivamente que

$$\begin{aligned}
&\text{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_k}, y) \leftrightarrow (y = 0);\\
&\text{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_k}, \overline{1}) \leftrightarrow (\overline{1} = 0);\\
&\text{PA} \vdash \neg(\overline{1} = 0) \rightarrow \neg\theta(\overline{n_1}, ..., \overline{n_k}, \overline{1}).
\end{aligned}$$

Pelo teorema 3.2.1, tem-se $\text{PA} \vdash \neg(\overline{1} = 0)$; portanto, via *modus ponens* temos que $\text{PA} \vdash \neg\theta(\overline{n_1}, ..., \overline{n_k}, \overline{1})$. Ou seja, $\neg\theta(x_1, ..., x_k, 1)$ representa fortemente $\neg R(x_1, ..., x_k)$, e temos o resultado. □

Lema 3.2.3. *Dados quaisquer $n \in \mathbb{N}$ e φ uma fórmula escrita na linguagem de* PA*, tem-se que*

$$\text{PA} \vdash \varphi(0) \wedge \varphi(\overline{1}) \wedge ... \wedge \varphi(\overline{n-1}) \wedge x < \overline{n} \rightarrow \varphi(x).$$

Demonstração. Novamente, precisamos fazer indução metateórica. Para $n = 0$, é suficiente que provemos que $\text{PA} \vdash x < 0 \rightarrow \varphi(x)$. Perceba, então, que $p \rightarrow (\neg p \rightarrow q)$ é uma tautologia.[10] Usando-a juntamente da suposição $x < 0$, e notando que o lema 3.1.1(1) nos fornece $\neg(x < 0)$, podemos tomar q como sendo $\varphi(x)$, e isso encerra o caso base. Dessa forma, suponha que o enunciado vale para um certo n. Usando o lema 3.1.1(2), temos que

$$\text{PA} \vdash (x < \overline{n+1}) \leftrightarrow (x < \overline{n} \vee x = \overline{n}). \tag{7}$$

Pelo teorema 3.1.2, obtemos:

$$\text{PA} \vdash (x = \overline{n}) \rightarrow [\varphi(x) \leftrightarrow \varphi(\overline{n})]. \tag{8}$$

Finalmente, usando a hipótese de indução e o que obtivemos em (7) e (8), chegamos ao resultado:

[10] Também chamada de "Lei de Pseudo-Scotus", ou *Princípio da Explosão*. Segundo essa tautologia, uma vez que encontramos alguma contradição dentro de um sistema (i.e., derivamos $p \wedge \neg p$, para alguma fórmula p), então tal sistema consegue provar *qualquer coisa* — e não só suas sentenças válidas, como é de se esperar num sistema correto.

$$\text{PA} \vdash \varphi(0) \wedge \varphi(\overline{1}) \wedge ... \wedge \varphi(\overline{n-1}) \wedge \varphi(\overline{n}) \wedge x < \overline{n+1} \rightarrow \varphi(x).$$

□

Lema 3.2.4. *Dados quaisquer $n \in \mathbb{N}$ e φ uma fórmula escrita na linguagem de* PA*, se para todo $i < n$ tem-se que* $\text{PA} \vdash \neg\varphi(\bar{i})$ *e* $\text{PA} \vdash \varphi(\overline{n})$*, então,*

$$\text{PA} \vdash \varphi(x) \wedge \forall y[y < x \rightarrow \neg\varphi(y)] \leftrightarrow (x = \overline{n}).$$

Demonstração. Denotando $\varphi(x) \wedge \forall y[y < x \rightarrow \neg\varphi(y)]$ por $\psi(x)$, temos, pelo teorema 3.1.2, que

$$\text{PA} \vdash x = \overline{n} \rightarrow \{\psi(x) \leftrightarrow [\varphi(\overline{n}) \wedge \forall y(y < \overline{n} \rightarrow \neg\varphi(y))]\}.$$

Por hipótese, $\text{PA} \vdash \neg\varphi(\bar{i})$ para todo $i < n$. Então, usando o lema 3.2.3 e a regra de generalização, obtemos:

$$\text{PA} \vdash \forall y[y < \overline{n} \rightarrow \neg\varphi(y)].$$

Pela outra hipótese, de que $\text{PA} \vdash \varphi(\overline{n})$, e notando que

$$[p \rightarrow (q \leftrightarrow r)] \wedge r \rightarrow (p \rightarrow q)$$

é uma tautologia, ao substituirmos p, q, r, respectivamente, por $x = \overline{n}$, $\psi(x)$ e $\psi(\overline{n})$, chegamos a

$$\text{PA} \vdash x = \overline{n} \rightarrow \psi(x). \tag{9}$$

Com isso, mostramos que em PA somos capazes de provar a "volta" da equivalência enunciada por este lema. Para a "ida", veja que, usando uma fórmula válida apropriada[11],

$$\text{PA} \vdash \forall y[y < x \rightarrow \neg\varphi(y)] \rightarrow [\overline{n} < x \rightarrow \neg\varphi(\overline{n})].$$

Assim, usando a hipótese de que $\text{PA} \vdash \varphi(\overline{n})$, obtemos (por contraposição):

$$\text{PA} \vdash \psi(x) \rightarrow \neg(\overline{n} < x). \tag{10}$$

Perceba, no entanto, que o lema 3.2.3 nos garante que $\text{PA} \vdash x < \overline{n} \rightarrow \neg\varphi(x)$, o que, por sua vez (e novamente pela contrapositiva), nos dá

$$\text{PA} \vdash \psi(x) \rightarrow \neg(x < \overline{n}). \tag{11}$$

Finalmente, a *lei da tricotomia*, enunciada no lema 3.1.1(3), juntamente de (10) e (11), nos fornecem:

$$\text{PA} \vdash \psi(x) \rightarrow x = \overline{n}.$$

Disso e de (7), segue o resultado. □

[11] Aqui, nos referimos a *Dictum de omni*, um princípio lógico que remonta à antiguidade clássica. Uma de suas mais famosas versões é: "Sócrates é homem; todo homem é mortal; logo, Sócrates é mortal."

Neste momento, temos toda a bagagem necessária para provar o

Teorema 3.2.5 (Teorema da Representabilidade)**.** *Toda função recursiva é representável na linguagem de* PA*.*

Demonstração. O que precisamos fazer é mostrar que (i) todas as funções básicas presentes na definição de $\mathcal{R}$ são representáveis (o que é justamente o enunciado do lema 3.2.1); e (ii) os métodos para obtenção de novas funções devem *preservar* tal representabilidade. Tome, então, g uma função k-ária e $h_1, ..., h_k$ funções j-árias tais que suas representações, na linguagem de PA, sejam dadas pelas fórmulas:

$$\varphi(y_1, ...y_k, z), \psi_1(x_1, ..., x_j, y_1), ..., \psi_k(x_1, ..., x_j, y_k),$$

respectivamente. Fazendo $f = \mathrm{Cn}[g, h_1, ..., h_k]$, nós *afirmamos* que a seguinte fórmula, $\theta(x_1, ..., x_j, z)$, representa a função f:

$$\exists y_1 ... \exists y_k [\psi_1(x_1, ..., x_j, y_1) \wedge ... \wedge \psi_k(x_1, ..., x_j, y_k) \wedge \varphi(y_1, ..., y_k, z)].$$

De fato: seja $f(n_1, ..., n_j) = m$, e para cada $i \in \{1, ..., k\}$, faça $h_i(n_1, ..., n_j) = t_i$. Então,

$$g(t_1, ..., t_k) = m,$$

de modo que tomando $1 \leq i \leq k$, ficamos com:

$$\mathrm{PA} \vdash \psi_i(\overline{n_1}, ..., \overline{n_j}, y_i) \leftrightarrow (y_i = \overline{t_i});$$
$$\mathrm{PA} \vdash \varphi(\overline{t_1}, ..., \overline{t_k}, z) \leftrightarrow (z = \overline{m}). \tag{12}$$

Pelo teorema 3.1.1, temos que

$$\mathrm{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_j}, z) \leftrightarrow \exists y_1 ... \exists y_k [y_1 = \overline{t_1} \wedge ... \wedge y_k = \overline{t_k} \wedge \varphi(y_1, ..., y_k, z)].$$

Observe, no entanto, que

$$\mathrm{PA} \vdash \exists y_i [y_i = \overline{t_i} \wedge \varphi(y_1, ..., y_k, z)] \leftrightarrow \varphi(y_1, ..., y_{i-1}, \overline{t_i}, y_{i+1}, ..., y_k, z),$$

o que deve nos dar

$$\mathrm{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_j}, z) \leftrightarrow \varphi(\overline{t_1}, ..., \overline{t_k}, z).$$

Dessa forma, usando (12), chegamos ao resultado:

$$\mathrm{PA} \vdash \theta(\overline{n_1}, ..., \overline{n_j}, z) \leftrightarrow (z = \overline{m}).$$

Para o processo de minimização, tome g uma função $(k+1)$-ária tal que exista x que verifique:

$$g(n_1, ..., n_k, x) = 0,$$

e suponha que g seja representada por $\psi(x_1, ..., x_k, y, z)$ em PA. Assumindo que $f = \mathrm{Mn}[g](x_1, ..., x_k)$, nós *afirmamos* que a seguinte fórmula, $\varphi(x_1, ..., x_k, y)$, representa f:

$$\psi(x_1, ..., x_k, y, 0) \wedge \forall x_{k+1} [x_{k+1} < y \rightarrow \neg \psi(x_1, ..., x_k, x_{k+1}, 0)].$$

Com efeito: faça $f(n_1, ..., n_k) = m$ e $g(n_1, ..., n_k, i) = t_i$, para cada $1 \leq i \leq k$. Assim, $\text{PA} \vdash \psi(\overline{n_1}, ..., \overline{n_k}, \overline{i}, z) \leftrightarrow (z = \overline{t_i})$, e substituindo z pelo numeral 0, obtemos:

$$\text{PA} \vdash \psi(\overline{n_1}, ..., \overline{n_k}, \overline{i}, 0) \leftrightarrow (0 = \overline{t_i}).$$

Perceba, no entanto, que $t_i \neq 0$ se $i < m$. Assim, caso seja $i < m$, teremos que

$$\text{PA} \vdash \neg\psi(\overline{n_1}, ..., \overline{n_k}, \overline{i}, 0),$$

e como $t_m = 0$, chegamos à conclusão que $\text{PA} \vdash \psi(\overline{n_1}, ..., \overline{n_k}, \overline{m}, 0)$. Usando o lema 3.2.4, alcançamos o resultado:

$$\text{PA} \vdash \varphi(\overline{n_1}, ..., \overline{n_k}, y) \leftrightarrow (y = \overline{m}).$$

Finalmente, temos o caso da recursão primitiva. De maneira ilustrativa, consideremos, inicialmente, apenas a situação em que g e h são uma função unária e ternária (respectivamente), e façamos $f = \text{Rp}[g, h]$. Pelo caráter indutivo de tal processo, será útil pensar nos valores de f como sendo as *entradas* de uma sequência σ. Assim, fazendo $c = f(m, n)$, devemos ter que

$$\begin{aligned}
&\text{entr}(\langle\sigma\rangle, 0) = f(0, n);\\
&\forall i < m[\text{entr}(\langle\sigma\rangle, i) = f(i, n) \rightarrow \text{entr}(\langle\sigma\rangle, \text{s}(i)) = f(\text{s}(i), n)];\\
&\text{entr}(\langle\sigma\rangle, m) = c.
\end{aligned}$$

Ou seja: pela definição de f via recursão primitiva, a entrada 0-ésima de σ é $g(n)$; e sendo $f(i, n)$ a i-ésima entrada, então $f(\text{s}(i), n) = h(i, f(i, n), n)$ é a $(i + 1)$-ésima entrada de σ.

Suponha, então, que g e h sejam representadas, respectivamente, pelas fórmulas $\varphi(x_1, x_2)$ e $\psi(y_1, y_2, y_3, y_4)$, e faça $\langle\sigma\rangle = s \in \mathbb{N}$. Uma vez que (é possível demonstrar que) a função entrada é representável na linguagem de PA[12] — digamos, por $\tau(i, s, z)$, significando que z é a i-ésima entrada da sequência codificada por s —, e considerando a discussão acima, temos que f é representada, em PA, pela fórmula $\theta(z_1, z_2, z_3) = \exists s(\phi)$, em que ϕ é a conjunção das seguintes fórmulas:

$$\begin{aligned}
&\exists u[\tau(0, s, u) \wedge \varphi(x, u)];\\
&\forall w < z_2 \exists u \exists v[\tau(w, s, u) \wedge \tau(\text{s}(w), s, v) \wedge \psi(z_1, w, u, v)];\\
&\tau(y, s, z_3).
\end{aligned}$$

Agora, a tarefa de *provar* que esta fórmula representa, de fato, o processo de recursão primitiva — no sentido dado pela definição 3.2.2 —, além de ser um desafio extremamente árduo[13], pode ser facilmente contornada, se fizermos algumas suposições

[12] Como entr = Cn[log, $\langle\sigma\rangle$, π], e como tais funções podem ser representadas em PA (outra demonstração que preferimos omitir), então entr também deve ser representável.

[13] "Um desafio extremamente árduo" dentro das nossas limitações, i.e., se usarmos apenas as ferramentas que possuímos. Uma prova direta para esse fato pode ser encontrada em (HÁJEK; PUDLÁK, 1998), p. 48-49.

adicionais. [Ademais, caso optássemos por provar tal representabilidade, estaríamos tratando apenas de um caso particular, com g unária e h ternária; por conseguinte, mesmo assim não teríamos uma demonstração completa do teorema.]

Aqui, devemos adiantar um resultado que exibiremos somente na seção subsequente — a saber, o *teorema das funções Beta*. Em suma, e para os nossos propósitos atuais, tal teorema nos diz que dada qualquer sequência finita $(a_1, ..., a_n)$ com n termos, sempre existirá uma função binária $\beta(w, i)$, com $0 \leq i \leq n$, tal que $\beta(w, 0) = n$ e $\beta(w, i) = a_i$, para $i > 0$. Neste caso, chamamos w de o *código da sequência* $(a_1, ..., a_n)$. O que queremos provar é que se uma função f de $k + 1$ argumentos for definida via recursão primitiva (a partir de, digamos, funções g k-ária e h $(k + 2)$-ária), então ela também pode ser definida através das relações $+, *, <, \mathrm{z}, \mathrm{s}, \mathrm{id}_i^n$ e $=$, usando apenas composição e minimização.

Com efeito: dado $\boldsymbol{x} = (x_1, ..., x_k)$, defina uma função auxiliar f' tal que $f'(\boldsymbol{x}, y) = w$ se e somente se w é o menor número natural que codifica uma sequência que satisfaz as seguintes condições:

$$\beta(w, 1) = g(\boldsymbol{x}), \text{ e}$$
$$\beta(w, \mathrm{s}(i)) = h(\boldsymbol{x}, \beta(w, i), i), \text{ para } i > 1.$$

[Ou seja, temos que f' retorna a sequência $(f(\boldsymbol{x}, 0), f(\boldsymbol{x}, 1), ..., f(\boldsymbol{x}, y))$. Perceba a semelhança dessa construção com a que demos anteriormente, quando identificamos o processo de recursão primitiva com uma sequência de aplicações de g e h. Tal semelhança, obviamente, não é mera coincidência: as funções Beta são responsáveis por *generalizar* a codificação de sequências numéricas finitas. Justamente essa generalização é que torna nossa atual tarefa tão mais fácil do que provar a representabilidade da fórmula θ anteriormente especificada.] Dessa maneira, f' pode ser escrita como:

$$f'(\boldsymbol{x}, y) = \mu w[\beta(w, 0) = k + 1 \wedge \beta(w, 1) = g(\boldsymbol{x}) \wedge$$
$$\wedge \forall i < y(\beta(w, \mathrm{s}(i)) = h(\boldsymbol{x}, \beta(w, i), i))].$$

Isso mostra que nenhuma aplicação de recursão primitiva é necessária para a obtenção de f'. Daí, por composição, obtemos:

$$f(\boldsymbol{x}, y) = \beta(f'(\boldsymbol{x}, y), y) = \mathrm{Cn}[\beta, f'],$$

e concluímos que a recursão primitiva pode ser definida somente por minimização e composição. Como já provamos que ambas preservam representabilidade, a prova está completa. □

Aplicando o lema 3.2.2 neste teorema, obtemos diretamente o seguinte resultado:

Corolário 3.2.1. *Toda relação recursiva é fortemente representável na linguagem de* PA*.*

Em suma, fica provado que o sistema aritmético PA é capaz de representar todas as relações *numéricas* efetivamente decidíveis. Nossa próxima tarefa, portanto, será determinar um método pelo qual *expressões metateóricas* sejam identificadas com *números naturais específicos*; usando o que descobrimos até aqui, finalmente teremos como saber o valor lógico de sentenças *sobre* PA.

3.3 ARITMETIZAÇÃO DA SINTAXE

A demonstração original de Gödel para seus teoremas de incompletude possui diversos argumentos brilhantes — alguns dos quais ainda teremos a oportunidade de discutir —, mas, olhando pelo lado técnico, nenhum se compara à ideia (muito simples, inclusive) de se fazer uma *correspondência entre símbolos e números*. Quando pensamos nos símbolos de um sistema formal — e, por conseguinte, em suas expressões bem formadas — como sendo números deste sistema, a noção de recursividade nas relações entre os últimos pode ser facilmente transferida para uma ideia análoga entre os primeiros. Ou seja, por meio desse artifício, conjuntos de expressões formais deverão ser efetivamente decidíveis se e somente se os conjuntos de seus respectivos números naturais o for. Como estamos descrevendo um sistema aritmético, chamaremos tal processo de *aritmetização*.

Há várias formas de se estabelecer tal correspondência — e geralmente, a escolha de uma em detrimento de outra é apenas uma questão de gosto, ou de conveniência com os propósitos da teoria desenvolvida. Todavia, sendo este um trabalho de cunho introdutório, achamos pertinente apresentar duas delas: a primeira será mais direta e de fácil compreensão — mas cujas demonstrações dos teoremas de que precisamos ficariam especialmente difíceis, caso optássemos por tal caminho. Na segunda abordagem, que presumirá alguns resultados mais complexos de teoria da recursão, será muito mais fácil nos convencermos, em particular, da recursividade de funções que exprimem sentenças metateóricas (que é o objetivo principal desta seção).

Antes de distinguir ambas as abordagens, será interessante mudar um pouco nossa caracterização de termos e fórmulas, presente nas definições 3.1.1 e 3.1.2. Designando símbolos quaisquer por ε (e suas variações subgrafadas), combinemos que toda fórmula ou termo pode ser escrita como $\varepsilon\varepsilon_1...\varepsilon_n$, em que ε é o símbolo de um conectivo, função, predicado, ou quantificador; para os ε_i não há restrição. Além de essa forma representar o caráter indutivo com que as fórmulas e termos são obtidos — o que, também, torna a aritmetização mais intuitiva —, com isso conseguimos eliminar a necessidade dos parênteses.[14]

Nesse sentido, deixemos nossa linguagem ainda mais enxuta, considerando como símbolos lógicos somente $\neg$, $\vee$ e $\exists$, uma vez que todos os outros podem ser defi-

[14] Assim, uma fórmula como (0 = 0) deve ser entendida, formalmente, como = 00.

nidos a partir destes. Em contrapartida, será útil adicionar o símbolo $\prime$ para determinar os subscritos das variáveis individuais, de modo que tenhamos $x_1 = x\prime, x_2 = x\prime\prime$,... e assim por diante. Dessa forma, dado um símbolo ε qualquer, denotaremos por SN(ε) seu número natural correspondente. A seguir, listamo-los:

$$\begin{array}{ll} \text{SN}(x) = 0 & \text{SN}(*) = 5 \\ \text{SN}(0) = 1 & \text{SN}(\neg) = 6 \\ \text{SN}(\prime) = 2 & \text{SN}(\vee) = 7 \\ \text{SN}(\text{s}) = 3 & \text{SN}(\exists) = 8 \\ \text{SN}(+) = 4 & \text{SN}(=) = 9 \end{array}$$

Em ambos os casos, usaremos a codificação acima.

3.3.1 Aritmetização por concatenação

Neste ponto, identificaremos cada número natural com a sequência que contém somente tal número (e.g., 4 = (4) e 378 = (378)). Assim, poderemos estender a noção de concatenação de sequências, introduzida ao final do capítulo anterior, para *concatenação de números*, usando o mesmo símbolo. Faremos isso dentro do sistema decimal de representação numérica, de modo que, e.g., tenhamos

$$2 \star 3 = 23 = 20 + 3 = 2 * 10^1 + 3 * 10^0.$$

Mais geralmente, tem-se que $m \star n = m * 10^{\ell(n)} + n$, em que $\ell(n)$ é o comprimento de n, i.e., o menor $k \in \mathbb{N}$ tal que $10^k > n$ (que é, claramente, uma função recursiva). Com isso, temos a primeira versão da ideia de *gödelização* de expressões formais:

Definição 3.3.1 (Números de Gödel, Versão I)**.** Dada $\epsilon = \varepsilon\varepsilon_1...\varepsilon_n$ uma expressão bem formada na linguagem de PA, chamaremos $\ulcorner\epsilon\urcorner$ de *o número de Gödel de* ϵ, e o calcularemos da seguinte maneira:

$$\ulcorner\epsilon\urcorner = \text{SN}(\varepsilon) \star \ulcorner\varepsilon_1\urcorner \star ... \star \ulcorner\varepsilon_n\urcorner,$$

em que os $\ulcorner\varepsilon_i\urcorner$ são números de Gödel dos subtermos e subfórmulas que compõem ϵ, calculados previamente.

Para fixar esse conceito, vamos calcular o número de Gödel da fórmula $\varphi(x)$, dada por

$$\neg\exists y[\text{s}(x) + y = 0 \vee x = y].$$

Como dissemos, é possível reescrever φ sem o uso de parênteses, colocando os símbolos de predicados, funções, conectivos e quantificadores mais à esquerda de cada subfórmula. A fim de melhorar a leitura, adicionamos alguns espaços vazios (que obviamente não contam como símbolos):

$$\neg \quad \exists y \quad \vee \quad = + \mathsf{s} x \, y \, 0 \quad = x \, y .$$

Assim, considerando $x = x_1$ e $y = x_2$ — de modo que $\mathrm{SN}(x) = \mathrm{SN}(x\prime)$ e $\mathrm{SN}(y) = \mathrm{SN}(x\prime\prime)$ —, temos sucessivamente que

$$\begin{aligned}
&\ulcorner x = y \urcorner = \ulcorner = xy \urcorner = 9 \star 0 \star 2 \star 0 \star 2 \star 2 = 902\,022 = a,\\
&\ulcorner \mathsf{s} x \urcorner = 3 \star 0 \star 2 = 302 = b,\\
&\ulcorner \mathsf{s} x + y \urcorner = 4 \star b \star 0 \star 2 \star 2 = 4\,302\,022 = c,\\
&\ulcorner \mathsf{s} x + y = 0 \urcorner = 9 \star c = 94\,302\,022 = d,\\
&\ulcorner \mathsf{s} x + y = 0 \vee x = y \urcorner = 7 \star d \star a = 794\,303\,022\,902\,022 = e,\\
&\ulcorner \varphi(x) \urcorner = 6 \star (8 \star 0 \star 2 \star 2 \star e) = 68\,022\,794\,302\,022\,902\,022.
\end{aligned}$$

É importante notar que tal processo *não é* uma bijeção, pois há (infinitos) números naturais que *não são* números de Gödel de expressão alguma na linguagem de PA (um para cada expressão que não seja bem formada, nesse caso). Agora, supondo que $a \in \mathbb{N}$ seja, de fato, um número de Gödel, então é possível percorrer o "caminho inverso", i.e., descobrir qual fórmula φ em PA é tal que $\ulcorner \varphi(x) \urcorner = a$. Com efeito: basta substituir cada algarismo de a pelo símbolo correspondente. Assim, concluímos que a aritmetização de expressões formais é capaz de exprimir, num sentido semelhante à *recursividade de relações entre números naturais*, a ideia de *processo efetivo de decisão entre as componentes formais do sistema axiomático em questão*.

A consequência mais relevante de se usar tal artifício, no entanto, é que podemos não somente designar números a todas as *fórmulas e termos* do sistema aritmético PA, mas que podemos fazer o mesmo para *sentenças sobre* PA. Dito de outra maneira, podemos aritmetizar a metamatemática em si, definindo funções (recursivas) na linguagem de PA que correspondam a ideias sobre o próprio sistema; a seguir, veremos que há funções capazes de dizer se algo é uma variável, uma fórmula, etc. Por conseguinte, teremos um método efetivo para decidir quanto à validade de tais sentenças metateóricas. Por motivos técnicos, apenas mostraremos como fazer essas correspondências no tópico a seguir.

3.3.2 Aritmetização por sequências

Enquanto o processo de aritmetização acima se baseia na simples justaposição de algarismos (que resulta num número natural em sua representação decimal), nossa próxima caracterização interpreta a mesma listagem de algarismos como uma sequência numérica, e depois a codifica.[15] Visando praticidade, no entanto, usaremos

[15] Ou seja, aqui trabalharemos de uma maneira praticamente inversa: lá, cada número era identificado com a sequência unária que o contém, de modo que a concatenação terminava o trabalho diretamente; nesta subseção, no entanto, teremos, e.g., que a aritmetização de $0 = 0$ é o *código da sequência* $s_3 = (9, 1, 1)$.

uma definição diferente da codificação $\langle\ \rangle$ de sequências apresentada na seção 2.2. (Basicamente, o problema prático de se codificar sequências numéricas através do teorema fundamental da aritmética, como fizemos no capítulo passado, surge na especificação de relações recursivas metalinguísticas — algo que trataremos logo a seguir —, onde precisamos *decodificar* os números que representam as fórmulas e termos da linguagem de PA.) Para isso, precisamos entender o que vem a ser uma *função Beta*:

Definição 3.3.2. Dizemos que $\beta : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ é uma *função Beta* se para toda sequência numérica $(a_0, a_1, ..., a_{n-1})$, tem-se que existe $w \in \mathbb{N}$ tal que

$$\begin{aligned}
&\beta(w, 0) = a_0,\\
&\beta(w, 1) = a_1,\\
&\vdots\\
&\beta(w, n-1) = a_{n-1}.
\end{aligned}$$

Perceba que, uma vez que sequências finitas são, por definição, funções entre números naturais, as funções Beta podem ser vistas como as versões binárias de suas respectivas sequências, em que a primeira entrada do argumento é fixa. Dessa maneira, podemos pensar nas constantes w como sendo as etiquetas, ou os códigos, que as marcam. Isto posto, enunciamos agora (sem demonstração)[16] o teorema que garante a existência de funções Beta:

Teorema 3.3.1. *É possível construir uma função Beta recursiva (e, portanto, representável em* PA*).*

Perceba que tal teorema garante a existência de *ao menos* uma função Beta, i.e., ao menos uma forma de se codificar qualquer sequência numérica finita (lembre-se: uma função Beta dá códigos a *todas* as sequências finitas). No mesmo sentido, da definição 3.3.2 vem que, também, existe *ao menos* um código para cada sequência; no entanto, sendo uma definição, não é possível dizer nem quais números, nem quantos deles, satisfazem a condição. Isto posto, para definirmos o código de $(a_1, ..., a_n)$, que possui n termos, fazemos o seguinte: primeiramente, consideramos a sequência $(a_0, a_1, ..., a_n)$, com $n+1$ termos, em que $a_0 = n$ (e não há o que nos impeça de fazê-lo, já que o teorema 3.3.1 garante a existência de Beta para qualquer sequência); com isso, teremos uma β que descreve *esta última sequência com* $n+1$ *termos*. A seguir, fazemos de w o menor número tal que $\beta(w, 0) = n$ e $\beta(w, i) = a_i$ para $0 < i \leq n$, e escrevemos:

$$w = \langle (a_1, ..., a_n) \rangle.$$

[16] O leitor pode encontrar duas ótimas provas desse teorema, inclusive bastante distintas uma da outra, em (BOOLOS, G.; JEFFREY, 1980), p. 162-163, e (SMULLYAN, 1992), p. 45-46.

Desse modo, temos uma nova definição para a codificação de sequências numéricas, diferente da apresentada no final do capítulo passado. Naturalmente, também podemos redefinir, relativamente à mesma sequência, as funções *comprimento* e *entrada*, colocando, para $0 < i \leq n$,

$$\text{lh}(w) = \beta(w, 0);$$
$$\text{entr}(w, i) = \beta(w, i).$$

As demais funções que dizem respeito a sequências numéricas, como a *truncagem* e a *concatenação*, podem ser redefinidas da mesma maneira.[17] Convém, no entanto, especificar o conjunto (ou melhor, a relação unária) de todas as sequências numéricas — i.e., de todos os *códigos* das mesmas:

$$\text{Seq}(w) \leftrightarrow \forall x < w[\text{lh}(x) \neq \text{lh}(w) \vee \exists i < \text{lh}(w)(\text{entr}(x, i) \neq \text{entr}(w, i))].$$

[Aqui, as desigualdades se justificam pois, denotando a sequência vazia — i.e., quando $n = 0$ — por s_0, é verdade que

$$w \neq \langle s_0 \rangle \rightarrow \text{lh}(w) < w \wedge \text{entr}(w, i) < w.]$$

Em posse dessas novas definições, podemos voltar ao processo de aritmetização. Primeiramente, podemos dar a segunda versão de números de Gödel:

Definição 3.3.3 (Números de Gödel, Versão II)**.** Dada $\epsilon = \varepsilon\varepsilon_1...\varepsilon_n$ uma expressão bem formada na linguagem de PA, chamaremos $\ulcorner\epsilon\urcorner$ de *o número de Gödel de* ϵ, e o calcularemos da seguinte maneira:

$$\ulcorner\epsilon\urcorner = \langle(\text{SN}(\varepsilon), \ulcorner\varepsilon_1\urcorner, ..., \ulcorner\varepsilon_n\urcorner)\rangle,$$

em que os $\ulcorner\varepsilon_i\urcorner$ são números de Gödel dos subtermos e subfórmulas que compõem ϵ, calculados previamente.

Quanto às expressões puramente aritméticas de PA, vê-se, claramente, como suas codificações via funções Beta devem ser mais simples do que as apresentadas na subseção anterior. Por exemplo, para a fórmula $x = y$, que lemos $= xy$ (ou, ainda, $=x\prime x\prime\prime$), temos que seu código será, simplesmente,

$$w = \langle(9, 0, 2, 0, 2, 2)\rangle,$$

[17] Finalmente, perceba que a codificação de sequências por funções Beta é, simplesmente, uma generalização para tal processo. Assim, a definição anterior de $\langle\ \rangle$, na qual usamos o Teorema Fundamental da Aritmética, é *apenas uma* das funções Beta que descrevem determinada sequência. É justamente neste sentido que o teorema 3.3.1 tanto nos ajuda: como ele nos dá, diretamente, um código w, não precisamos fazer contas!

i.e., o menor w tal que $\beta(w,0) = 6$, $\beta(w,1) = 9$, $\beta(w,2) = 0$, etc. Nosso objetivo, agora, é mostrar como as expressões metamatemáticas são, também, passíveis de codificação por sequências, ou seja, que elas têm seus respectivos números de Gödel. Para isso, basta mostrá-las recursivas, do que seguirá que são representáveis (e, portanto, "gödelizáveis").

Comecemos definindo que "algo, na linguagem de PA, é uma variável" — i.e., mostremos que há uma função que nos diz se w é o número de Gödel de uma variável. Podemos fazê-lo do seguinte modo:

$$\text{Var}(w) \leftrightarrow \ \text{tr}(w,1) = \langle(0)\rangle \wedge \forall y \leq n[y > 1 \rightarrow \text{entr}(w,y) = 2].$$

Ou seja, w será o código de uma variável se e somente se for o código de uma sequência $(a_1, ..., a_n)$ cujo primeiro termo é 0 (i.e., o código de x), e todos os termos seguintes são SN($\prime$). Dessa forma, $w = \ulcorner x_j \urcorner$, para algum j.

Combinando que Term(w) significa que w codifica uma sequência que representa um termo, temos:

$$\begin{aligned}
\text{Term}(w) &\leftrightarrow 0 = 0, \text{ se } w = \langle(\text{SN}(0))\rangle; \\
&\leftrightarrow \text{Term}(\text{entr}(w,2)), \text{ se } w = \langle(\text{SN(s)}, \text{entr}(w,2)\rangle; \\
&\leftrightarrow \text{Term}(\text{entr}(w,2)) \wedge \text{Term}(\text{entr}(w,3)), \\
&\quad \text{se } w = \langle(\text{SN}(+), \text{entr}(w,2), \text{entr}(w,3))\rangle \\
&\quad \vee w = \langle(\text{SN}(*), \text{entr}(w,2), \text{entr}(w,3))\rangle; \\
&\leftrightarrow \text{Var}(w).
\end{aligned}$$

[Traduzindo: w será o código de um termo quando for (*i*) o código de uma sequência cujo único termo é 0; ou (*ii*) o código de uma sequência cuja primeira entrada é $\ulcorner \text{s} \urcorner$, e a próxima é o código de um termo; ou (*iii*) o código de uma sequência cuja primeira entrada é ou $\ulcorner + \urcorner$ ou $\ulcorner * \urcorner$, sendo que as segunda e terceira entradas são códigos de termos; ou, finalmente, quando (*iv*) w codifica uma variável.] Por ter uma definição por casos, em que cada caso apresenta somente funções recursivas, essa função é recursiva.

De maneira similar, podemos determinar muitas outras relações que exprimem ideias metateóricas — e.g., teremos que Form(w) vale se e somente se w é o número de Gödel de uma fórmula escrita na linguagem de PA. Achamos desnecessário, no entanto, exibir *todas* as definições. Assim, elencamos abaixo apenas a mais pertinentes ao nosso trabalho — e especificamos as definições de duas delas em seguida. Note que, aqui, o símbolo $\Leftrightarrow$ ocorre numa linguagem híbrida do português com o sistema formal em que estamos trabalhando (e significa "se e somente se"):

- Var(a) $\Leftrightarrow$ a é o código (i.e., o número de Gödel) de uma variável da linguagem de PA;

- Term(a) $\Leftrightarrow$ a é o código de um termo (variável ou constante) da linguagem de PA;
- Form(a) $\Leftrightarrow$ a é o código de uma fórmula da linguagem de PA;
- Sub(a, b, c) = $m \Leftrightarrow$ m é o código do resultado de substituir a variável de código b pelo termo de código c na fórmula ou termo de código a;
- Liv(a, b) $\Leftrightarrow$ b é (o código de) uma variável livre na fórmula (de código) a;
- Sbstvl(a, b, c) $\Leftrightarrow$ a variável b é substituível pelo termo c na fórmula a;
- AxL(a) $\Leftrightarrow$ a é o código de um axioma lógico;
- AxI(a) $\Leftrightarrow$ a é o código de um axioma de identidade;
- AxP(a) $\Leftrightarrow$ a é o código de um axioma não-lógico;
- Ax(a) $\Leftrightarrow$ a é o código de um axioma;
- Sb(a, b) $\Leftrightarrow$ ambos a e b codificam fórmulas, e b é o resultado de se substituir uma variável em a por um termo qualquer;
- MP(a, b, c) $\Leftrightarrow$ a fórmula de código c pode ser obtida das fórmulas de código a e b via *modus ponens*;
- IUni(a, b) $\Leftrightarrow$ b é o resultado de se fazer a *introdução do quantificador universal* em a;
- IExi(a, b) $\Leftrightarrow$ b é o resultado de se fazer a *introdução do quantificador existencial* em a;
- EUni(a, b) $\Leftrightarrow$ b é o resultado de se fazer a *eliminação do quantificador universal* em a;
- EExi(a, b) $\Leftrightarrow$ b é o resultado de se fazer a *eliminação do quantificador existencial* em a;
- Gen(a, b) $\Leftrightarrow$ b codifica a fórmula obtida pela generalização da fórmula de código a.

[Como prometido, vamos definir formalmente duas dessas relações. Primeiramente, veja que a partir de Var podemos definir uma função que diz se algo é uma *fórmula atômica* — i.e., uma fórmula sem operadores lógicos —; neste caso, pela definição 3.1.2.(i), teremos uma fórmula atômica quando tivermos uma sequência de símbolos

em que o primeiro é o sinal de "igual a" e os outros dois símbolos representam *termos* de PA. Assim, devemos ter:

$$\begin{aligned}\mathrm{AtomForm}(a) \leftrightarrow\ & a = \langle \mathrm{entr}(a,1), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle \wedge \\ & \wedge\, \mathrm{entr}(a,1) = \mathrm{SN}(=) \wedge \mathrm{Term}(\mathrm{entr}(a,2)) \wedge \mathrm{Term}(\mathrm{entr}(a,3)).\end{aligned}$$

Seguindo a definição 3.1.2, podemos ver que a seguinte relação descreve todas as fórmulas de PA:

$$\begin{aligned}\mathrm{Form}(a) &\leftrightarrow \mathrm{Form}(\mathrm{entr}(a,2)),\ \text{se } a = \langle \mathrm{SN}(\neg), \mathrm{entr}(a,2)\rangle, \\ &\leftrightarrow \mathrm{Form}(\mathrm{entr}(a,2)) \wedge \mathrm{Form}(\mathrm{entr}(a,3)), \\ &\quad \text{se } a = \langle \mathrm{SN}(\vee), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle \vee \\ &\quad\ \vee\, a = \langle \mathrm{SN}(\wedge), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle \vee \\ &\quad\ \vee\, a = \langle \mathrm{SN}(\rightarrow), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle \vee \\ &\quad\ \vee\, a = \langle \mathrm{SN}(\leftrightarrow), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle, \\ &\leftrightarrow \mathrm{Var}(\mathrm{entr}(a,2)) \wedge \mathrm{Form}(\mathrm{entr}(a,3)), \\ &\quad \text{se } a = \langle \mathrm{SN}(\exists), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle \vee \\ &\quad\ \vee\, a = \langle \mathrm{SN}(\forall), \mathrm{entr}(a,2), \mathrm{entr}(a,3)\rangle, \\ &\leftrightarrow \mathrm{AtomForm}(a), \text{caso contrário.}\end{aligned}$$

(Perceba que não precisaríamos especificar os casos dos conectivos $\wedge$, $\rightarrow$ e $\leftrightarrow$, nem do quantificador $\forall$, pois estes são definíveis a partir de $\vee$, $\neg$ e $\exists$, como já frisamos no início desta seção. Mas uma definição como acima, embora redundante, nos parece mais didática, pois segue diretamente nossa primeira caracterização de fórmulas, presente na definição 3.1.2.)

Indo um pouco mais além na lista de relações metamatemáticas, abaixo definimos a introdução do quantificador existencial:

$$\begin{aligned}\mathrm{IExi}(a,b) \leftrightarrow\ & \exists x < a \exists y < a \exists z < a [\mathrm{Form}(x) \wedge \mathrm{Form}(y) \wedge \\ & \wedge\, \mathrm{Var}(z) \wedge a = \langle (7, 6, x, y)\rangle \wedge \\ & \wedge\, b = \langle (7, 6, \langle (8, z, x)\rangle, y)\rangle \wedge \neg \mathrm{Liv}(y, z)],\end{aligned}$$

lembrando que $6 = \mathrm{SN}(\neg)$ e $7 = \mathrm{SN}(\vee)$, e que a implicação $p \rightarrow q$ pode ser definida como $\neg p \vee q$ (que escrevemos $\vee \neg pq$). Como $8 = \mathrm{SN}(\exists)$, temos que a relação $\mathrm{IExi}(a,b)$ se dá quando existem três números $x, y, z < a$ tais que x, y codificam fórmulas φ_x e φ_y, z codifica uma variável — digamos, z_1 —, a codifica a implicação referida acima, b codifica a implicação $\exists z_1 \varphi_x \rightarrow \varphi_y$, e a variável codificada por z não é livre na fórmula codificada por y.]

Todas as relações acima listadas são derivadas sucessivamente uma das outras, partindo da função Var; assim, fica claro que todas elas são recursivas. Na prática, isso

significa que, para todas elas, temos uma maneira efetiva de decidir se determinado código se refere, ou não, ao componente sintático em questão. Além disso, perceba que, embora tal lista tenha um potencial infinito (pois, por meio da aritmetização, podemos "falar sobre" praticamente qualquer característica do sistema PA que pudermos identificar), o conjunto elencado acima engloba todos os indivíduos, axiomas e regras de inferência do sistema aritmético que estamos descrevendo. Neste ponto, portanto, o leitor não deve mais se surpreender com o fato de que *existe* uma relação entre números naturais a qual se dá quando *um deles é o número de Gödel de uma demonstração para a fórmula codificada pelo outro*.

Para defini-la formalmente, basta identificarmos um *prova em* PA com uma sequência finita de fórmulas, em que o primeiro termo deve ser um axioma e o último é a fórmula que buscamos provar — i.e., o *teorema*. Todos os outros termos da sequência são, também, axiomas, ou são obtidos destes pelas regras de inferência. Como cada um desses termos da sequência possui um número de Gödel, essa sequência, em si, pode ser vista como uma sequência numérica finita, do que segue que ela também possui um código; da discussão acima, alguma relação recursiva deve descrevê-la. Chamemos tal relação de Dem, de modo que $\mathrm{Dem}(a, b)$ vale se e somente se a é o número de Gödel de uma demonstração para a fórmula com número de Gödel b. Abreviando $\mathrm{entr}(a, i) = a_i$, fazemos:

$$\begin{aligned}\mathrm{Dem}(a,b) \leftrightarrow\ & \mathrm{Seq}(a) \wedge \mathrm{lh}(a) \neq 0 \wedge \\ & \wedge \forall i < \mathrm{lh}(a)\{\mathrm{Form}(a_i) \wedge \\ & \wedge [\mathrm{Ax}(a_i) \vee \exists j, k < i[\mathrm{Sb}(a_j, a_i)) \vee \\ & \vee \mathrm{MP}(a_j, a_k, a_i) \vee \mathrm{IExi}(a_j, a_i) \vee \\ & \vee \mathrm{EExi}(a_j, a_i) \vee \mathrm{IUni}(a_j, a_i) \vee \\ & \vee \mathrm{EUni}(a_j, a_i) \vee \mathrm{Gen}(a_j, a_i)]]\} \wedge \\ & \wedge \mathrm{entr}(a, lh(a) \ominus 1) = b.\end{aligned}$$

Isto posto, ainda podemos definir mais uma função, que traduz o primeiro conceito epistemológico que introduzimos — a saber, a ideia de *incerteza*. Com efeito: tomando a um número de Gödel qualquer, ao fazermos

$$\mathrm{Bew}(a) \leftrightarrow \exists x \mathrm{Dem}(x, a),$$

estamos aritmetizando a noção de que "a fórmula codificada por a é demonstrável no sistema PA".[18] Em suma, portanto, nossas indagações filosóficas presentes no primeiro capítulo encontraram, finalmente, um aspecto formal. Mais ainda, neste momento somos capazes de *falar* sobre tais indagações metateóricas *usando* a própria teoria (i.e., nossa *dúvida* quanto à demonstrabilidade de uma conjectura se reduz

[18] A escolha de notação vem de *beweisbar*, que significa "demonstrável" em alemão.

a codificar tal conjectura e, posteriormente, buscar por algum x que satisfaça uma equivalência como acima).

Vale notar, todavia, que o quantificador existencial em $\exists x\text{Dem}(x, a)$ é *ilimitado*. Isso ocorre porque, em princípio, não temos como saber o tamanho de uma demonstração (e, além do mais, geralmente há *diversas* maneiras de se demonstrar um teorema). Por conseguinte, Bew não é, necessariamente, uma função recursiva.[19] Como veremos adiante, isso justifica o fato de o conjunto dos números de Gödel de teoremas de PA não ser um conjunto recursivo (i.e., falando em termos de processos mecânicos, não é possível especificar uma máquina para cada conjectura sobre a qual buscamos uma decisão). Com isso em mente, podemos dizer que, enfim, chegamos ao clímax de nossa argumentação: a seguir, mostraremos como construir sentenças indecidíveis em PA, donde concluiremos que tal sistema é incompleto.

3.4 OS TEOREMAS DE INCOMPLETUDE DE GÖDEL

No início da seção precedente, enfatizamos como, ao menos no quesito técnico, a codificação dos elementos sintáticos de PA constitui a parte mais importante do argumento de Gödel. Tamanha relevância ficou ainda mais evidente alguns parágrafos atrás, quando o conceito de *prova* foi, enfim, definido formalmente. Mas, como o leitor deve ter percebido, embora a ideia de se atribuir números a símbolos seja, em essência, algo muito simples, o que vimos posteriormente foi uma torrente de demonstrações carregadas de notação e, senão complexas, bastante extensas (tanto que omitimos a maioria delas).

Em contrapartida, o passo final que precisamos dar é baseado num artifício bastante corriqueiro dentro das argumentações matemáticas — a saber, o processo de diagonalização, usado desde Cantor em seu já mencionado teorema. A rigor, a ideia é realmente muito simples, mas isso não torna o argumento, de maneira nenhuma, menos brilhante. Lembre, por exemplo, o que fizemos na seção 2.1, quando mostramos que determinada função é Turing-incomputável (não por coincidência, demos a ela o nome de *função diagonal*). O passo crucial, tanto lá como aqui, é aplicar, *na função*, o valor correspondente *à própria função*. A diferença é que, neste momento, o tal "valor" de uma função (ou, mais especificamente, de uma fórmula que a represente) tem uma caracterização precisa, dada pelo seu número de Gödel. Mais geralmente, dada uma expressão E na linguagem de PA, dizemos que a *diagonalização de E* é a expressão:

$$\exists x[(x = \ulcorner E \urcorner) \wedge E].$$

[19] A bem da verdade, Bew é uma relação *recursivamente enumerável*, i.e., pertencente à classe um nível acima de $\mathcal{R}$ na hierarquia aritmética. Isso significa, *grosso modo*, que Bew pode ser obtida de uma função recursiva ao se inserir, nesta, *apenas um* quantificador ilimitado. É o que vemos, claramente, em sua definição.

[Como, em geral, uma expressão unária $R(t)$ é equivalente a $\exists x(x = t \wedge R(x))$, vamos dizer que a diagonalização de E é, simplesmente, $E(\ulcorner E \urcorner)$.] Exemplificando, tome $E = E(x) \leftrightarrow x + y = \overline{2}$, e considere $m = \ulcorner E \urcorner$. Então, E representa a "reta discreta" $x = \overline{2} - y$, i.e., o conjunto dos $x \in \mathbb{N}$ que satisfazem tal equação, com y fixo (embora arbitrário). Já sua diagonalização, $E(\ulcorner E \urcorner) \leftrightarrow \overline{m} + y = 2$, representa o conjunto unitário que contém somente o número $y = \overline{2} - \overline{m}$ (ou, caso não exista tal número, E representa $\emptyset$).

Assim, o processo de diagonalização pode ser visto como a ideia de tratar as próprias expressões linguísticas como *elementos sintáticos* de uma nova expressão, cuja "base" (i.e., a fórmula em que inserimos seu número de Gödel) é a própria expressão sendo inserida. No exemplo acima, todavia, não obtivemos nada especial ao aplicar o número de Gödel da expressão a ela mesma (na verdade, apenas fizemos surgir uma nova sentença, derivada da original, que é verdadeira se e somente se for verdadeira quando seu próprio número de Gödel é aplicado em si mesma). O "pulo do gato" em nosso argumento acontece quando aplicamos esse conceito a uma relação definida a partir de $\mathrm{Dem}(x, y)$. Informalmente, a sentença que iremos construir diz, em linguagem aritmética, que *ela, a própria sentença que está sendo enunciada neste momento em itálico, não é demonstrável dentro do sistema*. Simbolicamente, teremos que $E \leftrightarrow E(\ulcorner E \urcorner)$. Em outras palavras, por meio da diagonalização, poderemos construir uma sentença que *faz referência a si mesma*. Por isso, a chamaremos de uma sentença *autorreferente*.[20]

3.4.1 Primeiro Teorema: construindo sentenças indecidíveis

Vejamos, então, passo a passo, como representar tal sentença na linguagem de PA. Primeiramente, definimos, via recursão primitiva, uma função que exprime a ideia

[20] Vale notar, entretanto, que o método de diagonalização, definido da maneira que fizemos, é *apenas uma* forma de se abstrair o conceito de autorreferência, específica para a linguagem aritmética. Em (SMULLYAN, 1994), define-se um método geral, chamado de *citação* (do inglês *quotation*). Nele, a distinção entre os atos de *usar* e *mencionar* uma expressão (i.e., uma palavra, ou qualquer componente sintático de um sistema) tem papel crucial — por exemplo, convencionando que a *menção* a uma expressão γ é a expressão "γ", devemos concordar que é falsa a sentença 'Arara tem cinco letras', enquanto que "'Arara" tem cinco letras' é verdadeira. Neste sentido, o processo de diagonalização de uma expressão pode ser visto como o resultado de, ao mesmo tempo, *usarmos* e *mencionarmos* tal expressão. Para obter autorreferência, inicialmente considere a frase 'Hilbert está lendo x', em que x varia sobre as expressões da língua portuguesa. Sua diagonalização, evidentemente, será

> Hilbert está lendo "Hilbert está lendo x".

Perceba, no entanto, que esta frase não é autorreferente, pois ela não diz que Hilbert está lendo ela mesma. Mas se considerarmos H como sendo a sentença 'Hilbert está lendo a diagonalização de x', então a seguinte sentença é autorreferente:

> Hilbert está lendo a diagonalização de "Hilbert está lendo a diagonalização de x".

De fato: tal sentença afirma que Hilbert está lendo a diagonalização de H, mas a diagonalização de H é, justamente, esta última sentença. O argumento de Gödel usa exatamente essa ideia — embora, evidentemente, o faça dentro de um sistema formal, e não em língua portuguesa.

de algo ser o número de Gödel de um numeral:

$$\begin{aligned}\mathrm{Num}(0) &= \langle(\mathrm{SN}(0))\rangle;\\ \mathrm{Num}(w+1) &= \langle(\mathrm{SN}(\mathrm{s}), \mathrm{Num}(w))\rangle.\end{aligned}$$

Assim, $\mathrm{Num}(k) = \ulcorner\overline{k}\urcorner$, para todo $k \in \mathbb{N}$. Em seguida, consideramos a relação binária:

$$R(a,b) \leftrightarrow \mathrm{Form}(a) \wedge \mathrm{Liv}(a,0) \wedge \mathrm{Dem}(b, \mathrm{Sub}(a,0,\mathrm{Num}(a))).$$

[Traduzindo numa linguagem híbrida, temos que tal relação vale se e somente se a codifica uma fórmula (digamos, ξ) na qual a variável x ocorre livremente, e b codifica uma prova para a fórmula resultante da substituição, em ξ, da variável x pelo código a (i.e., pelo número de Gödel da própria fórmula ξ).]

Evidentemente, R é uma relação recursiva, do que segue que é fortemente representável na linguagem de PA — digamos, pela fórmula $\varphi(x,y)$. Faça, então, $\psi(x,y)$ ser a fórmula $\forall y \neg\varphi(x,y)$, e considere $m = \ulcorner\psi\urcorner$. Substituindo x pelo numeral de m, obtemos a diagonalização de ψ, à qual designaremos o símbolo $\mathfrak{Dd}$:

$$\forall y \neg\varphi(\overline{m}, y) \leftrightarrow \psi(\overline{m}, y).$$

Perceba que $\mathfrak{Dd}$ é válida quando *nenhum* código representado por y é o número de Gödel de uma demonstração para a sentença cujo número de Gödel é representado por $\overline{m}$ — mas, nesse caso, m é o código da própria sentença $\mathfrak{Dd}$! Em outras palavras, $\mathfrak{Dd}$ é uma sentença autorreferente que representa, em PA, a ideia de *ela mesma não ser um teorema de* PA. É essa sentença, $\mathfrak{Dd}$, que provaremos ser indecidível. Daremos a ela o nome de *sentença de Gödel*.

Antes, no entanto, precisamos entender sob quais hipóteses os teoremas de incompletude se verificam. Como dissemos, a justificativa para PA ser incompleto é que seremos capazes de *construir*, efetivamente, uma fórmula que é verdadeira, mas que não pode ser demonstrada — nem refutada — usando as ferramentas disponíveis. Agora, quando construímos uma teoria axiomática como o sistema aritmético em questão, além de estabelecer um conjunto de axiomas potente o suficiente para provar tudo aquilo que desejamos provar, outro objetivo que sempre temos em mente é que tais axiomas não podem, de maneira nenhuma, ser contraditórios. Caso isso ocorresse, nossa teoria poderia provar, literalmente, qualquer coisa — ou seja, todas as sentenças verdadeiras de PA, *bem como suas negações, que são falsas*. Assim, fica claro que o segundo aspecto é o mais imprescindível: antes de nos perguntarmos se uma teoria matemática é *completa*, devemos *evitar os famigerados paradoxos*.

Isto posto, devemos assumir que PA é um sistema consistente. Para provar os teoremas, entretanto, precisamos de uma hipótese mais forte, chamada de ω-consistência:

Definição 3.4.1. O sistema PA é dito ser ω-*consistente* quando para toda fórmula $\varphi(x)$, tem-se que

$$\text{se PA} \vdash \varphi(0), \text{PA} \vdash \varphi(\overline{1}), ..., \text{PA} \vdash \varphi(\overline{n}), ..., \text{então PA} \nvdash \exists x \neg \varphi(x).$$

Ou seja, PA é ω-consistente se e somente se ele prova, individualmente, cada uma das fórmulas $\varphi(\overline{m})$, e *não prova* a existência de um número natural tal que $\neg\varphi$. Por outro lado, uma teoria será ω-*inconsistente* quando ela provar todas as fórmulas $\varphi(\overline{m})$, mas, contraditoriamente, também provar a *negação* daquilo que seria óbvio concluir (i.e., ela prova a negação de $\forall x \varphi(x)$). O próximo resultado justifica o fato de tal ideia ser mais forte do que a de consistência simples:

Teorema 3.4.1. *Se o sistema* PA *é* ω*-consistente, então é consistente.*

Demonstração. Suponha que PA é ω-consistente. Pela discussão que fizemos anteriormente, se PA fosse inconsistente, todas as fórmulas escritas em sua linguagem seriam teoremas. Então, basta provar que ao menos uma fórmula não é um teorema de PA. Com efeito: considere a fórmula $x = x$, que é um axioma lógico de PA. Então, tem-se que PA $\vdash x = x$. Por substituição, segue que para cada $n \in \mathbb{N}$, PA $\vdash \overline{n} = \overline{n}$. Assim, da ω-consistência de PA, concluímos que

$$\text{PA} \nvdash \exists x \neg (x = x).$$

□

Com isso, finalmente podemos provar o tão anunciado

Teorema 3.4.2 (Primeiro Teorema de Incompletude)**.** *Se* PA *é um sistema* ω*-consistente, então é incompleto.*

Demonstração. Tudo que precisamos provar é que a sentença de Gödel não é demonstrável em PA, nem sua negação. Em símbolos, temos que derivar PA $\nvdash \mathfrak{G}$ e PA $\nvdash \neg\mathfrak{G}$. Assim, por *reductio ad absurdum*, suponha que PA $\vdash \mathfrak{G}$, o que significa que $\mathfrak{G}$ possui uma prova em PA. Seja k o número de Gödel de tal prova. Recordando a relação R que deu origem a $\mathfrak{G}$ anteriormente, temos $R(m, k)$, em que m é o número de Gödel de $\mathfrak{G}$. Como R é fortemente representável pela fórmula φ, temos que

$$\text{PA} \vdash \varphi(\overline{m}, \overline{k}).$$

Mas da hipótese de que PA $\vdash \mathfrak{G}$, i.e., que PA $\vdash \forall y \neg\varphi(\overline{m}, y)$, obtemos:

$$\text{PA} \vdash \neg\varphi(\overline{m}, \overline{k}).$$

Ou seja, temos que PA é inconsistente; mas isso é impossível, já que PA é ω-consistente. Dessa forma, concluímos que PA $\nvdash \mathfrak{G}$.

Para o outro caso, veja que PA $\nvdash \mathfrak{G}$, que acabamos de provar, significa que não há sequer um número natural que codifica uma demonstração para a fórmula $\mathfrak{G}$. Assim,

para qualquer $n \in \mathbb{N}$ vale que $\neg R(m, n)$, e como φ representa tal relação, temos que para todo número natural n,

$$\mathrm{PA} \vdash \neg\varphi(\overline{m}, \overline{n}).$$

Daí, vem da ω-consistência de PA que $\mathrm{PA} \nvdash \exists y \neg[\neg\varphi(\overline{m}, y)]$, o que é equivalente a $\mathrm{PA} \nvdash \neg\forall y \neg\varphi(\overline{m}, y)$. Por sua vez, $\neg\forall\neg\varphi(\overline{m}, y)$ é a negação de $\mathfrak{Go}$, donde concluímos, finalmente, que $\mathrm{PA} \nvdash \neg\mathfrak{Go}$, e temos o resultado. □

Há, nessa prova, alguns pontos interessantes a se notar. Primeiro, veja que $\mathrm{PA} \nvdash \mathfrak{Go}$ estabelece *exatamente* o que a sentença $\mathfrak{Go}$ diz — i.e., que ela mesma não é demonstrável em PA. Logo, $\mathfrak{Go}$ é verdadeira. Nesse sentido, podemos derivar um argumento semântico para o teorema acima: supondo que $\mathfrak{Go}$ não fosse verdadeira, i.e., que ela tivesse uma prova em PA, então tal sistema provaria uma sentença falsa, o que é absurdo se considerarmos que PA é correto (i.e., um sistema cujos teoremas são apenas sentenças verdadeiras). Assim, $\neg\mathfrak{Go}$ deve ser falsa; como ela diz que pode, de fato, ser demonstrada em PA, então temos que $\neg\mathfrak{Go}$ também não é um teorema, e daí $\mathfrak{Go}$ é indecidível.

Outro aspecto é que a hipótese de ω-consistência não precisou ser usada na obtenção de $\mathrm{PA} \nvdash \mathfrak{Go}$. A bem da verdade, podemos construir, com as ferramentas que já temos, uma sentença diferente de $\mathfrak{Go}$ que também é indecidível, mas *sem precisar supor a condição mais forte de* ω*-consistência*. O responsável por essa construção foi Barkley Rosser (1907-1989), e por isso designaremos o símbolo $\mathfrak{Ro}$ a tal sentença. A maneira como procedemos é bastante similar ao que fizemos anteriormente: em primeiro lugar, considere a relação binária

$$P(a, b) \leftrightarrow \mathrm{Form}(a) \wedge \mathrm{Liv}(a, 0) \wedge \mathrm{Dem}(b, \mathrm{Sub}(\langle(6, a)\rangle, 0, \mathrm{Num}(a))).$$

Como na construção de $\mathfrak{Go}$, temos que a codifica uma fórmula (digamos, ξ) e que x ocorre livre em ξ. Aqui, b também será o código de uma demonstração; nesse caso, no entanto, não substituímos a na fórmula que este número codifica, mas em sua negação, $\neg\xi$. Evidentemente, P é recursiva, logo é fortemente representável em PA — digamos que pela fórmula $\vartheta(x, y)$. Lembrando a relação R definida anteriormente, e que φ a representa, seja $\chi(x)$ a seguinte fórmula:

$$\forall y\{\varphi(x, y) \rightarrow \exists z[z \leq y \wedge \vartheta(x, z)]\}.$$

Colocando $n = \ulcorner\chi(x)\urcorner$ e aplicando $\chi(n)$, temos a sentença de Rosser:

$$\forall y\{\varphi(\overline{n}, y) \rightarrow \exists z[z \leq y \wedge \vartheta(\overline{n}, z)]\}. \tag{13}$$

Agora, perceba que para cada $y \in \mathbb{N}$, temos que $R(n, y)$ é válida se e somente se y codifica uma prova para a fórmula cujo número de Gödel é n, i.e., a fórmula $\chi(n) = \mathfrak{Ro}$. Da mesma forma, temos $P(n, y)$ se e somente se y é o número de Gödel

de uma demonstração para $\neg\chi(n) = \neg Я$. Como a definição de Я, em (13), envolve a representação dessas duas ideias na linguagem de PA, vemos que a sentença de Rosser diz o seguinte:

> *Se há uma prova para* Я*, então também há um prova para* ¬Я*. Além disso, o número de Gödel da última é menor do que ou igual ao número de Gödel da primeira.*

Ou seja, temos que Я diz, sobre si mesma, que caso ela seja demonstrável em PA, então sua negação também deve ser. Logo, Я é autorreferente. Vejamos que, como Ɔ, também é indecidível.

Teorema 3.4.3 (Primeiro Teorema de Incompletude, versão de Rosser)**.** *Se* PA *é um sistema consistente, então é incompleto.*

Demonstração. Vamos proceder da mesma forma que no teorema 3.4.2; as únicas diferenças serão nossa hipótese e a complexidade da sentença que queremos provar ser indecidível. Assim, por contradição, suponha que PA $\vdash Я$. Destrinchando, temos que

$$\text{PA} \vdash \forall y\{\varphi(\overline{n}, y) \to \exists z[z \leq y \wedge \vartheta(\overline{n}, z)]\}. \tag{14}$$

Como Я é um teorema de PA, existe $k \in \mathbb{N}$ tal que k é o número de Gödel de uma prova para Я — ou seja, vale $R(n, k)$, do que segue que PA $\vdash \varphi(\overline{n}, \overline{k})$. Ora, mas substituindo y por $\overline{k}$ em (14), e eliminando o quantificador universal, temos que

$$\text{PA} \vdash \varphi(\overline{n}, \overline{k}) \to \exists z[z \leq \overline{k} \wedge \vartheta(\overline{n}, z)].$$

Portanto, via *modus ponens*, obtemos:

$$\text{PA} \vdash \exists z[z \leq \overline{k} \wedge \vartheta(\overline{n}, z)]. \tag{15}$$

Agora, como PA é consistente, ¬Я não pode ser um teorema; isso significa que para todo $y \in \mathbb{N}$, vale que $\neg P(n, y)$. Sendo ϑ a representação, em PA, da relação P, podemos concluir que PA $\vdash \neg\vartheta(\overline{n}, \overline{j})$, para todo $j \in \mathbb{N}$. Em particular, temos que

$$\text{PA} \vdash \neg\vartheta(\overline{n}, 0) \wedge \neg\vartheta(\overline{n}, \overline{1}) \wedge ... \wedge \neg\vartheta(\overline{n}, \overline{k}).$$

Mas pelo lema 3.2.3, é verdade que

$$\text{PA} \vdash \neg\vartheta(\overline{n}, 0) \wedge ... \wedge \neg\vartheta(\overline{n}, \overline{k}) \to \forall z[z \leq \overline{k} \to \neg\vartheta(\overline{n}, z)],$$

do que segue, novamente por *modus ponens*, que

$$\text{PA} \vdash \forall z[z \leq \overline{k} \to \neg\vartheta(\overline{n}, z)].$$

Isso, junto de (15), nos diz que PA é inconsistente, o que é absurdo. Assim, nossa hipótese PA $\vdash Я$ deve ser falsa, e concluímos que a sentença de Rosser não é demonstrável no sistema. Prossigamos mostrando que também não pode ser refutada.

Novamente por contradição, suponha que PA $\vdash \neg Я$. Reescrevendo, temos:

$$\mathsf{PA} \vdash \neg\forall y\{\varphi(\overline{n}, y) \to \exists z[z \leq y \wedge \vartheta(\overline{n}, z)]\}.$$

Seja, então, $l \in \mathbb{N}$ o número de Gödel de uma prova para $\neg Я$. Assim, vale $P(n, l)$, e daí $\mathsf{PA} \vdash \vartheta(\overline{n}, \overline{l})$. Como PA é consistente, tem-se que $\mathsf{PA} \nvdash Я$, i.e., nenhum número natural codifica uma demonstração para a sentença de Rosser, do que segue que $\neg R(n, y)$, qualquer que seja $y \in \mathbb{N}$. Temos, portanto, que $\mathsf{PA} \vdash \neg\varphi(\overline{n}, \overline{j})$, para todo $j \in \mathbb{N}$. Em particular,

$$\mathsf{PA} \vdash \neg\varphi(\overline{n}, 0) \wedge ... \wedge \neg\varphi(\overline{n}, \overline{l}).$$

Novamente usando o lema 3.2.3, e via *modus ponens*, temos que

$$\mathsf{PA} \vdash y \leq \overline{l} \to \neg\varphi(\overline{n}, y). \tag{16}$$

Por outro lado, perceba que, pela introdução do quantificador existencial, é verdade que

$$\mathsf{PA} \vdash [\overline{l} \leq y \wedge \vartheta(\overline{n}, \overline{l})] \to \exists z[z \leq y \wedge \vartheta(\overline{n}, z)].$$

Ora, acima já obtivemos $\mathsf{PA} \vdash \vartheta(\overline{n}, \overline{l})$, então também é verdade que

$$\mathsf{PA} \vdash \overline{l} \leq y \to \exists z[z \leq y \wedge \vartheta(\overline{n}, z)]. \tag{17}$$

Do lema 3.1.1(3), vem que $\mathsf{PA} \vdash \overline{l} \leq y \vee y \leq \overline{l}$. Juntando isso com o que obtivemos em (16) e (17), chegamos a

$$\mathsf{PA} \vdash \neg\varphi(\overline{n}, y) \vee \exists z[z \leq y \wedge \vartheta(\overline{n}, z)].$$

Finalmente, como $\neg p \vee q$ é equivalente a (ou melhor, é justamente a definição de) $p \to q$, basta usarmos a regra de generalização para obter:

$$\mathsf{PA} \vdash \forall y\{\varphi(\overline{n}, y) \to \exists z[z \leq y \wedge \vartheta(\overline{n}, z)]\},$$

i.e., $\mathsf{PA} \vdash Я$. Mas nosso sistema é consistente, então isso é impossível. Daí, concluímos que $\mathsf{PA} \nvdash \neg Я$, e a sentença de Rosser é indecidível. □

Note que, assim como a sentença de Gödel, Я é verdadeira, pois $\mathsf{PA} \nvdash Я$ significa que $\varphi(\overline{n}, y)$ é sempre falsa, logo a implicação presente em Я deve ser verdadeira, por vacuidade. Usando um argumento semântico, temos que caso Я fosse falsa, $\neg Я$ seria verdadeira, e ela diria que "minha negação é uma teorema de PA, mas eu não sou". Por separação, teríamos que uma sentença falsa (i.e., a negação de $\neg Я$) é um teorema de PA — o que é impossível, dado que PA é correto.

Ambas as provas que demos acima são, a bem da verdade, casos particulares de um ideia um pouco mais geral. Ao olharmos a construção das sentenças de Gödel e de Rosser, vamos observar que, antes de tudo, foi necessário construir relações recursivas que exprimem o processo de diagonalização. O resultado foi derivarmos *sentenças autorreferentes*, i.e., fórmulas fechadas E tais que $E \leftrightarrow P(\ulcorner E \urcorner)$, para determinadas propriedades P exprimíveis na linguagem de PA — no caso de Ↄ, P é a

propriedade de não ser demonstrável; neste sentido, o número de Gödel de Ɔ pode ser visto como um *ponto fixo* do predicado de não-demonstrabilidade. A forma geral que queremos estabelecer para o Primeiro Teorema de Incompletude se baseia, sobretudo, neste conceito.

A seguir, provaremos o resultado que sintetiza a ideia de autorreferência presentes nas sentenças de Gödel e de Rosser, ao estabelecer o fato de que toda fórmula aberta (numa única variável) de PA possui pontos fixos. Dessa forma, temos que Ɔ e Я são apenas casos especiais de tal resultado, quando o aplicamos a fórmulas que representam propriedades de demonstrabilidade dentro do sistema.

Lema 3.4.1 (Lema Diagonal)**.** *Para toda fórmula* $\varphi(x)$*, livre somente em* x*, existe uma sentença* ψ *tal que*

$$\mathsf{PA} \vdash \psi \leftrightarrow \varphi(\overline{\ulcorner \psi \urcorner}).$$

Demonstração. Primeiro, combinemos que $\mathbf{Sub}(x, y, z, w)$ é a representação de Sub na linguagem de PA. Fazendo $\mathrm{Sub}'(x, y, z) = \mathrm{Sub}(x, y, \mathrm{Num}(z))$ e representando tal relação por $\mathbf{Sub'}$, seja θ a seguinte fórmula:

$$\forall y[\mathbf{Sub'}(x, 0, x, y) \rightarrow \varphi(y)].$$

Ou seja, $\theta(x)$ diz que se y representa o resultado da substituição de uma variável pelo termo codificado por x na expressão codificada por x, então vale $\varphi(y)$. Seja, então, $m = \ulcorner \theta(x) \urcorner$, e faça ψ ser a diagonalização de θ, i.e., $\theta(\overline{m})$. Com isso, as seguintes equivalências são teoremas de PA:

$$\begin{aligned}
\psi &\leftrightarrow \theta(\overline{m}) \\
&\leftrightarrow \forall y[\mathbf{Sub'}(\overline{m}, 0, \overline{m}, y) \rightarrow \varphi(y)] \\
&\leftrightarrow \forall y[\mathbf{Sub'}(\overline{\ulcorner \theta(x) \urcorner}, 0, \overline{m}, y) \rightarrow \varphi(y)] \\
&\leftrightarrow \varphi(\overline{\ulcorner \theta(\overline{m}) \urcorner}) \\
&\leftrightarrow \varphi(\overline{\ulcorner \psi \urcorner}).
\end{aligned}$$

[Perceba que a penúltima equivalência se dá porque caso seja verdade que "y representar a substituição da variável livre em θ pelo termo m" implica $\varphi(y)$, então é verdade que a aplicação dessa substituição em φ vale.] □

Para provar o teorema em sua forma geral, considere a relação recursiva Dem, e seja **Dem** uma fórmula que a represente fortemente. Colocando **Bew** como sendo a fórmula $\exists x\mathbf{Dem}(x, y)$, temos que **Bew** representa a relação Bew(y). Tomando uma fórmula τ arbitrária e supondo que ela seja um teorema de PA, então é óbvio que a relação Bew($\ulcorner \tau \urcorner$) é válida (pois ela diz, exatamente, que τ é demonstrável no sistema), e daí $\mathbf{Bew}(\overline{\ulcorner \tau \urcorner})$ é, também, um teorema de PA — ou seja:

$$\text{se } \mathsf{PA} \vdash \tau, \text{ então } \mathsf{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}).$$

A recíproca, em geral, não é verdadeira. Perceba, no entanto, que o Lema Diagonal nos garante que $\neg$**Bew** possui um ponto fixo. Isso, junto da recíproca da implicação metateórica acima, serão as hipóteses do resultado que queremos mostrar:

Teorema 3.4.4 (Primeiro Teorema de Incompletude, Formal Geral)**.** *Seja* τ *uma sentença qualquer na linguagem de* PA*, e suponha que* PA $\vdash \tau \leftrightarrow \neg$***Bew***$(\overline{\ulcorner\tau\urcorner})$*. Então, tem-se que*

1. PA $\nvdash \tau$*;*

2. *se para toda sentença* ψ *na linguagem de* PA *for verdade que*

$$\textit{se } \text{PA} \vdash \textbf{\textit{Bew}}(\overline{\ulcorner\psi\urcorner}), \textit{ então } \text{PA} \vdash \psi,$$

   então PA $\nvdash \neg\tau$*.*

Demonstração. Por contradição, suponha que PA $\vdash \tau$. Então, do comentário feito anteriormente, vem que PA $\vdash$ **Bew**$(\overline{\ulcorner\tau\urcorner})$. Mas como τ é ponto fixo de $\neg$**Bew**, e **Bew** equivale a $\neg\neg$**Bew**, concluímos que PA $\vdash \neg\tau$, o que é absurdo, assumindo que PA é consistente. Daí, PA $\nvdash \tau$.

Para o segundo item, suponha que PA $\vdash \neg\tau$. Então, PA $\vdash \neg\neg$**Bew**$(\overline{\ulcorner\tau\urcorner})$, do que segue que PA $\vdash$ **Bew**$(\overline{\ulcorner\tau\urcorner})$. Isso, por hipótese, implica PA $\vdash \tau$, o que contradiz o primeiro item. Assim, PA $\nvdash \neg\tau$. □

Como um último comentário (por ora) sobre o Primeiro Teorema, perceba que tanto nos casos particulares quanto no caso geral, que acabamos de provar, a sentença que construímos tem a forma $\forall y[\eta(y)]$, para alguma fórmula η. Com efeito: em $\mathfrak{Ɔ}$, temos que η é $\neg\varphi(\overline{m}, y)$, com φ representando a relação R. Como PA $\nvdash \mathfrak{Ɔ}$, e $R(m, b)$ nos diz que b codifica uma prova para o resultado de se substituir, na fórmula de código m, o próprio número m, concluímos que para todo $n \in \mathbb{N}$, PA $\vdash \neg\varphi(\overline{m}, \overline{n})$, i.e., PA $\vdash \eta(\overline{n})$; no entanto, a fórmula com quantificador universal, i.e., $\mathfrak{Ɔ}$, não pode ser provada. Em outras palavras, cada uma das instâncias da sentença indecidível de Gödel, $\forall y[\eta(y)]$, é demonstrável em PA, e, portanto, decidível. [Já a sentença do caso geral é da forma $\neg\exists y$**Dem**$(y, \overline{\ulcorner\tau\urcorner})$, equivalente a $\forall y[\eta(y)]$, em que $\eta(y)$ é, simplesmente, $\neg$**Dem**$(y, \overline{\ulcorner\tau\urcorner})$.]

3.4.2 Segundo Teorema: a indecidibilidade da consistência

Prossigamos, então, com o Segundo Teorema. A bem da verdade, podemos vê-lo como um mero corolário do Primeiro, pois, assumindo ω-consistência, ele simplesmente nos indicará mais uma sentença indecidível em PA. O que o torna tão importante é, justamente, o conteúdo de tal sentença; chamando-a por Con_{PA}, ela diz que *o sistema* PA *é consistente*. Ou seja, assumindo que PA é consistente, vamos

concluir que é impossível, usando as ferramentas do sistema, provar esta hipótese (bem como refutá-la).

Como o fato de uma teoria ser consistente equivale ao fato de tal teoria não demonstrar qualquer contradição, parece bastante natural tentar definir $\mathrm{Con}_{\mathrm{PA}}$ como sendo a fórmula $\neg\mathbf{Bew}(\overline{c})$, em que $c \in \mathbb{N}$ codifica a negação de um teorema de PA. Ora, sendo $0 \neq \overline{1}$ demonstrável no sistema, basta, então, fazermos

$$\mathrm{Con}_{\mathrm{PA}} \leftrightarrow \neg\mathbf{Bew}(\overline{\overline{0 = \overline{1}}}).$$

Reforçando o que dissemos acima, poderíamos colocar qualquer outra afirmação absurda no lugar de $0 = 1$, como $2 + 2 = 5$ ou $1 < 0$, pois, quando aplicamos seus números de Gödel à função Bew, representada por **Bew**, temos uma sentença falsa. Mas será que essa é, essencialmente, a única maneira de se aritmetizar consistência?

Veja que o modo como definimos $\mathrm{Con}_{\mathrm{PA}}$ é, antes de tudo, uma maneira *indireta* de dizer que PA é consistente: embora todas as sentenças na forma $\neg\mathbf{Bew}(\overline{c})$ sejam equivalentes, nós *precisamos escolher* um absurdo em específico. Isso contrasta com Ↄ e Я, por exemplo, que foram obtidas, via diagonalização, de relações que descreviam explicitamente a noção de demonstrabilidade de alguma fórmula aritmética. Assim, o questionamento acima se mostra mais delicado do que aparenta, e, em decorrência de uma eventual resposta negativa, faz surgir outra dúvida: havendo outras formas de expressar consistência, será que alguma delas é demonstrável — ou, num cenário catastrófico — refutável em PA? Se esse for o caso, o Segundo Teorema de Incompletude é, ironicamente, "incompleto" (i.e., num sentido totalmente diverso de tal palavra: *neste* viés, tal teorema não garantiria a indecidibilidade de *todas* as sentenças que traduzem a consistência de PA).

Uma caracterização para a consistência de PA que difere da forma $\neg\mathbf{Bew}(\overline{c})$ é a seguinte:

$$\mathrm{Con}'_{\mathrm{PA}} \leftrightarrow \neg\exists x[\mathbf{Bew}(x) \wedge \mathbf{Bew}(\dot{\neg}x)],$$

em que $\dot{\neg}(x)$ representa a relação recursiva que leva o número de Gödel de uma fórmula ao número de Gödel de sua negação. Nesse caso, $\mathrm{Con}'_{\mathrm{PA}}$ diz que não existe um número que codifica uma fórmula que é, ao mesmo tempo, demonstrável e refutável. Perceba que isso, no entanto, traduz a mesma propriedade que $\neg\mathbf{Bew}(\overline{c})$: em PA, não podemos provar uma sentença e, depois, também provar sua negação. Assim, não é surpresa alguma que $\mathrm{Con}_{\mathrm{PA}}$ e $\mathrm{Con}'_{\mathrm{PA}}$ são equivalentes, i.e., que

$$\mathrm{PA} \vdash \mathrm{Con}_{\mathrm{PA}} \leftrightarrow \mathrm{Con}'_{\mathrm{PA}}.$$

O mesmo ocorre com várias outras sentenças que expressam a ideia de consistência.

Aqui, então, poderíamos ir além, e nos perguntar se não existe alguma outra expressão, menos óbvia, mais intrincada, que indique a consistência de PA. Em verdade, temos que sim, tais caracterizações existem; algumas delas, inclusive, *podem*

ser demonstradas em PA! Mas deixemos essa discussão para depois da demonstração. Para finalizarmos a parte introdutória ao Segunda Teorema, coloquemos apenas o seguinte: se quisermos que a relação representada por $\mathrm{Con}_{\mathsf{PA}}$ — i.e., $\neg\mathrm{Bew}(c)$, com $c \in \mathbb{N}$ codificando um absurdo — satisfaça algumas propriedades interessantes, então, de fato, a forma geral $\neg\mathbf{Bew}(\overline{c})$ é tudo o que gostaríamos de provar, quando dizemos que queremos que PA seja consistente.

As propriedades a que aludimos acima são chamadas de *condições de derivabilidade dos predicados de demonstrabilidade*. Elas sintetizam o que é esperado de uma função que diz se "algo é o número de Gödel de um teorema de PA". Colocando **D** como uma fórmula que representa um predicado qualquer como esse, temos:

(D1) Se $\mathsf{PA} \vdash \varphi$, então $\mathsf{PA} \vdash \mathbf{D}(\overline{\ulcorner \varphi \urcorner})$;

(D2) $\mathsf{PA} \vdash \mathbf{D}(\overline{\ulcorner \varphi \urcorner}) \to \mathbf{D}(\overline{\mathbf{D}(\overline{\ulcorner \varphi \urcorner})})$;

(D3) $\mathsf{PA} \vdash \mathbf{D}(\overline{\ulcorner \varphi \to \psi \urcorner}) \to [\mathbf{D}(\overline{\ulcorner \varphi \urcorner}) \to \mathbf{D}(\overline{\ulcorner \psi \urcorner})]$.

Como já explicamos anteriormente, a propriedade (D1) é satisfeita trivialmente se o predicado **D** expressa a relação de demonstrabilidade em PA; assim, temos que, em particular, **Bew** satisfaz tal propriedade. (Na verdade, usamos este fato na demonstração do último teorema.) Prosseguindo, (D2) é simplesmente a formalização de (D1), no sentido de que a implicação metateórica, presente no primeiro item, adquire um caráter puramente sintático, expressado pelo símbolo $\to$. Finalmente, (D3) mostra como a demonstrabilidade é preservada pela principal regra de inferência do sistema (*modus ponens*). A rigor, poderíamos provar que não só (D1), mas todas elas são satisfeitas por **Bew**. No entanto, isso seria, ao mesmo tempo, tedioso e demorado, de modo que apenas indicamos a leitura do segundo capítulo de (BOOLOS, G. S., 2008). Com isso, podemos provar o Segundo Teorema:

Teorema 3.4.5 (Segundo Teorema de Incompletude)**.** *Se* PA *é* ω*-consistente, então* $\mathrm{Con}_{\mathsf{PA}}$ *é indecidível.*

Demonstração. De início, nossa estratégia será mostrar que $\mathrm{Con}_{\mathsf{PA}}$ é equivalente a um ponto fixo de $\neg\mathbf{Bew}$. Pelo teorema 3.4.4, $\mathrm{Con}_{\mathsf{PA}}$ não será um teorema de PA. Seja, então, τ uma fórmula na linguagem de PA tal que $\mathsf{PA} \vdash \tau \leftrightarrow \neg\mathbf{Bew}(\overline{\ulcorner \tau \urcorner})$. Pelo Princípio de Explosão, temos que

$$\mathsf{PA} \vdash (0 = \overline{1}) \to \tau.$$

Assim, pela propriedade (D1), ficamos com

$$\mathsf{PA} \vdash \mathbf{Bew}(\overline{\ulcorner (0 = \overline{1}) \to \tau \urcorner}).$$

Mas isso, junto da propriedade (D3), deve nos dar

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner 0 = \overline{1} \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}),$$

e pela contrapositiva, obtemos $\text{PA} \vdash \neg\mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \neg\mathbf{Bew}(\overline{\ulcorner 0 = \overline{1} \urcorner})$. Agora, pela definição de τ, temos que $\text{PA} \vdash \tau \to \neg\mathbf{Bew}(\overline{\ulcorner \tau \urcorner})$. Essas duas últimas nos dão, imediatamente:

$$\text{PA} \vdash \tau \to \neg\mathbf{Bew}(\overline{\ulcorner 0 = \overline{1} \urcorner}).$$

Ou seja, $\text{PA} \vdash \tau \to \text{Con}_{\text{PA}}$, e temos um dos lados da equivalência que buscamos.

Para a recíproca, primeiro observe que a propriedade (D2), aplicada à fórmula τ, nos dá

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \urcorner}). \tag{18}$$

Novamente usando a definição de τ, temos que $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \neg\tau$ (usando contrapositiva). Assim, por (D1) e (D2), temos sucessivamente que

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \neg\tau \urcorner});$$

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \neg\tau \urcorner}). \tag{19}$$

Consequentemente, de (18) e (19) vem que

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \neg\tau \urcorner}). \tag{20}$$

Observando que $\neg\tau \to [\tau \to (\neg\tau \wedge \tau)]$ é um teorema de PA (pois é uma tautologia), usamos (D1) e (D3) para obter:

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \neg\tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \tau \wedge \neg\tau \urcorner}). \tag{21}$$

Mas veja que $[p \to (q \to r)] \to [(q \to p) \to (q \to r)]$ também é uma tautologia. Logo, fazendo as substituições apropriadas, temos que

$$\begin{aligned}\text{PA} \vdash &\{\mathbf{Bew}(\overline{\ulcorner \neg\tau \urcorner}) \to [\mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \tau \wedge \neg\tau \urcorner})]\} \to \\ &\to \{[\mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \neg\tau \urcorner})] \to [\mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \tau \wedge \neg\tau \urcorner})]\}.\end{aligned}$$

Agora, como as implicações acima, relativas a $p \to (q \to r)$ e $q \to p$, são exatamente o que obtivemos em (21) e (20), temos imediatamente que

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \tau \wedge \neg\tau \urcorner}). \tag{22}$$

Ora, $\tau \wedge \neg\tau$ é equivalente a $(0 = \overline{1})$, do que segue que $\text{PA} \vdash (\tau \wedge \neg\tau) \to (0 = \overline{1})$. Usando (D1) e (D3), ficamos com

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \wedge \neg\tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner 0 = \overline{1} \urcorner}).$$

Isso, juntamente de (22), nos dá $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \tau \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner 0 = \overline{1} \urcorner})$. Pela contrapositiva, obtemos:

$$\text{PA} \vdash \neg\mathbf{Bew}(\overline{\ulcorner 0 = \overline{1} \urcorner}) \to \neg\mathbf{Bew}(\overline{\ulcorner \tau \urcorner}),$$

e das definições de $\mathsf{Con}_{\mathsf{PA}}$ e τ, vem que $\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}} \to \tau$. Com isso, completamos a prova da recíproca, e finalmente temos que

$$\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}} \leftrightarrow \tau,$$

i.e., a sentença que expressa a consistência de PA é equivalente a um ponto fixo de $\neg\mathbf{Bew}$. Portanto, $\mathsf{PA} \nvdash \mathsf{Con}_{\mathsf{PA}}$.

Para ver que a consistência de PA tampouco pode ser *refutada*, primeiramente veja que, sob a *hipótese* de PA ser consistente (presente no enunciado deste teorema), temos que:

$$\textit{para qualquer } n \in \mathbb{N},\ \mathsf{PA} \vdash \neg\mathbf{Dem}(\overline{n}, \overline{\ulcorner 0 = \overline{1} \urcorner}),$$

em que $\mathbf{Dem}(x, y)$ é a fórmula que representa a relação Dem. Assim, caso *pudéssemos* provar a consistência de PA, teríamos que

$$\mathsf{PA} \vdash \forall y \neg\mathbf{Dem}(y, \overline{\ulcorner 0 = \overline{1} \urcorner}).$$

Ou seja, temos que a não-demonstrabilidade da consistência de PA se traduz como a *impossibilidade* de se obter a conclusão acima a partir do fato de que nenhum $n \in \mathbb{N}$, em particular, prova um absurdo. Por isso — i.e., por PA não provar algo que seria esperado de uma axiomática que se baseia nos números naturais e, principalmente, em sua *ordem* inerente —, chamamos PA de um sistema ω-*incompleto*.

Suponha, agora, num cenário equivalente à hecatombe da Matemática como um todo, que possamos provar que PA é *inconsistente*, i.e., que $\mathsf{PA} \vdash \neg\mathsf{Con}_{\mathsf{PA}}$. Escrevendo como acima, teríamos que $\mathsf{PA} \vdash \neg\forall y \neg\mathbf{Dem}(y, \overline{\ulcorner 0 = \overline{1} \urcorner})$ — ou, melhor ainda:

$$\mathsf{PA} \vdash \exists y \mathbf{Dem}(y, \overline{\ulcorner 0 = \overline{1} \urcorner}).$$

Ora, mas isso, junto ao fato de que todo $n \in \mathbb{N}$ prova que $\neg\mathbf{Dem}(\overline{n}, \overline{\ulcorner 0 = \overline{1} \urcorner})$, nos diz que PA é ω-*in*consistente, o que é absurdo. Assim, concluímos que

$$\mathsf{PA} \nvdash \neg\mathsf{Con}_{\mathsf{PA}},$$

e finalmente temos o resultado. □

Algo a se notar no procedimento de prova acima é que, de maneira geral, fomos capazes de mostrar a equivalência entre $\mathsf{Con}_{\mathsf{PA}}$ e qualquer sentença que enuncia sua não-demonstrabilidade (i.e., qualquer ponto fixo de $\neg\mathbf{Bew}$). Pela transitividade da equivalência, isso nos diz que dados quaisquer pontos fixos de $\neg\mathbf{Bew}$ — digamos, fórmulas φ e ψ tais que

$$\mathsf{PA} \vdash \varphi \leftrightarrow \neg\mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \text{ e } \mathsf{PA} \vdash \psi \leftrightarrow \neg\mathbf{Bew}(\overline{\ulcorner \psi \urcorner}),$$

deve-se ter que $\mathsf{PA} \vdash \varphi \leftrightarrow \psi$. Em particular, quaisquer duas sentenças de Gödel são equivalentes (bem como quaisquer duas sentenças de Rosser o são).

O leitor também deve perceber que, uma vez formulada a sentença que afirma a consistência de PA, podemos dar um aspecto realmente formal ao Primeiro Teorema. De fato: tudo que ele diz é que "se a Aritmética de Peano for consistente, então há alguma sentença indecidível dentro dela". Ora, tomando uma sentença indecidível qualquer — como Ɔ ou Я —, que sabemos existir, pois já provamos o Primeira Teorema, essa ideia metateórica pode ser descrita pela seguinte implicação:

$$\mathrm{Con}_{\mathrm{PA}} \rightarrow \neg\mathbf{Bew}(\overline{\ulcorner Ɔ \urcorner}) \wedge \neg\mathbf{Bew}(\overline{\ulcorner \neg Ɔ \urcorner}).$$

Portanto, o que segue do Primeiro Teorema é que tal implicação é um teorema de PA — i.e., que

$$\mathrm{PA} \vdash \mathrm{Con}_{\mathrm{PA}} \rightarrow \neg\mathbf{Bew}(\overline{\ulcorner Ɔ \urcorner}) \wedge \neg\mathbf{Bew}(\overline{\ulcorner \neg Ɔ \urcorner}).$$

Agora, como prometido, vamos exemplificar uma sentença que também expressa a consistência de PA, mas que *pode* ser decidida dentro do sistema. Primeiramente, chame de n_0 o número de Gödel da sentença $0 = \overline{1}$, e considere a fórmula $\mathbf{Dem}'(x, y)$, dada por

$$\mathbf{Dem}(x, y) \wedge \neg\mathbf{Dem}(x, n_0).$$

Como $0 = \overline{1}$ não é um teorema de PA, pode-se ver facilmente que para quaisquer $m, n \in \mathbb{N}$, a equivalência $\mathbf{Dem}(\overline{m}, \overline{n}) \leftrightarrow \mathbf{Dem}'(\overline{m}, \overline{n})$ é um teorema de PA. Em outras palavras, esta nova fórmula, $\mathbf{Dem}'$, também representa fortemente a relação Dem. Assim, chame por $\mathrm{Con}''_{\mathrm{PA}}$ a seguinte fórmula:

$$\neg\exists x\mathbf{Dem}'(x, \overline{n_0}),$$

que abrevia $\neg\exists x[\mathbf{Dem}(x, \overline{\ulcorner 0 = \overline{1} \urcorner}) \wedge \neg\mathbf{Dem}(x, \overline{\ulcorner 0 = 1 \urcorner})]$. Mas isso é perfeitamente demonstrável em PA, uma vez que estamos assumindo a Lei de Não-Contradição (presente da lógica de primeira-ordem embutida no sistema). Por conseguinte, temos que

$$\mathrm{PA} \vdash \mathrm{Con}''_{\mathrm{PA}},$$

o que prova que a consistência de PA, quando formulada dessa maneira, é, de fato, decidível em PA. Perceba, no entanto, que $\mathbf{Dem}'$ não satisfaz as condições (D1) e (D2); assim, embora tal sentença expresse a consistência de PA, não podemos dizer que ela é um *predicado de demonstrabilidade*.

Além do mais, mesmo que $\mathrm{Con}''_{\mathrm{PA}}$ seja um teorema, e que disso possamos concluir que "se PA é consistente, então podemos demonstrar um sentença que expressa tal consistência", de maneira nenhuma podemos inferir que a Aritmética de Peano é, de fato, consistente. Com efeito: se PA fosse *in*consistente, nós também poderíamos provar $\mathrm{Con}''_{\mathrm{PA}}$ — pois poderíamos provar qualquer coisa —, e isso com certeza não nos diria que PA é consistente. Como apontado em (SMULLYAN, 1992), acreditar na

consistência de um sistema simplesmente pelo fato de que ele consegue provar sua consistência é tão insensato quanto acreditar piamente nas palavras de uma pessoa que afirma nunca mentir.[21] No mesmo sentido, o fato de o Segundo Teorema nos dizer que PA não consegue decidir sobre sua própria consistência tampouco deve ser encarado como algo que impossibilite tal consistência. A bem da verdade, ele apenas nos diz que PA é incapaz de derivar uma certa sentença sobre si mesmo, bem como sua negação (lembre-se: a indecidibilidade de $\mathrm{Con}_{\mathrm{PA}}$ é apenas um caso particular de algo que provamos no Primeiro Teorema).

Para finalizar esta subseção, toquemos num ponto crucial dos teoremas de incompletude, concernente à extensão e abrangência dos mesmos. Pois bem: tudo que vimos até agora é que um sistema aritmético específico, que formaliza a teoria elementar dos números naturais (ao que demos o nome de PA), possui, sob a hipótese de ω-consistência, ao menos três sentenças indecidíveis — a saber, Ə, Я e $\mathrm{Con}_{\mathrm{PA}}$ (embora, como apontado acima, todas sejam equivalentes, uma vez que todas são pontos fixos da fórmula que exprime a não-demonstrabilidade). Sendo assim, o leitor poderia se perguntar: "Ora, se quisermos construir um sistema PA$'$, semelhante a PA, que seja completo, não bastaria adicionar uma dessas sentenças (e, por conseguinte, todas as três) como um novo axioma?". Em outras palavras, será que não é possível estender PA a um sistema PA$'$, cujo axioma ($P8$) é, por exemplo, Ə, de modo que PA$'$ não tenha sentenças indecidíveis?[22]

Infelizmente, a resposta é negativa. Mas é justamente isso que torna os teoremas de incompletude tão fortes.[23] Nesse sentido, dizemos que PA não é apenas incompleto, mas *essencialmente* incompleto, ou *incompletável*: mesmo que tomemos Ə como um axioma, de forma que seja uma sentença decidível em PA$'$, ainda teremos como interpretar PA$'$ da mesma forma que interpretamos PA (já que o modelo de ambas seria o modelo padrão, $\mathfrak{N}_0$). Assim, o mesmo método usado para construir Ə em PA pode ser aplicado para construir, em PA$'$, uma sentença indecidível `}$Ə. Podemos fazer isso indefinidamente, adicionado $\text{`Ə como um novo axioma da extensão PA$''$ de PA$'$, e produzindo uma nova sentença indecidível ``Ə, que se tornará um axioma da ex-

[21] Essa analogia pode ir muito longe. Numa outra obra do mesmo autor, referenciada em (SMULLYAN, 1987), muitos dos aspectos e consequências dos teoremas de incompletude são tratados por um viés *psicológico*. Ao considerar sistemas formais como sendo seres conscientes que argumentam (i.e., raciocinam dentro de suas cabeças) usando regras análogas às de inferência, e fazendo o paralelo entre "demonstrar" (uma noção intrínseca a sistemas formais) e "acreditar" (uma noção intrínseca à razão humana), podemos interpretar o Segundo Teorema como uma constatação do seguinte tipo, feita por um ser humano qualquer que raciocina sobre sua própria razão: "Mesmo que seja impossível eu me contradizer, *eu não consigo acreditar* que eu jamais vá entrar em contradição em algum raciocínio".

[22] A rigor, dada uma teoria (i.e., um sistema formal) T qualquer, dizemos que T' *estende* T (ou que T é uma *subteoria de* T') quando T' possui as mesmas constantes não-lógicas de T, e consegue provar todos os teoremas de T.

[23] A bem da verdade, a prova original de Gödel abrange uma classe muito maior de sistemas formais, não apenas aqueles que se baseiam em lógica de primeira ordem, pois se dá no campo da teoria de tipos (que Russell e Whitehead descrevem em seu *Principia*).

tensão PA''', etc. Como consequência, obtemos uma quantidade infinita (enumerável) de sistemas aritméticos incompletos.

A pergunta que naturalmente surge é se existem sistemas formais *menos* robustos que PA (i.e., algum fragmento deste sistema) para os quais os teoremas de incompletude *não* se aplicam. A resposta, como esperado, também é negativa (obviamente, nesse caso, se considerarmos sistema formais *aritméticos*, i.e., teorias que buscam definir propriedades dos números naturais). Acontece que as sentenças indecidíveis $\mathfrak{G}$ e $\mathrm{Con}_{\mathrm{PA}}$, cujas construções apresentamos, também podem ser formuladas em teorias mais fracas, como a aritmética de Robinson — devido a R.M. Robinson —, igualmente conhecida como *aritmética minimal*, ou Q. Nela, substituímos o esquema (enumerável) de axiomas de indução pelo único axioma

$$\forall x[x \neq 0 \rightarrow \exists y(x = \mathrm{s}(y))].$$

Essa é uma teoria bastante fraca: com ela, nem ao menos conseguimos provar que as operações elementares ($+$ e $*$) são comutativas. No entanto, pode-se provar que todas as relações recursivas são representáveis nesse sistema, de modo que, supondo-a consistente, deve ser incompleta. Além disso, como tal axioma é um teorema de PA, temos que Q é uma *subteoria* de PA. Pelo mesmo argumento usado no parágrafo anterior, concluímos que Q também é essencialmente incompleta.[24]

Com isso, é bastante plausível concluirmos que *qualquer* formalização de *qualquer* fragmento da Aritmética (considerando Q como a teoria mínima) deve ser incompleto. Mais ainda, todas devem ser *incompletáveis*. Evidentemente, por esse não ser nosso objeto de estudo, não fugiremos do escopo a ponto de mostrar como cada sistema aritmético formal (de primeira ou de segunda ordem) deve ser essencialmente incompleto;[25] acreditamos, no entanto, ter dado razões suficientes para o leitor se convencer de quão fortes (e amplos) são os teoremas de incompletude de Gödel.

3.4.3 Sobre a existência de Papai Noel, a indefinibilidade da verdade e outras consequências dos Teoremas

Há algumas conclusões que nos saltam aos olhos quando nos deparamos com os enunciados, bem como com os métodos de prova, dos teoremas de incompletude. Na mesma medida, é natural que vários questionamentos surjam quanto à aplicabili-

[24] É comum, na literatura, provar a incompletude de Q, no lugar de PA, pois, como o nome sugere, Q é a teoria mais enxuta possível para descrever relações aritméticas — e sendo ela incompleta, qualquer extensão da mesma (como é o caso de PA) também deve ser. Nossa escolha em descrever PA, e não Q, se deve principalmente ao fato de PA provar a maioria dos resultados elementares de teoria dos números (algo que evidenciamos ser Q incapaz de fazer).

[25] Em (SMULLYAN, 1992), por exemplo, prova-se que a aritmética *com exponenciação* também é incompleta. Mais geralmente, pode-se provar que a própria teoria dos conjuntos (com a axiomática ZFC, a mais comum para descrever teorias matemáticas) tem sentenças indecidíveis — e.g., a hipótese do *continuum*, ou a sentença que expressa sua consistência, $\mathrm{Con}_{\mathrm{ZFC}}$.

dade e consequências dos mesmos — e alguns deles podem levar a um desentendimento do que os Teoremas realmente dizem. Em outras palavras, parte das conclusões que percebemos como imediatas são, em verdade, bastante errôneas. Isso é perfeitamente compreensível, uma vez que tais resultados, exprimíveis por afirmações relativamente simples (ao ponto de facilmente poderem ser enunciados em linguagem natural), têm conteúdos imensamente significativos para uma área tão abrangente do conhecimento; por conseguinte, a popularização dos Teoremas, em semelhança ao que se vê acontecer com outras tantas construções teóricas que alcançaram o senso comum — como o Princípio de Incerteza de Heisenberg, ou a própria Teoria da Relatividade Geral de Einstein —, inevitavelmente leva a interpretações levianas.

Certamente, o desentendimento mais comum se dá quando alguém extrapola o escopo dos Teoremas, tentado aplicá-los a "sistemas" que não são aritméticos. Há incontáveis exemplos desse tipo de extrapolação; em especial, há quem diga que os teoremas de incompletude implicam a existência de Deus (por haver "verdades" no mundo que nós, humanos, não conseguimos estabelecer somente com nosso raciocínio), ou, ainda, que nenhum conjunto de leis de uma nação é capaz de reger todas suas questões sociais. Neste ponto do trabalho, o leitor prontamente deve perceber como a última afirmação, embora seja correta,[26] não tem qualquer relação com os resultados apresentados aqui. Os teoremas de incompletude são propriedades intrínsecas a sistemas *aritméticos*. Até mesmo em outras teorias que se provam incompletas e que não descrevem somente relações entre números naturais, como ZFC, as sentenças que conseguimos provar ser indecidíveis (pelos métodos que vimos[27]) são escritas em linguagem aritmética — ou seja, é na porção aritmética da teoria dos conjuntos que estabelecemos sua incompletude. Obviamente, não se espera que a Constituição Federal responda questões sobre números naturais. Dessa forma, não faz sentido algum tentar aplicar os teoremas de Gödel no campo jurídico.

O segundo exemplo, no entanto, possui diversas sutilezas. Pode-se dizer que a conclusão (errônea) sobre uma existência divina abarca boa parte das extrapolações levianas dos Teoremas (como a ideia de que uma Teoria de Tudo, perseguida há vários anos pelos físicos, é impossível de ser alcançada, pois teorias físicas são modelos matemáticos, logo estão sujeitas à incompletude, etc.). Mas também há um aspecto bastante pertinente neste debate, que ainda é objeto de muitas discussões filosóficas sobre os teoremas de incompletude, e que diz respeito à *capacidade mental* do ser humano — ou, mais especificamente, à possibilidade de emular nosso cérebro por al-

[26] Tendo em vista que uma sociedade está em constante desenvolvimento, um conjunto de leis abrangente nesse nível utópico seria não somente gigantesco (e, portanto, impraticável), como permanentemente mutável.

[27] A indecidibilidade da hipótese do *continuum* é demonstrada por métodos bastante diferentes (nem um pouco triviais) e não seguem dos teoremas de incompletude.

guma espécie de computador. Como o próprio Gödel apontou em sua Gibbs Lecture[28] de 1951 [cf. (WANG, 1997), p.185],

> Ou a mente humana supera todas as máquinas (i.e., ela pode decidir mais questões de teoria dos números do que qualquer máquina), ou existem questões de teoria dos números que não podem ser decididas pela mente humana. [E essas duas hipóteses não são mutuamente excludentes.]

Esse comentário — que Solomon Feferman (1928-2016) apresenta como "a dicotomia de Gödel" em seu artigo intitulado *Are there absolutely unsovable problems?*[29] — pode ser visto, de fato, como uma consequência da incompletude dos sistemas aritméticos. Ao menos quando consideramos "máquina" no sentido dado por Turing (que, ainda hoje, é a essência de todo computador físico), o que os Teoremas nos dizem é que, ao tentarmos identificar nossos argumentos matemáticos (frutos do raciocínio e trabalho árduo humano) com um programa de computador (que verifica mecanicamente se dada sentença é ou não um teorema), inevitavelmente iremos descobrir que algumas conjecturas matemáticas não possuem respostas. Ou seja, se as abstrações de processos efetivos de decisão que conhecemos são as únicas formas de se modelar o raciocínio matemático humano, então realmente há questões sobre números naturais que nunca poderão ser demonstradas, não importa quanto nosso intelecto evolua, nem quanto tempo nossa espécie sobreviva. Por outro lado, se máquinas de Turing, funções recursivas, algoritmos de Markov, etc., *não são* suficientes para emular toda nossa capacidade intelectual (com respeito às abstrações matemáticas), então a mente humana deve ser mais robusta do que qualquer computador.

Em qualquer um dos casos — i.e., sendo a mente humana simulável por máquinas ou não —, podemos nos indagar se existem problemas tão difíceis que nem mesmo nosso intelecto, nem qualquer outra abstração mais poderosa de máquinas, podem decidi-los.[30] A estes, dá-se o nome de problemas (ou sentenças) *absolutamente indecidíveis*. Nesse sentido, as sentenças que construímos (Ð, Я e $\mathrm{Con}_{\mathrm{PA}}$) *não são* indecidíveis em absoluto, pois podem ser trivialmente demonstradas nas extensões de PA que adicionam tais sentenças como axiomas (ou podem ser decididas em outras teorias, diferentes de PA).[31] É isso que Feferman quer dizer com a "dicotomia" de Gödel; de qualquer forma, o próprio autor compartilha a opinião que, de fato, nenhuma

[28] Prêmio concedido anualmente pela Sociedade Americana de Matemática (AMS); o premiado profere uma palestra.

[29] Cf. (FEFERMAN, 2006a).

[30] E caso a mente humana possa ser emulada por uma máquina, então os teoremas de incompletude garantem a existência de problemas desse tipo — i.e., que nunca poderão ser decididos via papel, caneta, tentativa, erro e inspiração.

[31] Isso não contradiz o fato de que PA é essencialmente incompleto, i.e., tal que quaisquer de suas extensões sejam incompletas, como apontado anteriormente. O que ocorre é que as sentenças mencionadas *podem ser* demonstradas ou refutadas em alguma dessas extensões.

máquina é capaz de modelar a mente humana, e que, mesmo assim, devem existir sentenças tão fortemente indecidíveis que não importa quais axiomas escolhamos, ou em qual teoria as interpretemos, elas continuarão sendo indemonstráveis e irrefutáveis.[32]

Em seu artigo *On the Question of Absolute Undecidabilty*[33], Peter Koellner elenca três das maiores candidatas a sentenças absolutamente indecidíveis: na análise matemática, (i) a sentença que diz que todo conjunto projetivo de números reais é Lebesgue-mensurável; na teoria dos conjuntos, (ii) a hipótese do *continuum* (HC), por várias vezes aqui mencionada, e (iii) a sentença que diz que todo conjunto é construtível (afirmação esta que implica HC). Para os fins deste trabalho, é desimportante entender o que (i) e (iii) significam — até porque existem várias outras candidatas à insolubilidade absoluta (como a conjectura de Kaplanski, ou o problema de Whitehead na teoria de grupos). Aqui, o que realmente importa, como o próprio Koellner aponta (p. 6), é que "[a] dificuldade está no fato de que quando passamos para além da aritmética, indo à análise ou à teoria dos conjuntos, o acréscimo substancial de recursos expressivos levanta a possibilidade de existirem sentenças que não são decidíveis em *nenhum* nível". HC é uma candidata bastante atraente, tanto por sua simplicidade (em relação às outras) quanto por seu histórico; aventada por Cantor, ela é tão antiga quanto a própria teoria dos conjuntos em que se insere. Sua indecidibilidade em ZFC, como já apontamos, não é fruto dos teoremas de incompletude — e, mais especificamente, não pode ser alcançada por métodos tão corriqueiros como a diagonalização. Ademais, tal prova de independência levou décadas para ser alcançada (a primeira parte tendo sido demonstrada por Gödel em 1938, e a segunda, por Paul Cohen em 1963; mesmo versando sobre o mesmo objeto de estudo, os métodos utilizados por cada um são bastante diferentes entre si). Também por esses motivos é que ela se apresenta como uma postulante tão forte à insolubilidade absoluta. Entretanto — e *infelizmente*, para aquele que vos escreve —, uma discussão sobre provas de independência em teorias dos conjuntos configuraria uma fuga total de nosso escopo.[34] Cabe ressaltar, todavia, que as diferenças entre as sentenças apresentadas neste parágrafo e as que construímos anteriormente não ficam restritas a seus níveis de complexidade.[35] Em particular, assumindo que ZFC é consistente, a indecidibilidade de HC *não implica* que seu valor lógico é verdadeiro. Este ponto nos faz voltar à discussão sobre as (más-)interpretações dos teoremas de incompletude.

Ao longo deste capítulo, por várias vezes buscamos enfatizar a importância de certos passos nas construções e argumentações que levam aos Teoremas — como

[32] Como Gödel notou em sua fala acima (no comentário entre colchetes), a afirmação que ele propõe não é uma disjunção exclusiva.

[33] Cf. (KOELLNER, 2006).

[34] O leitor interessado encontra maravilhosas referências nos (já clássicos) textos de (SMULLYAN; FITTING, 1996) e (JECH, 2014).

[35] Aqui, leia-se "complexidade" no sentido de que as teorias ZFC e de análise real têm uma capacidade expressiva muito maior do que PA.

quando apresentamos formas de se aritmetizar a sintaxe de PA, ou quando enunciamos o lema diagonal, que generaliza a ideia de autorreferência. Ainda assim vale salientar, mais uma vez, como a hipótese de *consistência* é essencial para todo o argumento: sem levá-la em conta, damos margem a interpretações errôneas do tipo "Os teoremas de incompletude nos mostram que *existem sentenças verdadeiras* que nunca poderão ser demonstradas nem refutadas". Mas veja que não há garantia alguma de que a indecidibilidade de uma sentença (digamos, em PA) implica que seu valor lógico é verdadeiro na interpretação padrão — ou seja, a menos que *saibamos* que PA é consistente, não podemos afirmar que Ɔ é verdadeira. A rigor, a prova do Segundo Teorema apenas nos diz que $\mathrm{Con}_{\mathrm{PA}}$ é equivalente a Ɔ (ou a qualquer ponto fixo de $\neg$**Bew**), do que segue que Ɔ é verdadeira *se e somente se* $\mathrm{Con}_{\mathrm{PA}}$ o for. Ocorre que há teorias que sabemos ser consistentes (por métodos que extrapolam os da teoria em questão), e outras para as quais ainda não encontramos uma resposta; para as primeiras, podemos dizer que sim, os teoremas de incompletude garantem a veracidade das sentenças indecidíveis especificadas.

No sentido inverso, a consistência de uma teoria *não garante* que sentenças falsas não possam ser demonstradas. Por exemplo, se supusermos PA consistente, ao adicionarmos $\neg\mathrm{Con}_{\mathrm{PA}}$ como um axioma, a extensão resultante demonstra trivialmente que PA é *in*consistente (pois é exatamente isso que $\neg\mathrm{Con}_{\mathrm{PA}}$ diz). Portanto, tal extensão também deve ser inconsistente, e como $\mathrm{Con}_{\mathrm{PA}}$ é indecidível em PA, continuará sendo-lo na teoria estendida. Logo, essa extensão de PA também deve permanecer consistente (dada a hipótese de que PA é consistente, e observando que $\neg\mathrm{Con}_{\mathrm{PA}}$ e $\mathrm{Con}_{\mathrm{PA}}$ *não são*, ao mesmo tempo, demonstráveis nessa extensão). Desse modo, $\neg\mathrm{Con}_{\mathrm{PA}}$ é um "teorema" falso dessa extensão.

Dito isso, tal má-interpretação também é uma espécie de extrapolação dos Teoremas. Dessa vez, no sentido técnico, pois o que se vê é uma *omissão* da hipótese de consistência, em favorecimento do desejo (inerente a todo matemático) de que suas teorias nunca estabeleçam contradições. Em suma, trata-se de um desentendimento bastante justificável — até porque teorias inconsistentes "provam qualquer coisa", logo são totalmente desinteressantes.

Voltemos nossa atenção, portanto, às questões que *de fato* se seguem dos teoremas de incompletude. Pela discussão anterior, Ɔ é um exemplo de sentença que, *mesmo que seja verdadeira*, não pode ser decidida na aritmética de Peano. Assim, sob a hipótese de consistência de PA, existem sentenças verdadeiras que nunca poderão ser demonstradas (e que, ainda bem, tampouco serão refutadas). Isso parece nos indicar uma lacuna entre o que é "verdade" e o que "pode ser demonstrado" numa teoria: por um lado, há sentenças verdadeiras que nunca receberão o status de "teorema"; por outro, nada garante que dentre seus teoremas não haja afirmações falsas. Ora, por meio da aritmetização da sintaxe, fomos capazes de construir uma fórmula

que expressa a *demonstrabilidade* de qualquer sentença; será que é possível fazer o mesmo para o conceito metateórico de *veracidade* de uma sentença? A presente discussão parece apontar para uma resposta negativa.

Vejamos que, com efeito, a "verdade" não pode ser definida em termos aritméticos.[36] Se realmente existisse uma função recursiva que diz que determinado número natural é o número de Gödel de uma sentença aritmética verdadeira — digamos que representada pela fórmula $\mathbf{Tr}(x)$ na linguagem de PA — , então com certeza poderíamos usar a fórmula $\neg\mathbf{Tr}(x)$ para exprimir a propriedade de algo *não ser* uma sentença aritmética verdadeira. Como toda fórmula em PA tem um ponto fixo, disso seguiria que

$$\mathsf{PA} \vdash \varphi \leftrightarrow \neg\mathbf{Tr}(\overline{\ulcorner\varphi\urcorner}),$$

para alguma fórmula φ na linguagem de PA. Mas, segundo nosso conceito restrito de "verdade" (cf. nota de rodapé 36, abaixo), dizer que φ é verdadeira é o mesmo que afirmar a própria φ.[37] Em símbolos,

$$\mathsf{PA} \vdash \varphi \leftrightarrow \mathbf{Tr}(\overline{\ulcorner\varphi\urcorner}).$$

Juntando com o que obtivemos anteriormente, chegamos a

$$\mathsf{PA} \vdash \mathbf{Tr}(\overline{\ulcorner\varphi\urcorner}) \leftrightarrow \neg\mathbf{Tr}(\overline{\ulcorner\varphi\urcorner}),$$

o que, sob a hipótese de consistência de PA, é completamente absurdo. Com isso, fica provado o *Teorema da Indefinibilidade da Verdade*, de Tarski[38] (1933):

Teorema 3.4.6. *Não existe fórmula alguma na linguagem de* PA *que represente o conceito de verdade em tal teoria.*

Perceba que, em semelhança a boa parte dos últimos argumentos que expusemos, a prova do teorema acima se baseia na existência de pontos fixos, dada pelo lema diagonal. Uma outra consequência deste resultado surge da indagação sobre a

[36] Perceba que, aqui, estamos tratando o conceito de "verdade" pelo ponto de vista da prática matemática comum — i.e., no sentido encontrado na ideia de que uma sentença é verdadeira se e somente se, quando *interpretada* como um afirmação sobre números naturais, obtiver o valor lógico "verdadeiro" em tal modelo. Em particular, dizer que a conjectura de Goldbach é verdadeira é apenas uma outra forma de afirmar seu próprio conteúdo, i.e., que "todo número par maior do que dois pode ser escrito como a soma de dois primos". Ou seja, em nossa discussão, não estamos interessados nos aspectos metafísicos do conceito de verdade — e tal abordagem inevitavelmente tocaria em pontos que a maioria dos matemáticos comuns preferem evitar. Talvez por uma espécie de instinto de sobrevivência intelectual, deixar de discutir a "real existência" dos objetos abstratos com que trabalhamos é uma maneira que encontramos para continuar fazendo matemática.

[37] Ou seja, se tal propriedade existir, então *toda* fórmula é um ponto fixo de $\mathbf{Tr}$.

[38] Alfred Tarski (1901-1983), um dos maiores lógico-matemáticos do séc. XX, ao lado de Gödel, Church e Turing (os quais já receberam os devidos reconhecimentos neste trabalho.) Seu nome ressoa no senso comum matemático em virtude do Paradoxo de Banach-Tarski, um resultado deveras contra-intuitivo, e que se segue do Axioma da Escolha. Além da indefinibilidade da verdade, podemos elencar como seu resultado mais expressivo (em nosso contexto) a decidibilidade da geometria clássica.

decidibilidade de uma fórmula que diz, sobre si mesma, *ser demonstrável em* PA (em contraste com $\mathfrak{D}$, que diz o oposto). Tecnicamente, queremos saber se um ponto fixo $H \leftrightarrow \mathbf{Bew}(H)$ é demonstrável ou refutável (ou nenhuma dessas alternativas) em PA.[39]

Intuitivamente, pelo conteúdo de tal sentença, é natural que esperemos que $\text{PA} \vdash H$, pois, por definição, ou ela é verdadeira e demonstrável em PA, ou é falsa e refutável. No entanto, uma rápida análise não consegue desvendar muita coisa: se for o caso de H ser demonstrável, então a condição (D1) dos predicados de demonstrabilidade nos diz que $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner H \urcorner})$; a mesma conclusão poder ser obtida via *modus ponens* aplicada a $\text{PA} \vdash H \leftrightarrow \mathbf{Bew}(H)$ e $\text{PA} \vdash H$. Assim, a hipótese de demonstrabilidade de H não nos leva a uma contradição (como a hipótese de demonstrabilidade de $\mathfrak{D}$ o faz), e não podemos concluir que H não é demonstrável. (Obviamente, isso *não prova* que H *é* demonstrável em PA.) A suposição contrária, de que H é refutável no sistema, tampouco nos leva a conclusões absurdas, de modo que uma prova direta sobre a decidibilidade de H não parece ser possível.

De fato, o argumento para a demonstrabilidade de H (já revelando o *spoiler* de que ela realmente é derivável no sistema) passa por um resultado muito mais geral; em particular, o que chamamos de *Teorema de Löb* implica no próprio Segundo Teorema de Incompletude (como veremos adiante):

Teorema 3.4.7. *Dada qualquer sentença H na linguagem de* PA*, se* $\text{PA} \vdash \boldsymbol{Bew}(\overline{\ulcorner H \urcorner}) \rightarrow H$*, então* $\text{PA} \vdash H$*.*

Ou seja, para deduzirmos a demonstrabilidade de H, basta que, sob a hipótese de H ser demonstrável, consigamos provar H dentro do sistema. Embora tal teorema possua uma prova elementar e direta, seu caráter técnico e puramente simbólico pode ser bastante nebuloso para uma primeira leitura, de modo que o leitor, em geral, não consegue captar prontamente a essência do raciocínio. Por conta disso, é comum introduzi-lo na forma de uma anedota lógica — um tanto quanto lúdica —, que basicamente nos mostra que Papai Noel existe. Considere, então, a sentença autorreferente *PN*, escrita na linguagem natural do português brasileiro:

> "Se *PN* é verdadeira, então Papai Noel existe".

Assumindo que *PN* é verdadeira, de sua definição segue que

> "Se *PN* é verdadeira, então Papai Noel existe" é verdadeira,

e por *modus ponens* temos imediatamente que

> "Papai Noel existe" é verdadeira.

[39] A escolha pela letra H se dá em alusão a Leon Henkin (1921-2006), lógico estadunidense que primeiro questionou sobre a decidibilidade de uma sentença que afirma sua própria demonstrabilidade, em 1952.

Assim, da hipótese que *PN* é verdadeira, fomos capazes de concluir que Papai Noel existe. Em outras palavras, fica deduzido que

"Se *PN* é verdadeira, então Papai Noel existe",

que é justamente o que a sentença *PN* afirma. Logo, *PN* deve ser verdadeira, e, por mais absurdo que possa parecer, Papai Noel com certeza existe!

O argumento de Löb é análogo ao que acabamos de derivar; trocando o predicado "é verdadeira" por "é demonstrável" e "Papai Noel existe" por qualquer letra sentencial que desejarmos[40] (digamos, A), vemos que é possível derivar A apenas sob a hipótese de demonstrabilidade da sentença que diz que a demonstrabilidade de A implica A (já que, em paralelo, derivamos a existência de Papai Noel apenas sob a hipótese de veracidade da sentença que diz, sobre si mesma, que sua veracidade implica na existência do bom velhinho[41]). Vejamos, de maneira formal, como chegar a esse resultado:

Demonstração do teorema 3.4.7. Sejam H qualquer sentença na linguagem de PA e $\mathbf{Bew}(x)$ o predicado de demonstrabilidade que definimos na seção anterior. Suponha que $\mathrm{PA} \vdash \mathbf{Bew}(\overline{\ulcorner H \urcorner}) \to H$, e faça $\mathbf{D}(y)$ ser a seguinte fórmula:

$$\mathbf{Bew}(y) \to H.$$

Aplicando o lema diagonal a $\mathbf{D}$, tem-se φ na linguagem de PA tal que

$$\begin{aligned} \mathrm{PA} &\vdash \varphi \leftrightarrow [\mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \to H] \\ &\vdash \varphi \to [\mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \to H]. \end{aligned}$$

Pelas propriedades (D1) e (D3) dos predicados de demonstrabilidade, obtemos:

$$\mathrm{PA} \vdash \mathbf{Bew}\{\overline{\ulcorner \varphi \to [\mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \to H]\urcorner}\}$$

$$\mathrm{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \varphi \to [\mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \to H]\urcorner}) \to [\mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \to \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \to H\urcorner})].$$

Juntando as duas, por *modus ponens* vem que

[40] Perceba como não há nada de especial com a escolha pela existência de Papai Noel: poderíamos ter optado por "Se essa sentença é verdadeira, então o leitor é a pessoa mais bonita do mundo", e o mesmo raciocínio nos levaria à constatação de que o leitor é realmente a pessoa mais bonita do mundo. Mais do que isso, poderíamos substituir "Papai Noel existe" por sua negação, e chegaríamos à conclusão de que Papai Noel *não* existe. Assim, independentemente do fato de crermos ou não na entidade máxima do espírito natalino capitalista, seríamos levados a uma contradição. Na literatura, isso leva o nome de *Paradoxo de Curry*, um reminiscente do Paradoxo de Russell.

[41] É importante enfatizar que o paralelo entre a anedota lógica natalina e o Teorema de Löb não passa de um artifício pedagógico. Em particular, diferentemente do que ocorre na linguagem natural, seria impossível derivar um paradoxo do tipo de Curry (vide nota anterior) na linguagem de PA, pois *não há* algo que possamos chamar de "predicado da verdade" na Aritmética de Peano, como acabamos de ver no Teorema de Tarski.

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow H \urcorner})$$

Agora, novamente por (D3), temos que

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow H \urcorner}) \rightarrow [\mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\ulcorner \varphi \urcorner) \urcorner}) \rightarrow \mathbf{Bew}(\overline{\ulcorner H \urcorner})].$$

Por transitividade, estas duas últimas nos dão:

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow [\mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\ulcorner \varphi \urcorner) \urcorner}) \rightarrow \mathbf{Bew}(\overline{\ulcorner H \urcorner})].$$

Usando (D2), temos que

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow \mathbf{Bew}(\overline{\ulcorner \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \urcorner}),$$

e destas duas últimas conseguimos chegar, novamente por *modus ponens*, a

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow \mathbf{Bew}(\overline{\ulcorner H \urcorner}).$$

Juntando isso com a hipótese inicial, i.e., $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner H \urcorner}) \rightarrow H$, mais uma aplicação de *modus ponens* deve nos dar

$$\begin{aligned} \text{PA} &\vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow H \\ &\vdash \varphi, \end{aligned}$$

pois φ é ponto fixo de $\mathbf{D}(y) \leftrightarrow [\mathbf{Bew}(y) \rightarrow H]$. Finalmente, uma última aplicação de (D1), juntamente de *modus ponens* com o fato de que $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \rightarrow H$, nos dá

$$\begin{aligned} \text{PA} &\vdash \mathbf{Bew}(\overline{\ulcorner \varphi \urcorner}) \\ &\vdash H, \end{aligned}$$

e a prova está completa. □

Como $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner H \urcorner}) \leftrightarrow H$ implica $\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner H \urcorner}) \rightarrow H$, segue-se imediatamente do Teorema de Löb que a sentença de Henkin é, de fato, um teorema de PA (bem como qualquer ponto fixo da fórmula representativa do predicado de demonstrabilidade). Ademais, se supusermos que PA é consistente e que $\text{PA} \vdash \neg\mathbf{Bew}(\overline{\ulcorner 0 = \mathbf{1} \urcorner})$, i.e., que PA prova sua própria consistência, então do Princípio da Explosão vem que

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner 0 = \mathbf{1} \urcorner}) \rightarrow A,$$

qualquer que seja a sentença A. Em particular,

$$\text{PA} \vdash \mathbf{Bew}(\overline{\ulcorner 0 = \mathbf{1} \urcorner}) \rightarrow (0 = \mathbf{1}),$$

e pelo Teorema de Löb temos que $\text{PA} \vdash (0 = \mathbf{1})$. Mas isso contradiz a hipótese de que PA é consistente, e daí se segue o Segundo Teorema de Incompletude.

Como apontado anteriormente, a *in*definibilidade da verdade, em conjunto com a definibilidade da demonstrabilidade, nos impedem de afirmar que toda sentença aritmética verdadeira possui uma prova. Por outro lado, a não ser que um sistema

seja presumidamente *correto*, não podemos dizer que *todos* seus teoremas são verdadeiros (i.e., que demonstrabilidade implica veracidade). Ora, mas em suma, o que o teorema 3.4.7 nos diz é que para determinarmos a demonstrabilidade de alguma sentença, basta que mostremos que ela é verdadeira sempre que for demonstrável. Assim, a hipótese do Teorema de Löb pode ser vista como uma sentença reflexiva de PA, que para cada fórmula fechada φ afirma que "se φ é demonstrável em PA, então φ é verdadeira".[42] Para sistemas que não são presumidamente corretos, embora não possamos concluir que toda sentença demonstrável é verdadeira — i.e., que toda instância do princípio de reflexão acima é demonstrável —, ao menos podemos dizer que se para cada instância dessas isso for verdade, então cada φ dessas é, na verdade, um teorema de PA. A contrapositiva do Teorema de Löb, i.e.,

$$\text{"se PA} \nvdash H\text{, então PA} \nvdash \mathbf{Bew}(\overline{\ulcorner H \urcorner}) \rightarrow H\text{"}$$

ilustra de maneira mais clara esse aspecto — só *não irão* satisfazer o princípio de reflexão aquelas sentenças que *não são* demonstráveis.

Antes de finalizarmos este capítulo, cabem ainda alguns comentários sobre certa característica marcante dos Teoremas de Incompletude — evidenciada pelo Teorema de Löb, mas que gera discussões desde a prova original de Gödel. Como vimos, não obstante a gama de construções teóricas interessantes de que tivemos que lançar mão para chegar aos Teoremas, parece ser unânime a concordância de que o passo intermediário, de construir a função recursiva Bew(x), é o ponto-chave da argumentação (ao menos quando estamos tratando de provas de incompletude que usam autorreferência[43]).

Entretanto, o predicado de demonstrabilidade não é peculiar somente pelo fato (técnico) de abstrair, em termos recursivos e aritméticos, a noção primordial da Teoria de Prova; como já apontamos, tal peculiaridade é o que mune um sistema formal da capacidade de *refletir* sobre suas próprias deduções, da mesma forma que um ser humano consciente racionaliza sobre as coisas em que acredita (o que fica ainda mais evidente na hipótese do Teorema de Löb). Por conta disso, outro aspecto interessante de Bew(x) é que ela pode ser vista como uma *modalidade* (ou um *operador modal*). [No âmbito da lógica modal, o que um tal operador faz é *qualificar* o valor de verdade de uma sentença (de maneira análoga ao que um verbo modal auxiliar faz com o verbo principal, em linguagens naturais que os admitem). As modalidades mais emblemáticas são as de necessidade (algo ser "necessariamente" verdadeiro) e de possibilidade (algo ser "possivelmente" verdadeiro), cujos signos são a caixa $\Box$ e o losango $\Diamond$,

[42] Chamamos essa forma geral de sentença de um *princípio de reflexão*, justamente porque elas indicam uma ponderação de PA sobre as coisas que ele mesmo pode derivar.

[43] O protagonismo de sentenças autorreferentes (i.e., que precisam fazer uso da diagonalização), embora bastante comum, não é um passo *crucial* em provas de incompletude. O leitor pode encontrar uma demonstração alternativa (devido a George Boolos, 1989), via noções algorítmicas, em (MURAWSKI, 1999), p. 155-156.

respectivamente. Mas há diversas outras, como temporalidade e condicionalidade — ou a própria demonstrabilidade, que estamos discutindo.]

Nesse sentido, fazendo as devidas correlações entre sentenças aritméticas (i.e., escritas na linguagem de PA) e sentenças modais (i.e., escritas na linguagem do sistema modal em questão), podemos interpretar Bew(x) como $\Box A$, em que A é a tradução, na linguagem da lógica de demonstrabilidade, da sentença cujo número de Gödel é x. Em particular, o Teorema de Löb pode ser parafraseado na seguinte forma:

$$\Box(\Box A \to A) \to \Box A,$$

qualquer que seja a sentença A. Tal esquema é tomado como um *axioma* da lógica de demonstrabilidade, e por essa razão esse sistema modal é comumente chamado de GL, em alusão a Gödel e Löb. É possível mostrar, dada uma interpretação adequada dos elementos de PA em GL, que este sistema modal é completo no sentido aritmético — i.e., que GL demonstra tudo que PA demonstra, e vice-versa. Por respeito ao escopo de nossa tese, não despenderemos os comentários necessários para uma plena compreensão deste assunto — nosso interesse está mais relacionado às interpretações e aplicações *puramente matemáticas* dos teoremas de incompletude, de modo que um estudo da demonstrabilidade, vista como operador modal, não nos levaria ao ponto que queremos chegar sem que tergiversássemos.[44] Há em GL, entretanto, um resultado de grande interesse aritmético (mesmo que formulado e deduzido em termos puramente lógicos), e que fazemos questão de mencionar.

Conforme já apontamos, há provas de incompletude que não dependem de autorreferência. O *teorema do ponto fixo* da lógica de demonstrabilidade estabelece essa verdade em termos puramente modais. Digamos que uma fórmula φ de GL é p-modalizada se toda ocorrência da variável p em φ ocorre dentro do escopo de um operador de demonstrabilidade — e.g., $\varphi = \neg\Box p$ ou $\varphi = \Box(p \to q)$. O teorema do ponto fixo então se lê:

Teorema 3.4.8. *Seja φ um fórmula p-modalizada qualquer de* GL*. Então, existe uma fórmula ψ em que p não ocorre em lugar algum, e cujas outras variáveis ocorrem todas também em φ, tal que*

$$\mathrm{GL} \vdash \psi \leftrightarrow \varphi(\psi).$$

Chamamos ψ de o ponto fixo de φ. Ademais, este ponto fixo é único.

Perceba como isso realmente mostra que é *desnecessária* a condição — pressuposta, pelo senso comum — de autorreferência. De fato, o leitor pode ser levado

[44] Todavia, o leitor interessado pode encontrar ótimas referências de lógica modal (e, em particular, de lógica de demonstrabilidade) em (HUGHES; CRESSWELL, 1996) e (BOOLOS, G. S., 2008)

à descrença quanto a tal afirmação, pois "$\psi \leftrightarrow \varphi(\psi)$" claramente nos diz que o conteúdo de ψ é "ψ possui a propriedade φ", o que configura autorreferência. No entanto, nenhum argumento ou construção (sintáticos ou semânticos) que pressuponham autorreferência são utilizados para derivar este teorema; a única ferramenta de que precisamos lançar mão é a *diagonalização* — e, como já havíamos notado anteriormente, uma sentença pode ser uma diagonalização sem ser autorreferente.

Em geral, as provas para este teorema nos indicam um algoritmo para computar tais pontos. Por exemplo, para $\varphi = \neg\Box p$, temos que $\psi = \neg\Box\bot$ (em que $\bot$ é o "conectivo" 0-ário que representa uma contradição qualquer). Assim, pelo teorema acima, obtemos:

$$\mathsf{GL} \vdash \neg\Box\bot \leftrightarrow \neg\Box(\neg\Box\bot).$$

Considerando apenas a "ida" desta equivalência, vemos que dada uma sentença que afirma sua própria não-demonstrabilidade, conseguimos obter uma sentença que afirma a consistência de PA (i.e., $\psi = \neg\Box\bot$), mas que não pode ser demonstrada. Isso é justamente o Segundo Teorema de Incompletude. Há outras instâncias do teorema do ponto fixo que são bastante interessantes;[45] no entanto, para nossos propósitos, já temos o suficiente.

[45] Cf. (BOOLOS, G. S., 2008), p. 105.

4 INCOMPLETUDE CONCRETA

Todo o aparato teórico que viemos construindo desde o início deste trabalho — especialmente nos capítulos 2 e 3, onde o rigor formal se mostrou mais presente — teve sua razão de ser na vontade do autor de explanar a essência, o contexto e o método que permeiam o conceito de incompletude na teoria de números. Nesse sentido, embora os teoremas de Gödel sejam aplicáveis a qualquer sistema formal que abarque uma porção mínima de aritmética[1], a abordagem de tais resultados no caso particular de PA, além de ser de simples compreensão, parece ser a mais próxima da realidade prática da maioria dos matemáticos. (Afinal, este é um trabalho destinado a leigos de exatas e a matemáticos comuns — i.e., graduandos, pós-graduandos e pesquisadores que se interessam por objetos matemáticos usuais — e não somente ao nicho fundamentalista, que inclui lógicos e conjuntistas, e que é o local onde questões como a incompletude surgem mais naturalmente.)

Isto posto, podemos dar como concluída a tarefa de desmistificar os Teoremas de Gödel; em particular, Ð, Я e $\mathrm{Con}_{\mathrm{PA}}$, que são as maiores fontes de desentendimento quando num primeiro contato com os Teoremas, agora têm significados matemáticos bastante precisos. No entanto, nosso segundo (e principal) objetivo ainda carece tanto de forma quanto de conteúdo.

Como nos endereçamos focadamente ao público matemático comum, o primeiro ponto a ser discutido é o fato de que Ð, Я e $\mathrm{Con}_{\mathrm{PA}}$, embora formuladas na linguagem de PA, possuem significados *metalinguísticos* — i.e., versam *sobre* o sistema em questão. Assim, argumenta-se que tais sentenças não surgem naturalmente nas áreas mais gerais do pensamento matemático — ou, de maneira mais jocosa, diz-se que estas são frutos de meros “truques lógicos”. Não é obra do acaso, portanto, que os teoremas de incompletude venham sendo negligenciados por grande parte da comunidade matemática. No fim das contas, tudo que o matemático comum deseja é (i) que suas teorias não sejam inconsistentes nem redundantes, e (ii) que suas conjecturas possam ser provadas (ou, no mínimo, refutadas). A indecidibilidade da consistência não resolve o item (i), ao passo que a falta de exemplos de sentenças indecidíveis não-metateóricas não sugere a necessidade de receio de que suas conjecturas — não-metateóricas por natureza — nunca possam ser decididas.

O objetivo deste capítulo é mostrar que, de fato, *existem* motivos para tais receios. Isso ocorre justamente porque, embora o senso comum ainda não tenha absorvido as informações necessárias para nutrir esse tipo de medo, há *diversas* sentenças indecidíveis que ocorrem naturalmente na Matemática (i.e., fora do nicho lógico-matemático). Nosso foco, no entanto, dar-se-á nos teoremas de Paris-Harrington (PH), Paris-Kirby (PK) e Harvey Friedman (HF), já que eles estão inseridos no contexto

[1] Ou seja, qualquer sistema formal que estenda a aritmética minimal, Q, mencionada anteriormente.

da aritmética de Peano, eixo de nosso estudo.[2]

A bem da verdade, e tentando não ser leviano, tais exemplos só começaram a brotar no final da década de 1970, sendo o primeiro deles PH, de 1977 — ou seja, mais de quarenta anos após a prova de Gödel. Isso, aliado ao fato de que nenhum desses exemplos de indecidibilidade é uma conjectura famosa (como a de Goldbach, ou a hipótese de Riemann), talvez justifique a persistente indiferença com que os Teoremas são tratados.

O segundo ponto de discussão é mais sutil, e deve ser nosso ponto de partida; antes de exemplificar e esboçar provas de sentenças indecidíveis que "ocorrem naturalmente nas áreas mais comuns da Matemática", ou, ainda, que "sejam puramente matemáticas", precisamos entender o que isso quer dizer. Já de início, surge algo peculiar, pois tal indagação pode ser circular e contraditória: se é possível fazer este tipo de pergunta, então o que podemos dizer sobre uma *definição puramente matemática* do conceito "puramente matemático"? A questão parece mais filosófica do que matemática, e abre caminho para a intuição e a subjetividade.

4.1 MATEMÁTICA CONCRETA VS. MATEMÁTICA ABSTRATA

Antes de mais nada, mesmo o título da presente seção já se mostra controverso; afinal, não seria a própria Matemática, em sua totalidade, uma área *intrinsecamente abstrata* do conhecimento? Sob esse ponto de vista, parece infrutífera uma distinção entre sentenças/teorias matematicamente concretas e abstratas. Mas a verdade, como já apontamos, é que *algumas* sentenças (como $\mathfrak{Con}$ e $\mathrm{Con}_{\mathsf{PA}}$) e *certas* teorias (como a metamatemática) diferem substancialmente das sentenças e teorias comuns do pensamento matemático. Dessa forma, não há dúvidas quanto à classificação destas: $\mathfrak{Con}$ e $\mathrm{Con}_{\mathsf{PA}}$ são sentenças matemáticas abstratas. O problema surge nos casos limítrofes.

O melhor exemplo é a teoria dos conjuntos (que, aqui, tomamos como sendo ZFC ou ZF, i.e., a axiomática de Zermelo-Fraenkel com ou sem o axioma da escolha, respectivamente). Seu status é, ao mesmo tempo, fundacional e puramente matemático (logo, abstrato e concreto), pois tanto descreve uma base para as mais variadas teorias, quanto tem seus significados matemáticos próprios. Colocando de maneira ilustrativa: grande parte dos objetos matemáticos mais usuais podem ser descritos em termos de conjuntos (ou classes, ou categorias); por outro lado, o conceito abstrato de conjunto — e toda estrutura possivelmente derivável de tal conceito — é, ele próprio, considerado um objeto matemático.

Nesse sentido, podemos tomar a hipótese do *continuum* (HC) como um desses casos limítrofes. Sua formulação, em linguagem natural, é "todo subconjunto infinito de números reais está em bijeção ou com o conjunto dos inteiros ou com o próprio

[2] Mesmo assim, ainda teremos algum espaço para discutir a indecidibilidade de sentenças alheias a PA, como a hipótese do *continuum* e o axioma da escolha.

conjunto dos reais". Ora, conjuntos numéricos e suas cardinalidades são assuntos recorrentes em *diversas* áreas da Matemática, seja em definições, axiomas ou hipóteses de teoremas; além disso, HC claramente não fala *sobre* ZFC — nem de coisas que podem ocorrer *dentro* da teoria dos conjuntos, como "algo ser demonstrável em ZFC", ou "ZFC ser consistente" —, logo tal sentença não compartilha a "artificialidade" presente nos exemplos metalinguísticos, como $\mathfrak{Con}$ e $\mathrm{Con}_{\mathrm{ZFC}}$. Entretanto, há pouca concordância quanto à *naturalidade* com que HC surge nas teorias matemáticas. Harvey Friedman aponta[3] que tal discordância se dá não somente pelo fato de que HC tem poucas aplicações diretas fora da teoria dos conjuntos. Mais do que isso, seu nível de generalidade é alto demais se comparado às sentenças que comumente preocupam os matemáticos em suas pesquisas. Em outras palavras, HC está mais próxima das sentenças metateóricas que vimos anteriormente do que das conjecturas de interesse matemático usual.

Até aqui, tivemos pouco sucesso na tentativa de distinção entre Matemática Concreta e Matemática Abstrata. No entanto, podemos ao menos concordar num ponto: sentenças metalinguísticas (como $\mathfrak{Con}$) e de menor aplicabilidade fora dos fundamentos da Matemática (i.e., da lógica e da teoria dos conjuntos, como HC) se enquadram como abstratas. Em busca de uma definição menos subjetiva, Friedman estabelece a ideia de concretude através de *medidas de Borel*, um conceito comum na topologia.[4] Mais formalmente, diz-se que

Definição 4.1.1. Uma sentença é *matematicamente concreta* se e somente se é uma sentença concernente a conjuntos Borel-mensuráveis, ou a funções de posto finito entre espaços métricos completos separáveis.

Respeitando o escopo desse trabalho, não vamos definir — muito menos entrar em detalhes sobre — os conceitos topológicos presentes na definição acima. Isso porque, embora tal definição se baseie em conceitos próprios da topologia geral — que, provavelmente, não é capaz de descrever a Matemática como um todo —, existe um resultado da teoria de conjuntos descritiva[5] que mostra a equivalência entre a classe das funções de posto finito em espaços métricos completos separáveis e a classe das funções aritméticas descritas na teoria da recursão (que é a base para nossa estimada Teoria da Prova; esta, sim, devidamente esmiuçada em nosso trabalho). Outra motivação para a definição 4.1.1 está no fato, apontado por Friedman, de que a imensa maioria da atividade matemática comum está voltada aos conceitos introduzidos em

[3] O material a que nos referimos não foi publicado, mas um rascunho do que viria a ser tal monumental obra está disponível para acesso livre na página do autor: https://u.osu.edu/friedman.8/foundational-adventures/boolean-relation-theory-book/. O apontamento se encontra na página 12 do referido texto.

[4] *Grosso modo*, uma medida de Borel quantifica o *tamanho* de um conjunto, dentro de uma álgebra específica. Em $\mathbb{R}$, a medida usual é aquela que associa o intervalo $[a, b] \subset \mathbb{R}$ ao número real $b - a$.

[5] Cf. Friedman, p. 54.

tal definição, ou a níveis teóricos abaixo do que estes pressupõem. Em bom português: a concretude matemática, definida como acima, além de ser *realmente* uma definição matemática — pois há nada de subjetivo nela —, também está de acordo com a classe de relações aritméticas efetivamente decidíveis, objeto de estudo deste trabalho; mais ainda, tal definição consegue abarcar todo tipo de sentença que um matemático comum classificaria como *seu* objeto de estudo.

Neste ponto, portanto, vale notar como a própria definição 4.1.1, através da qual Friedman buscou descrever o fenômeno de incompletude concreta, se baseia na posição *limítrofe* que algumas questões de teoria dos conjuntos — HC e AC especialmente — ocupam em nossa tentativa de distinguir o que é concreto e o que é abstrato dentro da Matemática. Em última instância, o fato (i.e., o teorema afirmando) que *a hipótese do continuum para conjuntos Borel-mensuráveis é decidível em ZFC*[6] também motiva tal definição. Assim, segundo Friedman, a forma mais natural de se definir o que é "natural" (ou "puro", ou "usual") dentro da Matemática é restringir nosso campo de estudo a conjuntos e funções que tornam demonstráveis algumas sentenças de aplicabilidade notadamente "abstrata" (ou "não usual" sob o ponto de vista do matemático comum) na teoria dos conjuntos (e que são indecidíveis quando avaliadas para conjuntos quaisquer).

De toda maneira, questões inerentes à teoria dos conjuntos sempre foram alvos de calorosas discussões (principalmente no que se refere a AC). Assim, se o fato de HC e AC serem indecidíveis implica (ou não) que "existem sentenças de cunho 'puramente matemático' que não podem ser demonstradas", a resposta — qualquer que seja ela — abala menos as fundações da Matemática do que os próprios conteúdos de tais sentenças. A indecidibilidade de sentenças matematicamente concretas *na aritmética de Peano*, no entanto, chama bem mais a atenção. Enquanto ZFC tem um aspecto dual (i.e., concomitantemente fundacional e puramente matemático, como apontado acima) e só foi admitido mais recentemente no universo da Matemática (através de Cantor), a teoria de números, por sua vez, é muito mais elementar, remonta a milênios de desenvolvimento teórico bem estabelecido, e claramente fala apenas de números naturais (objetos mais específicos do que coleções quaisquer de indivíduos quaisquer). Como mostraremos a seguir, os exemplos que trazemos têm caráter *evidentemente* concreto, ocorrendo de maneira natural nas respectivas teorias. Ademais, eles se encaixam na definição formal de Friedman.

4.2 INCOMPLETUDE CONCRETA EM PA

A história que queremos contar teve seu marco inicial na publicação, em 1977, do artigo intitulado *A Mathematical Incompleteness in Peano Arithmetic*, de Jeff Pa-

[6] Cf. (HORTON *et al.*, 2003), disponível em http://www.units.miamioh.edu/sumsri/sumj/2003/Continuum1.pdf

ris e Leo Harrington. Nele, certa sentença da análise combinatória (PH) se mostra indecidível em PA, apesar de ser verdadeira (e demonstrável em teorias mais fortes que PA). O que se seguiu foi uma grande quantidade de provas de indecidibilidade de outras sentenças, em sua maioria também combinatórias, cujos interesses eram notoriamente comuns na atividade matemática em geral.

Pouco tempo depois, em 1982, Paris figurou novamente na vanguarda da incompletude aritmética concreta, quando, ao lado de Laurie Kirby, provou a indecidibilidade de uma versão finita do Teorema de Goodstein. Diferentemente dos casos anteriores, que versam sobre princípios combinatórios, tal sentença se dá na teoria elementar de números. Mais importante que isso, é um problema que pode ser formulado visualmente através de uma árvore (i.e., um *grafo* não orientado em que cada dois vértices estão conectados por um único caminho), e isso permite que sua exposição seja bastante lúdica e didática. Em suma, PK (como combinamos chamá-lo) é o exemplo perfeito para a constatação da incompletude concreta.

Finalmente, temos HF, o teorema de Friedman.[7] Ele pode ser visto como uma versão do Teorema de Kruskal, e é mais um exemplo que pode ser formulado usando árvores. No entanto, os métodos para obtenção de HF — mais especificamente, a função que avalia o crescimento da árvore em questão — ultrapassam os níveis de PK. Isso implica, entre outras coisas, o fato de que HF continua sendo indecidível mesmo em teorias que provam PK.

Apesar de haver muitos outros exemplos de incompletude concreta em PA, focaremos nossas atenção nesses três. Mesmo assim, não vamos nos aventurar na tarefa de argumentar rigorosamente suas provas, pois isso fugiria demais de nosso escopo, e temos um limite de espaço.[8] Nosso maior interesse está em dar as motivações para tais teoremas serem verdadeiros, e, principalmente, em evidenciar seus aspectos puramente matemáticos.

De toda forma, para melhor compreensão de PK e HF, um conhecimento basilar de teoria dos números ordinais é requerido. Àqueles com pouco ou nenhum entendimento sobre o assunto, recomendamos a leitura do Apêndice A no final dessas páginas, após as referências bibliográficas. Neste excerto em questão, também se encontra um tratamento resumido de ordinais *contáveis*, e, mais especificamente, as construções necessárias para se entender as funções de rápido crescimento que descrevem os conjuntos presentes em PH e HF.

---

[7] Vale notar a nomenclatura que estamos usando: quando chamamos PH, PK e HF de “teoremas”, estamos nos referindo aos resultados que provam a indecidibilidade das respectivas sentenças, e não às sentenças em si. Obviamente, as sentenças referentes a PH, PK e HF *não são teoremas* (em PA).

[8] O leitor interessado pode tentar acompanhá-las em sua totalidade nas fontes originais; as provas não são tão complexas.

4.2.1 PH: a total desordem é impossível

Como todo bom princípio da análise combinatória, PH trata de eventos randômicos. Neste e em muitos outros casos dessa área da matemática discreta, o objetivo é encontrar padrões, ou algum tipo de "ordem", em situações completamente aleatórias. PH, que é uma versão ligeiramente modificada de um teorema bastante geral, devido a Frank Ramsey (1903-1930), pode ser ilustrado através da organização de uma festa.

Suponhamos que o leitor (em tempos normais[9]) decida realizar um evento comemorativo — uma "celebração ao caos afetivo" —, e que, com essa temática em mente, convide pessoas quaisquer, escolhendo-as de maneira aleatória, mas sob a condição de que cada dois convidados devem ou se dar muito bem, ou se dar muito mal. O problema então se configura da seguinte forma: qual a quantidade (mínima) necessária de convidados para que se tenha certeza de que um subgrupo de certo tamanho contenha somente pessoas que se dão bem ou que se dão mal? O Teorema de Ramsey, em sua versão finita, vem nos dizer que, por exemplo, se nossa festa for um pequeno encontro de *seis* pessoas que, duas a duas, ou se amam ou se odeiam, então existe um subgrupo de *três* dessas pessoas, todas as quais ou se amam ou se odeiam. (Claramente, uma das rodas de conversa nessa festa tende a ser ou muito afetuosa ou muito conflituosa.)

Mas vejamos como formular o problema matematicamente. Dado A um conjunto finito qualquer (não necessariamente numérico), cuja cardinalidade é $\geq n$, chamamos $B \subset A$ de um *n-subconjunto de A* se e só se B possui exatamente n elementos. Em seguida, reunindo todos os n-subconjuntos de A no conjunto $[A]^n$, e dado um número natural $m \leq n$, podemos formar uma *m-partição* dessa reunião, de forma que cada $B \in [A]^n$ caia dentro de uma (e apenas uma) das classes $C_1, ..., C_m$. Por exemplo, se a cardinalidade de A é igual a n, então há apenas uma classe na partição de $[A]^n$, e apenas um B, de modo que tenhamos $B = A$ e $[A]^n = \{B\} = \{A\}$; se a cardinalidade for $n+1$, haverá $n+1$ classes, i.e., $m = n+1$, e assim por diante, seguindo um argumento combinatório.

Finalmente, dado $H \subset A$ qualquer, cuja cardinalidade é $\geq n$, dizemos que H é *homogêneo* para uma m-partição de $[A]^n$ quando existe uma classe C_i dessa partição tal que C_i tem como membros todos os n-subconjuntos de A cujos elementos estão todos em H. Podemos enxergar tal propriedade — "visualmente", nesse caso – sob o ponto de vista de uma *coloração*. De fato: em nosso exemplo, tomando o universo A como um conjunto que tem a quantidade necessária para satisfazer o problema (i.e., encontrar um grupo de três pessoas que ou se amam ou se odeiam), podemos pensar em "amar" como sendo, digamos, a cor azul, e "odiar" como sendo a cor azul. Assim, dado que dois convidados quaisquer ou se amam ou se odeiam, a ideia principal do problema será *colorir* cada grupo de duas pessoas com uma dessas duas cores, e um

[9] O presente trabalho foi inteiramente produzido em meio à pandemia de Covid-19.

subconjunto $H \subset A$ será homogêneo quando todas as duplas possíveis de convidados *que estão nesse subconjunto* tiverem a mesma coloração. O Teorema de Ramsey, então, se lê:

Teorema 4.2.1. *Sejam A um conjunto e $k, m, n \in \mathbb{N}$ quaisquer, com $k \geq n$. Então existe $p \in \mathbb{N}$, $p \geq n$, tal que se A possui ao menos p elementos, para cada m-partição de $[A]^n$ deve existir $H \subset A$ um conjunto homogêneo contendo ao menos k elementos.*

Traduzindo para o caso da nossa pequena festa em celebração ao caos, temos $p = 6$ (o número de convidados), $m = n = 2$ (pois os convidados muito se gostam ou muito se desgostam, e tal relação é binária), e $k = 3$ (i.e., numa festa com seis convidados, certamente há um subgrupo de 3 indivíduos que ou se odeiam ou se amam). Agora, para ver que $p = 6$ é, de fato, solução desse problema, primeiramente tome a_1 um dos convidados. Então, dos cinco restantes, há pelo menos três deles que ou o convidado a_1 ama ou odeia; chame-os de convidados a_2, a_3, a_4, e suponha (sem perda de generalidade) que a_1 ama todos eles. Haverá, portanto, dois casos: se (i) a_2 ama a_3 ou a_2 ama a_4 ou a_3 ama a_4, então ou a tripla (a_1, a_2, a_3) ou a tripla (a_1, a_2, a_4) ou a tripla (a_1, a_3, a_4) é tal que cada dois de seus membros se amam. Por outro lado, se (ii) a_2 *odeia* a_3 e a_2 odeia a_4 e, também, a_3 odeia a_4, então a tripla (a_2, a_3, a_4) é tal que cada dois de seus membros se odeiam, e temos o resultado.

Para $p = 18$ convidados, teríamos $k = 4$; i.e., para que sempre houvesse um grupo de quatro pessoas que nutrem ódio ou amor pelos outros três, precisaríamos chamar ao menos dezoito indivíduos. Para 5 convidados, é impossível tomar 3 deles sem que eles que formem um trio heterogêneo (i.e., tal que dois não se gostem entre si mas nutram afeto pelo terceiro, ou vice-versa). As figuras abaixo ilustram, por meio de grafos, que a solução $p = 6$ é realmente válida, pois existe um triângulo cujos lados têm a mesma cor (vermelha ou azul), e também, o fato de que $p = 5$ não pode ser solução do mesmo problema (pois não há um tal triângulo):

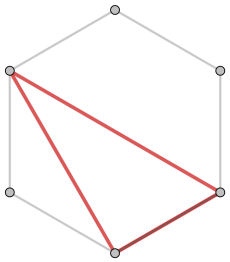


Figura 3 – Solução $p = 6$ para $k = 3$.

Mas a importância do Teorema de Ramsey não se restringe a *puzzles* com anedotas sobre festas. Em uma de suas instâncias, ele implica o princípio combinatório mais elementar de todos. Tomando $n = 1$, $k = 2$ e m arbitrário, então $p = m + 1$ nos dá

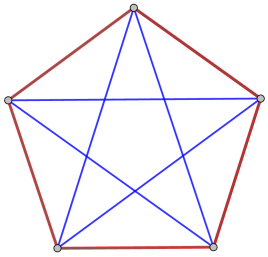


Figura 4 – Para $k = 3$, $p = 5$ não é solução.

exatamente o Princípio das Gavetas de Dirichlet (também conhecido como o *Princípio das Casas dos Pombos*). Com efeito: nessa configuração, o teorema 4.2.1 afirmaria que se um conjunto A com $m+1$ elementos for particionado em m classes, então uma dessas classes deve possuir ao menos dois elementos de A.

Agora, o teorema 4.2.1 é demonstrável em PA. Para se convencer disso, pensemos em termos mecânicos: como há apenas uma quantidade finita de n-subconjuntos de $A = \{0, ..., p-1\}$, finitas maneiras de particionar A em m classes, e finitos k-subconjuntos de $[A]^n$, tudo que um computador precisaria fazer para verificar a existência de A seria analisar cada partição de $[A]^n$ e procurar por um subconjunto homogêneo de cardinalidade k. Caso em alguma partição se encontrasse algum conjunto não-homogêneo, bastaria diminuir a cardinalidade de A para $p-1$, e prosseguir dessa forma até que se encontrasse tal k mínimo. [Os limites práticos desse tipo de rotina não nos importam, como é de praxe na teoria da recursão; todavia, a forma como a magnitude p aumenta tem papel central na justificativa para a indecidibilidade de PH, como veremos adiante.] Uma prova formal para o teorema 4.2.1, no entanto, teria de passar pela versão infinita do mesmo, aliada a um resultado combinatório conhecido como o Lema de König, e isso é algo que não cabe em nossos propósitos.[10]

O enunciado de PH surge de uma pequena modificação no Teorema de Ramsey. Dado $A \subset \mathbb{N}$ um conjunto finito, chamamos B de um *conjunto relativamente grande* se sua cardinalidade for maior ou igual a seu menor elemento. Assim, por exemplo, $\{1, 3\}$ é relativamente grande, ao passo que $\{100, 1001\}$ não o é. Com isso, trocando "H homogêneo" por "H relativamente grande e homogêneo", temos a sentença de Paris-Harrington:

PH. *Sejam A um conjunto e $k, m, n \in \mathbb{N}$ quaisquer, com $k \geq n$. Então existe $p \in \mathbb{N}$, $p \geq n$, tal que se A possui ao menos p elementos, para cada m-partição de $[A]^n$ deve existir $H \subset A$ um conjunto homogêneo relativamente grande contendo ao menos k elementos.*

Usando o mesmo método de prova indicado para o teorema 4.2.1, conclui-se facilmente que PH é um teorema. Ou seja, PH é decidível, desde que se trabalhe

[10] O leitor curioso pode encontrar essa prova em (BOOLOS, G. *et al.*, 2007), p. 320-325.

dentro de uma teoria capaz de expressar conceitos infinitários, i.e., onde as variáveis possam percorrer não somente indivíduos, mas também subconjuntos infinitos desse espaço infinito. Ocorre que a prova original de Ramsey para a versão finita de seu teorema não pressupunha sua versão infinita, de forma que o argumento indicado anteriormente é inválido no contexto do Programa de Hilbert, que impõe métodos finitários. Tal prova original, entretanto, é válida em PA. PH por sua vez, desconhece (e sempre irá desconhecer) uma prova na Aritmética de Peano, apesar de poder ser formulada em sua linguagem.[11]

O argumento exposto pelos autores em (HARRINGTON; PARIS, 1977) é, senão tecnicamente, ao menos essencialmente simples. Em primeiro lugar, constrói-se uma teoria T a partir de PA — adicionando, à sua linguagem, infinitas novas constantes $c_0, c_1, ...$, e estabelecendo axiomas específicos para elas (inclusive um novo princípio de indução); as regras de inferência e as constantes individuais originais, bem como as operações básicas, permanecem as mesmas. Como T possui entidades diferentes dos números naturais (que são obtidos de uma *única* constante, o zero, via função sucessor), dizemos que T possui um modelo não-*standard*. O segundo passo consiste em mostrar que $\mathsf{PA} \vdash \mathsf{Con}_T \rightarrow \mathsf{Con}_{\mathsf{PA}}$, o que pode ser facilmente obtido, uma vez que se mostre que o modelo de T também é um modelo de PA. Finalmente, a indecidibilidade de PH é obtida ao se provar que $\mathsf{PA} \vdash \mathsf{PH} \rightarrow \mathsf{Con}_T$, pois disso se segue que

$$\mathsf{PA} \vdash \mathsf{PH} \rightarrow \mathsf{Con}_{\mathsf{PA}},$$

e como a consistência de PA é indecidível em PA, PH também deve sê-lo.

Alguns dos passos intermediários dessa prova têm uma forte relação com funções de rápido crescimento. Mais especificamente, tais funções buscam descrever o tamanho do conjunto inicial pressuposto em PH (i.e., o número p de convidados, no caso da festa). Essa discussão também tem a ver com o comentários que fizemos sobre o Teorema de Ramsey ser decidível em PA, quando invocamos um computador capaz de verificar o tamanho de p para quaisquer $k, m, n \in \mathbb{N}$. O fato de p depender de k, m, n pode gerar conjuntos tão gigantescos que a função que os descrevem nem ao menos estão bem definidas em PA.

Gina Kolata e Craig Smorýnski, numa série de artigos sobre o trabalho de Friedman quanto à incompletude concreta, esboçam vários exemplos de funções que crescem muito rapidamente.[12] Como esse conceito também é central na indecidibilidade de PK e HF, será interessante mostrar de que forma se dá a velocidade com que essas funções crescem.

Como dissemos acima, o tamanho de p depende de k, m, n. No entanto, pode-se estabelecer um *limite inferior* para as funções que descrevem a variação desse

[11] C.f. (HÁJEK; PUDLÁK, 1998), p. 112.
[12] C.f. (HARRINGTON *et al.*, 1985).

tamanho. Esse limite, como era de se esperar, traduz-se numa função de crescimento absurdamente veloz. *Grosso modo*, pode-se dizer que tal função cresce tão rapidamente quanto a função de Ackermann (mencionada no capítulo 2, quando da discussão sobre funções que não são recursivas primitivas). Sua definição, por dupla recursão, é relembrada a seguir:

$$f(0, y) = y + 1;$$
$$f(\mathrm{s}(x), 0) = f(x, 1);$$
$$f(\mathrm{s}(x), \mathrm{s}(y)) = f(x, f(\mathrm{s}(x), y)).$$

Os primeiros valores de f são $f(0,0) = 1$, $f(1,0) = f(0,1) = 2$, $f(1,1) = 3$. Mas basta tomar os pares $(3,2)$ e $(3,4)$ para constatar sua explosiva taxa de crescimento; temos:

$$f(3,2) = 2^{2^2} = 2^4 = 16;$$
$$f(3,4) = 2^{2^{2^{2^2}}} = 2^{65536}.$$

Enquanto $f(3,2)$, e todos os valores anteriores a este, têm menos de três dígitos, $f(3,4)$ tem mais do que dezenove mil. Indo um pouco adiante, para se ter uma ideia, $f(6,6)$ é um valor tão gigantesco que seu número de dígitos ocuparia mais de uma página. A extrapolação para argumentos mais distantes da origem de $\mathbb{N}^2$ faz f atingir valores cada vez mais altos — tão altos que nosso intelecto parece não dar conta de acompanhar tal progressão. De certa maneira, essa sensação de impotência é compartilhada pelo próprio sistema PA; a função que descreve o tamanho p, diferentemente da função de Ackermann (que obviamente mora na linguagem de PA), cresce mais rapidamente do que *qualquer* função computável — inclusive f. Mas essa taxa de crescimento, como veremos adiante, é extremamente baixa se comparada à taxa de crescimento da função associada à sentença proposta por Friedman. Em partes, isso justifica o fato de PH ser decidível em extensões de PA nas quais HF continua sendo impossível de se demonstrar.

4.2.2 PK: simulando uma batalha hercúlea

A busca por incompletude concreta em PA teve seu apogeu em 1982, quando Jeff Paris e Laurie Kirby exibiram duas sentenças indecidíveis intimamente relacionadas entre si, no sentido de que os métodos combinatórios usados em ambas as provas são muito parecidos. A primeira delas é uma versão restrita do Teorema de Goodstein. Como tal resultado mora na teoria de números, ela passou a ser considerada o primeiro exemplo de sentença indecidível fora dos campos da metateoria e da análise combinatória (que, até então, vinha sendo a única área da Matemática a se mostrar concretamente incompleta).

A segunda sentença é combinatória por natureza. Seu enunciado envolve um jogo/disputa que se dá numa árvore que se ramifica muito rapidamente; tal jogo pode ser imaginado como a representação matemática da luta entre Hércules e a Hidra de Lerna.[13] Acontece que ambas as sentenças (que, a partir de agora, chamaremos de PK1 e PK2, respectivamente) podem ser formuladas na linguagem de PA, e como os argumentos que levam a suas indecidibilidades são quase os mesmos, costuma-se abordar apenas a segunda (até porque seu apelo visual e mitológico são bem mais interessantes do que o Teorema de Goodstein sozinho).

Mas vejamos, ao menos, como formular PK1. Primeiramente, dados $m, n \in \mathbb{N}$, com $n > 1$, definimos o que vem a ser a *representação de m na base n*, da seguinte forma iterada: na primeira iteração, escreva m como uma soma de potências únicas de n (i.e., cada número natural aparece como expoente de n no máximo uma vez, e o que é sempre possível fazer, dado o Teorema Fundamental da Aritmética). Por exemplo, se $m = 266$ e $n = 2$, temos $266 = 2^8 + 2^3 + 2^1$. O procedimento deve continuar até que a representação *estabilize*, i.e., até que todos os expoentes iniciais (e os expoentes dos expoentes iniciais, etc.) sejam escritos como uma soma de potências de n. Assim, a representação de $m = 266$ se torna estável em $m = 266 = 2^{2^{2+1}} + 2^{2+1} + 2^1$. (É claro que, por simplicidade, optamos por escrever 1 no lugar de 2^0.)

Em seguida, definimos o número $G_n(m)$ de maneira que $G_n(m) = 0$ se $m = 0$; se esse não for o caso, $G_n(m)$ será o resultado da substituição de cada n na representação de m em base n, como acima, por seu sucessor, $n + 1$, seguida de uma subtração do número 1. Em nosso exemplo recorrente, teremos $G_2(266) = 3^{3^{3+1}} + 3^{3+1} + 2$. Finalmente, a *sequência de Goodstein para m* será dada por:

$$m_0 = m,\ \ m_1 = G_2(m_0),\ \ m_2 = G_3(m_1),\ \ m_3 = G_4(m_2), ...$$

Como ilustração, temos os primeiros quatro termos da sequência de Goodstein para 266 (aqui, o símbolo $\approx$ significa "aproximadamente"):

$$\begin{aligned}
266_0 &= 266 = 2^{2^{2+1}} + 2^{2+1} + 2^1; \\
266_1 &= 3^{3^{3+1}} + 3^{3+1} + 2 \approx 10^{38}; \\
266_2 &= 4^{4^{4+1}} + 4^{4+1} + 1 \approx 10^{616}; \\
266_3 &= 5^{5^{5+1}} + 5^{5+1} \approx 10^{10^3}.
\end{aligned}$$

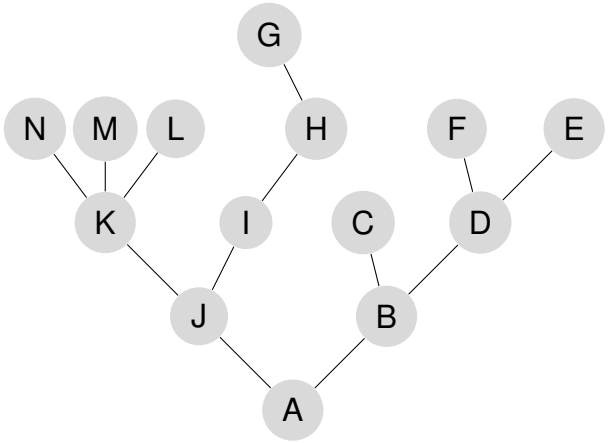
Perceba como os valores da sequência tendem a números exorbitantes. Nesse sentido, o Teorema de Goodstein afirma algo bastante contra-intuitivo:

[13] A mitologia grega nos conta que, para obter sua imortalidade, Hércules — o semi-deus filho de Zeus e Alcmena — deveria completar doze penosas tarefas para o rei de Micenas. Uma delas (talvez a mais emblemática da lenda) seria matar a hidra, uma espécie de monstro aquático com numerosas cabeças de serpente. Venenosa ao ponto de matar homens apenas com o odor de seu hálito, a hidra ainda contava com um absurdo poder de regeneração — se uma cabeça fosse cortada, novas cópias prontamente surgiriam.

Teorema 4.2.2. *Para quaisquer $m, n \in \mathbb{N}$, com $n > 1$, a sequência de Goodstein para m eventualmente chega a zero, i.e., existe $k \in \mathbb{N}$ tal que $m_k = 0$.*

A sentença acima é facilmente exprimível em PA, e uma possível formulação se daria no molde "$\forall m \forall n \exists k(m > 1 \wedge n > 1 \wedge m_k = 0)$". No entanto, ela não pode ser demonstrada neste sistema, embora seja verdadeira e demonstrável em outros (como ZFC, que é onde se dá a teoria dos números ordinais). O motivo para o teorema 4.2.2 valer (em ZFC) está no fato de que toda cadeia descendente de números ordinais é finita. O truque para se chegar nesse ponto é substituir cada n na representação de m (como acima) pelo ordinal ω. Vejamos, então, como tal argumento se dá no contexto de PK2.

A batalha hercúlea que simularemos se baseia numa *árvore finita*. Uma árvore é simplesmente um grafo com um nó fixado (chamado de *raiz*), de onde saem todos os caminhos. Desse modo, cada um dos nós da árvore está ligado à sua raiz. No entanto, tal ligação deve se dar através de um único caminho, i.e., por uma única sucessão de segmentos e nós. Abaixo, ilustramos um exemplo de árvore, que será a representante da configuração inicial da Hidra de Lerna na disputa contra Hércules (as letras romanas servirão apenas como auxílio na descrição do jogo, logo adiante):

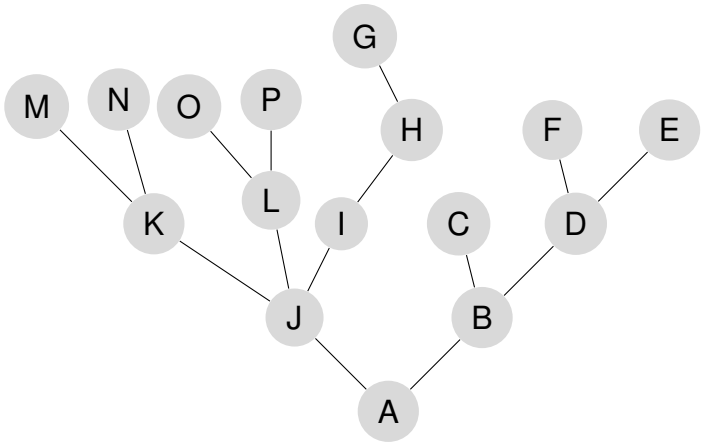


Adicionando um pouco mais de nomenclatura, sugestivamente chamamos todos os nós mais acima no grafo, ligados a apenas um segmento — e que não sejam a raiz —, de *nós superiores*; uma *cabeça* da hidra se constituirá de um nó superior mais o segmento que o antecede. Por exemplo, os caminhos K-N, B-C e D-E determinam cabeças.

A batalha, então, se dá de acordo com a seguinte regra: no n-ésimo estágio da luta ($n \geq 1$), Hércules deve cortar uma (e apenas uma) das cabeças da hidra; em seu lugar, nascem n "novas cabeças". A definição do que vem a ser uma "nova cabeça" depende da cabeça sendo cortada — ou, mais especificamente, dos segmentos e nós que a antecedem. Partindo do nó que antecede aquele que foi eliminado (i.e., do nó ligado ao segmento que também constituía a cabeça decepada), desça um segmento, em direção à raiz, até o próximo nó. A partir *deste* nó, devem crescer n cópias do

subgrafo logo acima (i.e., de todos os nós e segmentos que restaram, acima de tal nó, após a *n*-ésima decapitação). Ou seja, a "nova cabeça" será tal subgrafo. No entanto, se a "descida" em questão levar diretamente à raiz da árvore (i.e., se, em algum estágio, a cabeça a ser cortada estiver diretamente ligada à raiz por um único segmento), então nenhuma cabeça deve crescer. Essa regra é o que impede que a hidra sempre vença.

Para clarear as ideias, vejamos um exemplo. Se a primeira cabeça a ser cortada for aquela referente ao nó L, então o "nó de partida", como o chamamos acima, deve ser o nó K. Daí, deve-se descer um segmento em direção à raiz, chegando ao nó J. A nova cabeça, portanto, será constituída dos caminhos J-K-M e J-K-N (a parte que restou acima do nó de onde a nova cabeça brotará), e como esse é o primeiro estágio da batalha, apenas uma cópia de tal parte da árvore deve brotar do nó J. A nova configuração é ilustrada abaixo:



Continuando o processo para mais duas iterações, se a segunda cabeça cortada for aquela do nó C, então *duas* cópias do subgrafo com caminhos A-B-D-E e A-B-D-F devem brotar do nó A (a raiz); finalmente, cortando fora a cabeça do nó M, devem brotar três cópias do subgrafo J-K-N a partir do nó J. Sucessivamente, teremos o seguinte (destacando em azul as novas cabeças):

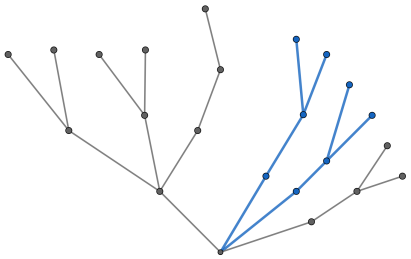


Figura 5 – Configuração da hidra após duas decapitações.

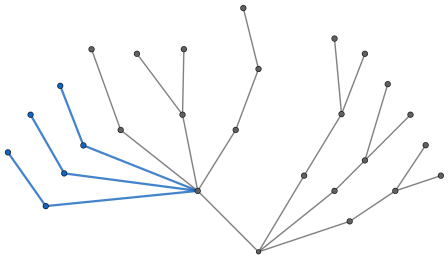


Figura 6 – Configuração da hidra após três decapitações.

Evidentemente, nossas escolhas sobre qual cabeça deveria ser cortada em cada estágio foram totalmente arbitrárias. De fato, a ideia é que não se imponha qualquer restrição quanto à tática que nosso herói deve usar para vencer a hidra; a única regra é que os cortes ocorram apenas em nós superiores. A princípio, portanto, Hércules pode escolher qualquer estratégia. E por *estratégia*, entende-se qualquer função que atribua a n (o estágio da luta) uma determinada cabeça a ser decepada. Assim, uma *estratégia vencedora* é uma tal função que leve à derrota da hidra. Isso acontece quando, após uma quantidade finita de decapitações, apenas a raiz da árvore permanecer intacta. A parte mais divertida dessa anedota — a despeito da forma extremamente rápida com que a árvore se ramifica — é que Hércules nem ao menos precisa bolar um plano de ataque, pois

Teorema 4.2.3. *Toda estratégia é uma estratégia vencedora.*

Para ter uma ideia de como isso é verdade, devemos entrar no campo da teoria de números ordinais. É aqui que PK1 e PK2 se assemelham. O conceito-chave é a *Forma Normal de Cantor*, e uma definição para tal conceito pode ser encontrada no Apêndice A. No contexto de sequências de Goodstein, fazemos o seguinte: dada a representação de m em base n (conforme visto anteriormente), substitua todas as ocorrências de n pelo ordinal ω, e chame tal número de $o_n(m)$. Em nosso exemplo recorrente, como $266 = 2^{2^{2+1}} + 2^{2+1} + 2^1$, temos que

$$o_2(266) = \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega^1.$$

Fato é que todo número ordinal tem sua Forma Normal de Cantor (vide teorema A.1.2). Mais do que isso, na prova original de Paris e Kirby, vê-se que toda sequência de Goodstein $a_0, a_1, a_2, ...$ tem uma sequência de formas normais correspondente, dada por

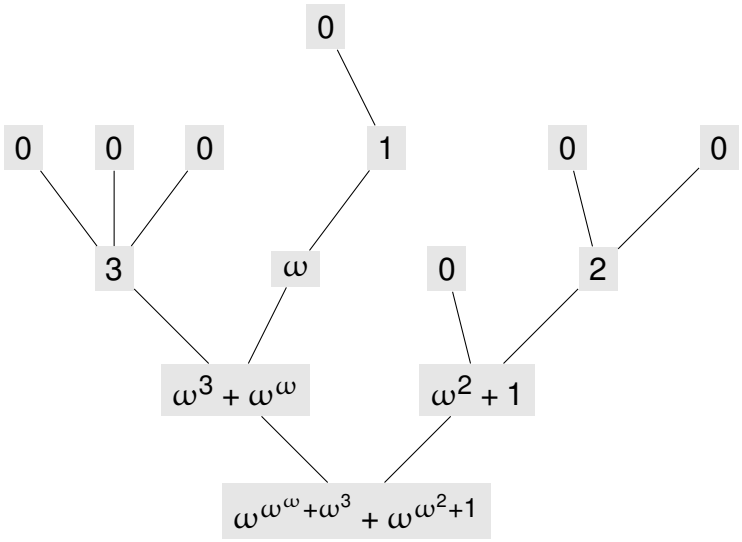
$$o_n(a_0),\ o_{n+1}(a_1),\ o_{n+2}(a_2), ...$$

Assim, tomando a sequência de Goodstein de 266 (vide p. 110), sua sequência de formas normais correspondente seria:

$$\omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega,$$
$$\omega^{\omega^{\omega+1}} + \omega^{\omega+1} + 2,$$
$$\omega^{\omega^{\omega+1}} + \omega^{\omega+1}, ...$$

Claramente, essa é uma sequência que só decresce. O mesmo ocorre para quaisquer $m, n > 1$; i.e., qualquer sequência de formas normais de Cantor correspondentes a termos de alguma sequência de Goodstein é estritamente decrescente (c.f. (KIRBY; PARIS, 1982), p. 289). Mas, daí, usando o fato de que toda sequência estritamente decrescente de ordinais é finita (um resultado fundamental da teoria de ordinais que utiliza indução transfinita abaixo de ε_0), segue-se diretamente o Teorema de Goodstein (i.e., nosso teorema 4.2.1). A prova para o teorema 4.2.3 segue de maneira muito parecida (com a vantagem do apelo visual). Vejamos como proceder.

Para cada nó da árvore que representa a hidra, associamos um número ordinal da seguinte maneira: os nós superiores recebem o número (ordinal) zero; a todos os demais, associa-se $\omega^{\alpha_1} + ... + \omega^{\alpha_n}$, em que os α_i são os ordinais atribuídos aos nós imediatamente acima daquele que está sendo considerado. Na configuração inicial de nossa hidra, obtemos o seguinte (atentando à convenção $\omega^0 = 1$):



O *ordinal da hidra*, portanto, será o ordinal atribuído à sua raiz. Como o leitor pode tentar conferir, os ordinais da hidra nos estágios posteriores são, respectivamente,

$$\omega^{\omega^{\omega}+\omega^2\cdot 2} + \omega^{\omega^2+1};$$
$$\omega^{\omega^{\omega}+\omega^2\cdot 2} + \omega^{\omega^2} \cdot 3;$$
$$\omega^{\omega^{\omega}+\omega^2+\omega\cdot 4} + \omega^{\omega^2} \cdot 3.$$

Perceba que essa sequência é estritamente decrescente (embora, é claro, os expoentes vão diminuindo muito lentamente). Para provar essa constatação — bastante evidente, em nosso exemplo particular — no caso de uma hidra arbitrária, Paris e Kirby tomam uma estratégia σ qualquer, e definem uma operação, $[\alpha]_\sigma(n)$, que associa o ordinal da hidra após $n-1$ estágios ao ordinal da hidra após n estágios. Daí, prova-se que para qualquer ordinal α, se $0 < \alpha < \varepsilon_0$, então

$$[\alpha]_\sigma(n) < \alpha. \tag{23}$$

Porque toda sequência decrescente de ordinais é finita, tem-se o resultado — i.e., toda estratégia é vencedora.

Agora, veja que ao lançarmos mão dos métodos de teoria dos números ordinais, tal prova não pode ser conduzida em PA. Mais do que isso, PA nem ao menos é capaz de *expressar* a sentença do teorema 4.2.3, simplesmente pelo fato de que estratégias, em geral, são objetos *infinitários*. Isso ocorre porque, uma vez que podemos codificar cada hidra através de sequências de números naturais,[14] estratégias devem ser funções de $\mathbb{N}$ em $\mathbb{N}$; todavia, existem $2^{\aleph_0}$ funções nos números naturais (uma quantidade incontável, portanto), ao passo que a linguagem de PA, finita por natureza, só é capaz de descrever uma quantia de termos e fórmulas na ordem de $\aleph_0$ (enumerável, portanto). Nesse sentido, para chegarmos à formulação de PK2, precisamos nos ater a estratégias *recursivas*:

PK2. *Toda estratégia recursiva é uma estratégia vencedora.*

Como PK2 é uma restrição do teorema 4.2.3, e ambos são exprimíveis em PA, torna-se imperativo, sob a hipótese de PA ser correto, que PA $\nvdash \neg$PK2. A prova de PA $\nvdash$ PK2 segue linhas muito parecidas com a prova de indecidibilidade de PK1 (como era de se esperar, já que ambas são demonstradas de maneiras quase idênticas em ZFC). Em suma, o ponto-chave continua sendo a Forma Normal de Cantor, mas algumas outras construções são necessárias; em especial, Paris e Kirby se apoiam, em ambas as provas, no trabalho de Ketonen e Solovay[15] sobre funções de crescimento rápido. Trata-se, no entanto, de um aparato técnico que vai muito além dos nossos propósitos; aqui, basta apontarmos alguns dos aspectos gerais.

Em primeiro lugar, como no caso de PH, modelos não-*standard* desempenham um papel importante nas provas de indecidibilidade de PK1 e PK2; mais especificamente, no caso de PK1, a suposição PA $\vdash$ PK1 nos leva a um absurdo dentro de PA (sob o modelo não-*standard* sendo considerado). Para a indecidibilidade de PK2,

[14] C.f. (MURAWSKI, 1999), p. 166-167. Basicamente, devemos associar uma sequência numérica a cada nó da árvore; a codificação de sequências (por funções Beta, concatenação, ou qualquer outro método efetivamente decidível) garante que existe um (e apenas um) número natural para cada árvore possível. O processo, então, pode ser revertido, i.e., tomando-se um número qualquer, pode-se decodificá-lo, de modo a reconstruir a árvore correspondente.

[15] C.f. (KETONEN; SOLOVAY, 1981).

constrói-se uma estratégia recursiva, τ, para a qual não vale (23) (acima); seguindo os mesmos passos que levam à não-demonstrabilidade de PK1, prova-se que τ não pode ser vencedora, e segue-se que PK2 é indecidível.[16]

O segundo ponto de interesse conversa mais com a metateoria (e uma relação direta de certo resultado desta com PK2) do que com o problema em si. O resultado em questão é a prova de Gerhard Gentzen (1909-1945) de que PA é consistente (em termos de indução transfinita para ordinais abaixo de ε_0, e assumindo a consistência da teoria de números ordinais).[17] Segundo (HAMANO; OKADA, 2006), cada etapa da prova de Gentzen para a consistência de PA pode ser interpretada como uma quantidade finita de etapas do jogo da hidra proposto por Paris e Kirby. Isso sugere uma estreita relação entre a Teoria da Prova e um resultado puramente combinatório. Somando-se a tudo isso o fato de que a prova de PH passa por proposições envolvendo sentenças metateóricas, podemos concluir que, embora nossos exemplos tenham conteúdos "puramente matemáticos", os argumentos que justificam suas insolubilidades estão fortemente amarrados à lógica e à teoria dos conjuntos — i.e., à Matemática Abstrata. Mas, antes que o próximo exemplo seja bem explorado, essas digressões podem esperar um momento.

4.2.3 HF: levando a incompletude concreta níveis além de PA

Como mencionado acima, a sentença de Harvey Friedman é especial porque a função que descreve os objetos envolvidos nela tem crescimento (muito) superior à função correspondente a PH (que já é uma função absurdamente "veloz"). Aqui, exploraremos mais detalhadamente o papel de tais crescimentos no fenômeno de indecidibilidade.

Primeiramente, contudo, vejamos seu conteúdo. Assim como PH e PK2 — e a maioria dos exemplos de incompletude concreta em PA — HF mora no campo da análise combinatória. Também à semelhança das outras sentenças, HF surge de uma pequena modificação de um teorema bastante poderoso — nesse caso, uma restrição a objetos finitos no *Teorema de Kruskal*.[18] Como em PK2, os objetos de interesse

[16] Mais especificamente, por (23) não ser satisfeita, uma suposta demonstração de que τ é uma estratégia vencedora levaria a outro absurdo dentro de PA, relativo às mesmas construções (devido a Ketonen-Solovay) do absurdo da prova de PK1, mencionado acima.

[17] O leitor não deve se espantar com a existência de tal resultado, pois ele não contradiz, em absoluto, qualquer instância dos teoremas de incompletude — a bem da verdade, Gentzen faz uso dos mesmos em sua prova. *Grosso modo*, pode-se dizer que a prova de Gentzen é mais um caso daquelas provas de consistência que necessitam de um modelo específico assumidamente consistente (como a prova de Hilbert para a consistência da geometria elementar, onde se assume a consistência da teoria de números reais; nesse caso, o modelo assumidamente consistente é da aritmética das funções recursivas primitivas, ou Aritmética de Skolem, um sistema formal incomparável com PA, i.e., "nem mais forte nem mais fraco" que este). Para um tratamento palatável da prova de Gentzen, veja (STILLWELL, 2010), p. 130-133.

[18] A bem da verdade, Friedman construiu *diversos* exemplos de sentenças indecidíveis a partir do Teorema de Kruskal. Nós apresentaremos apenas uma delas.

são árvores finitas; aqui, no entanto, usaremos uma definição ligeiramente diferente (e mais formal).

Uma *árvore finita* deve ser entendida como um conjunto parcialmente ordenado de pontos tal que (i) todo elemento não-minimal tem um único antecessor; e (ii) existe um único elemento minimal (a *raiz*). Atribuímos o par $(\mathfrak{t}, <)$ ao nome de uma árvore, mas em geral se omite a ordem, escrevendo-se apenas "$\mathfrak{t}$".

Dadas duas árvores finitas $\mathfrak{t}_1$ e $\mathfrak{t}_2$, chamamos de um *implante homeomórfico* qualquer função injetiva $f : \mathfrak{t}_1 \longrightarrow \mathfrak{t}_2$ que preserve a ordem. Assim, num implante homeomórfico como esse, para quaisquer pontos $a, b \in \mathfrak{t}_1$, se $a < b$, então $f(a) < f(b)$ em $\mathfrak{t}_2$, i.e., pontos "mais abaixo" em $\mathfrak{t}_1$ são levados em pontos "mais abaixo" em $\mathfrak{t}_2$. Se uma f como essa existir, então $\mathfrak{t}_1$ é *homeomorficamente implantável* em $\mathfrak{t}_2$, e escrevemos $\mathfrak{t}_1 \preccurlyeq \mathfrak{t}_2$. O Teorema de Kruskal, então, se lê:

Teorema 4.2.4. *Seja* $\langle \mathfrak{t}_1, \mathfrak{t}_2, ... \rangle$ *uma sequência infinita de árvores finitas. Então, existem* $i, j \in \mathbb{N}^*$*, com* $i < j$*, tais que* $\mathfrak{t}_i \preccurlyeq \mathfrak{t}_j$*.*

A formulação de HF surge ao restringirmos tanto os tamanhos das árvores quanto o tamanho da sequência correspondente:

HF. *Dado um número natural k qualquer, com* $k > 1$*, existe* $n \in \mathbb{N}$ *tal que se* $\langle \mathfrak{t}_1, ..., \mathfrak{t}_n \rangle$ *for uma sequência finita de árvores finitas tais que cada* $\mathfrak{t}_i$ *tem cardinalidade máxima de* $k + i$*, então existem* $i, j \in \mathbb{N}^*$*, com* $i < j \leq n$*, tais que* $\mathfrak{t}_i \preccurlyeq \mathfrak{t}_j$*.*

Assim como no caso do Teorema de Ramsey, a relevância do Teorema de Kruskal reflete seu caráter puramente matemático. Uma aplicação importante é devido a Andrzej Ehrenfeucht e resulta na boa-ordenação de certo conjunto de polinômios exponenciais. Trata-se, obviamente, de um resultado não tão proeminente quando o Princípio de Dirichlet (que, lembre-se, é um caso particular do Teorema de Ramsey), mas seu conteúdo algébrico é evidentemente concreto.

Os argumentos para HF tanto ser verdadeira quanto indecidível seguem linhas paralelas às respectivas provas sobre PH. Em suma, a veracidade de HF segue do Lema de König (que fala justamente sobre árvores), e sua indecidibilidade é consequência do fato de que $\mathrm{PA} \vdash \mathrm{HF} \rightarrow \mathrm{Con}_{\mathrm{PA}}$. No entanto, HF também implica a consistência de PA+PH, i.e., a teoria resultante da adição de PH a PA como um novo axioma. Mais ainda, HF implica a própria PH [c.f. (HARRINGTON *et al.*, 1985), p. 384.], e disso podemos concluir que HF é *mais forte* do PH, tanto no sentido matemático quanto metateórico. Em particular, a indecidibilidade de HF se verifica em teorias mais fortes que PA, e também em teorias mais fortes que PA+PH.

Agora, a indecidibilidade de PH (ou, melhor dizendo, a dificuldade de se dar uma eventual prova construtiva deste possível teorema[19]) pode ser mensurada pela

[19] Pois, em certo sentido, é isso que "uma sentença ser indecidível em dada teoria" significa: ela é indecidível via métodos finitários, i.e., não há prova construtiva para ela em tal teoria.

função que descreve o crescimento do conjunto em questão. O mesmo tipo de análise pode ser feito para HF — i.e., existe uma função que descreve o crescimento das árvores envolvidas. Mas se a função referente a PH já crescia mais rapidamente do que qualquer função computável, o que podemos dizer sobre HF?

Como de praxe neste capítulo, tal problema nos leva à teoria de números ordinais. O motivo para isso mora no fato de que o Teorema de Kruskal pode ser enunciado apenas em termos desta teoria, uma vez que toda árvore é, em essência, um conjunto parcialmente ordenado.[20] *Grosso modo*, a hierarquia das funções de crescimento rápido se estabelece ao associarmos ordinais específicos a cada uma delas: quanto maior o ordinal, maior a taxa de crescimento. O método consiste em criar funções entre ordinais contáveis, i.e., ordinais relativos a conjuntos de cardinalidade enumerável. A partir de funções com ordinais relativamente baixos, constroem-se outras, cujos valores são ordinais maiores que os anteriores, e assim por diante. Nesse sentido, à função referente a PH associa-se o ordinal ε_0.

O caso de HF é outro: não se sabe exatamente qual ordinal devemos associar a ela. Entretanto, muitos limites superiores e inferiores já foram encontrados. O limite mais baixo é Γ_0, que, como apontado no Apêndice A, é o menor ordinal contável em cuja definição precisamos apelar para o universo *in*contável. A construção formal dessa hierarquia de funções se encontra no referido texto; aqui, exploraremos a conexão entre Γ_0 e o Teorema de Kruskal (o que vai explicar por que HF é indecidível).

Para isso, vamos definir uma função entre o conjunto de todas as árvores finitas e o conjunto de todos os ordinais $< \Gamma_0$, que seja sobrejetiva e preserve a ordem. Como cada ordinal $< \Gamma_0$ terá uma árvore correspondente, e árvores de mesmo tamanho serão levadas em ordinais de mesmo tamanho, o Teorema de Kruskal (que versa sobre boa ordenação em árvores) vai implicar a boa ordenação dos ordinais contáveis $< \Gamma_0$. Colocando de outra maneira, vamos concluir que quando o Teorema de Kruskal é válido, o Teorema de Indução Transfinita para ordinais até Γ_0 também é válido. Como tal resultado não é demonstrável em diversas teorias[21], vamos concluir que o Teorema de Kruskal, bem como suas outras versões construídas por Friedman (HF inclusa), são indecidíveis em PA.

Chame, então, de $\mathfrak{T}$ o conjunto de todas as árvores finitas, e de $\mathcal{O}(\Gamma_0)$ o conjunto de todos os ordinais $< \Gamma_0$. Definimos $h : \mathfrak{T} \longrightarrow \mathcal{O}(\Gamma_0)$ de maneira recursiva, através de zero, +, e as funções $g(\alpha, \beta)$:

[20] Mais especificamente, o Teorema de Kruskal pode ser enunciado em termos de *boa-quasi-ordenação*, e não apenas em termos de boa-ordem. Uma *quasi*-ordem é uma relação reflexiva e transitiva, enquanto uma ordem parcial, além disso, é antissimétrica. Um conjunto *bem*-quasi-ordenado é aquele tal que para toda sequência infinita $\langle x_i \rangle$, tem-se $x_i < x_j$ para todo $i < j$. Daí, o Teorema de Kruskal se lê: "A quasi-ordem induzida pela implantabilidade homeomórfica de árvores finitas é uma boa-quasi-ordem".

[21] Lembre-se: segundo Gentzen, a indução transfinita em ordinais até ε_0, um ordinal muito menor do que Γ_0, é indecidível em PA.

1. se $\mathfrak{t}$ é a árvore com apenas um nó (i.e., somente a raiz), então $h(\mathfrak{t}) = 0$;
2. se $\mathfrak{t}$ é tal que sua raiz possui apenas um sucessor (i.e., há apenas um nó imediatamente acima da raiz), então $\mathfrak{t}$ é formada pela raiz mais uma sub-árvore $\mathfrak{t}'$; faça $h(\mathfrak{t}) = h(\mathfrak{t}')$;
3. se a raiz de $\mathfrak{t}$ possui exatamente dois sucessores, então há exatamente duas sub-árvores, digamos $\mathfrak{t}'$ e $\mathfrak{t}''$; sem perda de generalidade, suponha $h(\mathfrak{t}') \geq h(\mathfrak{t}'')$, e faça $h(\mathfrak{t}) = h(\mathfrak{t}'') + h(\mathfrak{t}')$;
4. se a raiz de $\mathfrak{t}$ possui exatamente três sucessores, teremos três sub-árvores, $\mathfrak{t}'$, $\mathfrak{t}''$ e $\mathfrak{t}'''$; assumindo que $h(\mathfrak{t}') \geq h(\mathfrak{t}'') \geq h(\mathfrak{t}''')$ e que $h(\mathfrak{t}') < g(h(\mathfrak{t}'), h(\mathfrak{t}''))$, faça $h(\mathfrak{t}) = g(h(\mathfrak{t}''), h(\mathfrak{t}'))$;
5. se a raiz de $\mathfrak{t}$ possui exatamente três sucessores, assumindo que $h(\mathfrak{t}') \geq h(\mathfrak{t}'') \geq h(\mathfrak{t}''')$ e que $h(\mathfrak{t}') = g(h(\mathfrak{t}'), h(\mathfrak{t}''))$, faça $h(\mathfrak{t}) = h(\mathfrak{t}') + h(\mathfrak{t}'')$;
6. se a raiz possui quatro o mais sucessores, assumindo que $h(\mathfrak{t}') \geq h(\mathfrak{t}'') \geq h(\mathfrak{t}''') \geq h(\mathfrak{t}'''')$, faça $h(\mathfrak{t}) = g(h(\mathfrak{t}'), h(\mathfrak{t}''))$.

Por mais que esta seja uma definição bastante longa e de difícil leitura, a prova de que h é sobrejetiva depende unicamente do fato, provado no teorema A.2.7, que todo ordinal $< \Gamma_0$ pode ser obtido recursivamente de 0, + e g. As cláusulas acima, nesse sentido, cuidam de cobrir tal demanda. Para provar que h preserva a ordem, cada caso precisa ser analisado separadamente. De todo modo, não mais do que um processo de indução transfinita e alguns dos resultados presentes no Apêndice A são necessários. Assim, temos o tão desejado teorema:

Teorema 4.2.5. *A função $h : \mathfrak{T} \longrightarrow \mathcal{O}(\Gamma_0)$ é sobrejetiva e preserva a ordem — i.e., para qualquer $\alpha \in \mathcal{O}(\Gamma_0)$, existe $\mathfrak{t} \in \mathfrak{T}$ tal que $\alpha = h(\mathfrak{t})$, e para quaisquer $\mathfrak{t}_1, \mathfrak{t}_2 \in \mathfrak{T}$, se $\mathfrak{t}_1 \preccurlyeq \mathfrak{t}_2$, então $h(\mathfrak{t}_1) \leq h(\mathfrak{t}_2)$.*

Mais simples de se demonstrar — e, também, mais importante para os nossos fins — é o seguinte corolário:

Corolário 4.2.1. *O Teorema de Kruskal implica que $\mathcal{O}(\Gamma_0)$ é bem ordenado pela relação $\leq$.*

Demonstração. Assuma o contrário. Então, existe uma sequência infinita $\langle \alpha_n \rangle$ tal que $\alpha_{i+1} < \alpha_i$ para todo $i \geq 1$. Da sobrejetividade de h, deve existir uma sequência infinita de árvores $\langle \mathfrak{t}_n \rangle$ tal que $h(\mathfrak{t}_i) = \alpha_i$, qualquer que seja $i \geq 1$. Pelo Teorema de Kruskal, existem índices i, j tais que $i < j$ e $\mathfrak{t}_i \preccurlyeq \mathfrak{t}_j$, e como h preserva a ordem, temos que

$$\alpha_i = h(\mathfrak{t}_i) \leq h(\mathfrak{t}_j) = \alpha_j.$$

Mas isso contradiz a hipótese de que $\alpha_j < \alpha_i$ se $i < j$. Logo, tal hipótese é absurda, e concluímos que $\mathcal{O}(\Gamma_0)$ é bem ordenado por $\leq$. □

Abreviando as boas ordenações de $\mathcal{O}(\Gamma_0)$ e $\mathfrak{T}$ por $\mathrm{BO}(\Gamma_0)$ e $\mathrm{BO}(\mathfrak{T})$, respectivamente, e a boa-quasi-ordenação de $\mathfrak{T}$ por $\mathrm{BQO}(\mathfrak{T})$, o corolário 4.2.1 se lê:

$$\mathrm{BQO}(\mathfrak{T}) \to \mathrm{BO}(\Gamma_0).$$

Escrevendo as coisas desse jeito, por contraposição, se $\mathrm{BO}(\Gamma_0)$ não pode ser demonstrada em certo sistema formal, então tal teoria tampouco prova o Teorema de Kruskal. Vejamos, portanto, algumas teorias em que isso ocorre.

Como afirmamos anteriormente, Gentzen foi capaz de provar que a indução transfinita até o ordinal ε_0 não é válida para PA. Isso é equivalente a dizer que existem sequências *decrescentes* de ordinais $< \varepsilon_0$, i.e., que o fato de o conjunto dos ordinais menores que ε_0 ser bem ordenado não pode ser demonstrado em PA. Como $\varepsilon_0 < \Gamma_0$, isso implica que o Teorema de Kruskal é indecidível em PA. Mas há exemplos muito mais interessantes.

Os três sistemas que Friedman investigou mais atentamente foram ACA_0, ATR_0 e Π^1_1-CA_0, que são subteorias da aritmética de segunda ordem. Em contraste com PA, que é de primeira ordem, neles consideramos, também, variáveis sobre *conjuntos* de números naturais, e os esquemas de indução se tornam um único axioma. Além disso, em cada um deles se considera alguma forma do que chamamos de *axioma de compreensão*:

$$\exists X \forall n[n \in X \leftrightarrow \varphi(n)],$$

em que X é uma variável de conjunto, e φ é uma fórmula qualquer de segunda ordem, desde que X não ocorra livre dentro dela.

O sistema ACA_0, que abrevia "Arithmetical Comprehension Axioms", considera o axioma de compreensão restrito a fórmulas aritméticas, e pode ser visto como uma extensão conservativa de PA. Aqui, construções teóricas muito mais robustas que a aritmética elementar podem ser obtidas, como a teoria de funções contínuas e sequências convergentes (uma parte da análise real). Em particular, o Teorema de Bolzano-Weierstrass[22] é válido em ACA_0.

O sistema ATR_0, que abrevia "Arithmetical Transfinite Recursion" inclui uma versão do axioma de compreensão que o extrapola para qualquer ordinal contável. Esta é uma teoria bastante poderosa, capaz de descrever grande parte do conteúdo matemático visto num curso de graduação. Por exemplo, pode-se construir o cálculo integral, a teoria de boa-ordenação para ordinais contáveis, a topologia de espaços métricos separáveis, conjuntos de Borel, etc.

[22] "Toda sequência limitada de números reais possui uma subsequência convergente".

Por fim, o sistema Π^1_1-CA é obtido quando consideramos as fórmulas φ do axioma de compreensão como sendo fórmulas Π^1_1, i.e., fórmulas com (apenas) quantificadores universais, estes agindo sobre variáveis de conjuntos. Dos três sistemas, este é o mais forte, e, além das construções teóricas possíveis que citamos acima, usando Π^1_1-CA podemos falar de abstrações extremamente complexas, como teoria descritiva de conjuntos, e outras tantas áreas fundamentais da Matemática, como análise combinatória, álgebra e análise (em espaços dos mais variados).

O que Friedman provou foi, primeiramente, que BO(Γ_0) não é demonstrável em ATR_0; depois, que $\mathsf{ACA}_0 \vdash \mathrm{BQO}(\mathfrak{T}) \to \mathrm{BO}(\Gamma_0)$. Como ACA_0 é uma subteoria de ATR_0, ele concluiu que BQO($\mathfrak{T}$) não é demonstrável em ATR_0. Isso, dada a quantidade de Matemática que se pode construir nesse sistema, já demonstra como o Teorema de Kruskal é um exemplo grandioso de incompletude concreta (e, claro, como o trabalho de Friedman é brilhante).

Ocorre que BQO($\mathfrak{T}$) é uma sentença Π^1_1 (i.e., da forma $\forall X \varphi$), logo sua indecidibilidade em sistemas de segunda ordem, como os que estamos discutindo, não representa algo tão impressionante. Nesse sentido, o feito mais notável de Friedman foi determinar a não-demonstrabilidade de HF, e outras versões "miniaturizadas" de BQO($\mathfrak{T}$), que são sentenças Π^0_2 (i.e., da forma $\forall x \exists y \varphi$, e, portanto, de *primeira* ordem), em sistemas muito mais robustos do que o necessário para formulá-las. Em particular, ele provou que HF é indecidível em ATR_0, e que uma extensão da mesma é indecidível em Π^1_1-CA.

Aqui, finalizamos nossa discussão sobre a relação entre o fenômeno de indecidibilidade e funções de rápido crescimento, bem como a exposição de exemplos de incompletude concreta (que esperamos ter tornado evidente). Comentários mais objetivos e sintetizados serão tecidos na conclusão do trabalho, para onde seguimos imediatamente.

5 CONCLUSÃO

Façamos, agora, uma breve síntese daquilo que buscávamos ao início deste trabalho, e daquilo que conseguimos alcançar. Diferentemente do que diz o senso comum, no entanto, devemos atentar ao fato de que conclusões, em geral, não são marcadas apenas por afirmações categóricas. Muitas vezes, no final de uma pesquisa, sobram mais *perguntas* do que respostas. Este parece ser o nosso caso.

Em primeiro lugar, como já frisamos algumas vezes, a tese deste trabalho de conclusão de curso, bem como seu escopo e seu público-alvo, são bastante amplos. A ideia, de fato, era que qualquer pessoa familiarizada com os conceitos básicos de lógica-matemática entendesse tanto o contexto quanto o conteúdo dos Teoremas de Incompletude de Gödel. Para isso, foi necessário estabelecer o ambiente metamatemático por inteiro — desde a concepção dos processos efetivos de decisão, que caracterizam os meios através dos quais podemos derivar as sentenças válidas de uma dada teoria matemática, até o estudo de um sistema aritmético formal específico.

Nesse sentido, até o final do terceiro capítulo, nosso labor foi mais exaustivo do que complicado. Todavia, acreditamos ter cumprido de maneira satisfatória nosso primeiro objetivo: de forma bastante detalhada, construímos não apenas uma, mas três sentenças indecidíveis na aritmética elementar. Justamente por ser esta a teoria de números mais usual (que aprendemos na escola e que vem sendo construída desde que a humanidade aprendeu a contar), tais exemplos demonstram de maneira clara o impacto das descobertas de Gödel. Afinal, se nem mesmo o terreno matemático em que qualquer leigo consegue caminhar está livre de incertezas, o que dizer, então, das áreas mais abstratas do conhecimento matemático? [É claro que essa não foi uma pergunta que tentamos responder, pois nos ativemos a PA. Mas ela atesta a força dos Teoremas.]

Afora os conteúdos autorreferentes das sentenças de Gödel e de Rosser, que motivam a distinção entre o “concreto” e o “abstrato” dentro da Matemática, suas formulações são notadamente aritméticas. O mesmo acontece com a sentença que afirma a consistência de PA, mas, como vimos, seu caso é bem mais delicado, pois depende da fórmula que escolhemos pra representar o fenômeno de contradição. Além disso, a indecidibilidade da consistência está amarrada às propriedades que o operador de demonstrabilidade sendo usado satisfaz (no sentido que vimos em nossa breve discussão sobre lógica modal). Tais dificuldades nos dizem que a indecidibilidade de $\mathrm{Con}_{\mathrm{PA}}$, em particular, ou de qualquer outra sentença equivalente a ela, não determinam, em absoluto, a impossibilidade de se provar a consistência de PA.

Isso vai de encontro com um dos mitos/má interpretações mais difundidos sobre os teoremas de incompletude. Diz-se, ocasionalmente, que o Segundo Teorema veio para destruir todas as esperanças depositadas no Programa de Hilbert, pois “é

impossível atestar a consistência de um sistema aritmético". Tal tipo de afirmação não só caracteriza leviandade — pois, em linguagem natural, o mais correto seria dizer que "é impossível atestar a consistência de um sistema aritmético *através do próprio sistema*" —, como, também, é uma afirmação falsa. A prova de Gentzen, por exemplo, nos diz que PA *é* consistente (embora deixe em aberto a hipótese de que a Aritmética de Skolem o seja). Soma-se a tudo isso o fato de que o Programa de Hilbert nunca foi completamente estabelecido (a começar pela estranha distinção entre "finitário" e "*in*finitário"), e vê-se que, de fato, sobram mais perguntas do que respostas. O estado da arte do Programa de Hilbert, portanto, é uma das linhas de pesquisa que este trabalho desencadeia.

Outro afluente importante se dá no tratamento do predicado de demonstrabilidade como um operador modal. Como vimos brevemente durante a discussão sobre as consequências (lógicas) diretas dos teoremas de incompletude, sob essa perspectiva, questões bastante sofisticadas se tornam muito mais legíveis, e teoremas fortes brotam como meros casos particulares de resultados muito mais gerais. Em suma, ao trabalharmos numa lógica específica para atacar problemas de decidibilidade, podemos elevar o nível de abstração enormemente. As possibilidades são incontáveis. Ademais, tendo em vista o histórico recente de não mais que quatro décadas, é uma área ainda bastante prolífica e, portanto, interessante de se aventurar.

Também de grande interesse são as questões concernentes aos *meios* necessários para se determinar incompletude. Nosso foco se deu em sentenças autorreferentes, pois este foi o método que o próprio Gödel encontrou para obter sua prova original, mas desde então foram construídas demonstrações das mais variadas. Uma, devido a George Boolos, utiliza apenas o conceito de máquinas de Turing. Nesse sentido, vale ressaltar que poderíamos ter optado por um tratamento (equivalente) em termos de máquinas — mas, aí, nosso protagonista seria Emil Post, não Kurt Gödel. As construções teóricas que surgem de um tratamento mais mecânico são, também, inestimáveis, e um ótimo caminho a se seguir a partir do presente estudo. De qualquer forma, análises sobre o papel da autorreferência e, mais especificamente, sobre o conceito de diagonalização no fenômeno de indecidibilidade são ótimos temas de pesquisa; de fato, elas também conversam intimamente com as ideias presentes na lógica de demonstrabilidade mencionada acima.

Finalmente, quanto à nossa tese, temos poucos comentários *categóricos* a tecer; as exposições de exemplos de indecidibilidade concreta, por si mesmas, atestam o fato de que não apenas sentenças "artificais", construídas por "meros truques lógicos", são indemonstráveis em certos sistemas aritméticos. Nosso receio inerente, enquanto matemáticos, de que algumas conjecturas jamais poderão ser decididas é, portanto, plenamente justificável.

Aqui, no entanto, sobram *muito* mais dúvidas do que respostas. O primeiro

ponto problemático, evidentemente (e a despeito do que afirmamos acima), se dá na distinção entre matemática concreta e matemática abstrata. Em nossa breve discussão sobre o assunto, nem de longe fomos capazes de delinear o limite entre estas duas de maneira totalmente clara. Em particular, para muitos autores (este incluso), a hipótese do *continuum* e o axioma da escolha representam casos de incompletude concreta tão evidentes quanto HF, PK e PH. O trabalho pioneiro de Harvey Friedman, portanto, merece um estudo bem mais detalhado do que nossa curta visita abordou.

Um segundo ponto problemático diz respeito aos métodos utilizados para se determinar a não-demonstrabilidade de sentenças matematicamente concretas. Como vimos, se, por exemplo, PH fosse demonstrável, então $\mathrm{Con}_{\mathsf{PA}}$ também o seria. Isso amarra o fenômeno de incompletude concreta à incompletude abstrata. Nesse sentido, um possível estudo futuro seria investigar quão apertado é esse nó, i.e., quais são os motivos para precisarmos de sentenças metalinguísticas para determinarmos a indecidibilidade de sentenças que ocorrem naturalmente na Matemática.

Agora, se por um lado podemos dar como vencido o desafio de provar para a comunidade matemática comum, notadamente negligente e cética com os Teoremas de Incompletude de Gödel, que tais resultados impactam, sim, *diretamente* sua atividade profissional cotidiana, por outro lado, até onde vai o conhecimento daquele vos escreve, nenhuma *grande* conjectura, como a de Goldbach, ou a hipótese de Riemann, se mostrou indecidível. [Na literatura mais antiga sobre o assunto, o exemplo em voga era o Último Teorema de Fermat (que por 358 anos foi apenas uma conjectura, até finalmente ser provado por Andrew Wiles, em 1995, quando passou a ser chamado de Teorema de Fermat-Wiles).] A questão que importa aqui é que, a menos que consigamos provar ou refutar determinada sentença, esta continuará sendo um problema em aberto, e há problemas em aberto mais interessantes ou complicados do que outros.

Dito isto, essencialmente, a única diferença entre PH, PK e HF e as conjecturas de Goldbach e de Riemann é que estas ainda estão em processo de decisão, enquanto que aquelas já tiveram sua indecidibilidade determinada. Fato é que ambas podem ser formuladas na linguagem da aritmética (Goldbach por motivos óbvios, e a hipótese de Riemann por ser equivalente a uma sentença que envolve apenas uma soma, uma desigualdade, uma função logarítmica e outra exponencial; este resultado consta em (LAGARIAS, 2002)). Assim, podemos atacar essas conjecturas com todo o arsenal que construímos (e provavelmente mais algumas coisas, como teoria de modelos), e tentar prová-las indecidíveis. De qualquer maneira, tanto essa investigação quanto as investigações mais pertinentes do ponto de vista matemático usual, i.e., o ato de tentar prová-las ou refutá-las, são caminhos interessantes, e um não diminui o sentido de ser do outro.

REFERÊNCIAS

BARKER, Stephen F. **Philosophy Of Mathematics**. Englewood Cliffs: Prentice Hall, Inc., 1964. (Foundations of Philosophy Series). ISBN 6419005.

BOOLOS, George; BURGESS, John; JEFFREY, Richard. **Computability and Logic**. 5ª edição. Cambridge: Cambridge University Press, 2007. ISBN 9780521701464.

BOOLOS, George; JEFFREY, Richard. **Computability and Logic**. 2ª edição. Cambridge: Cambridge University Press, 1980. ISBN 0521234794.

BOOLOS, George; JEFFREY, Richard. **Logic, logic, and logic**. Cambridge: Harvard University Press, 1999. ISBN 9780674537675.

BOOLOS, George S. **The Logic of Provability**. Nova Iorque: Cambridge University Pess, 2008. ISBN 9780521483254.

BUTTON, Tim; WALSH, Sean. **Philosophy and Model Theory**. Nova Iorque: Oxford University Press, 2018. ISBN 9780198790402.

DAVIS, Martin. **Computability and Unsolvability**. Nova Iorque: Dover Publications, 1985. ISBN 9780486614717.

FEFERMAN, Solomon. Are There absolutely unsolvable problems? Gödel's Dichotomy. **Philosophia Mathematica**, v. 14, n. 2, p. 134–152, jun. 2006a. ISSN 00318019. DOI: 10.1093/philmat/nkj003.

FEFERMAN, Solomon. The Nature and Significance of Gödel's Incompleteness Theorems. **Lecture for the Princeton Institute for Advanced Study Gödel Centenary Program**, 2006b. Disponível em: http://math.stanford.edu/~feferman/papers.html.

FRANZÉN, Torkel. **Gödel's theorem: an incomplete guide to its use and abuse**. Wellesley: A K Peters, Ltd., 2005. ISBN 9781568812380.

FRANZÉN, Torkel. **Inexhaustibility: A Non-Exhaustive Treatment**. Wellesley: A K Peters/CRC Press, 2008. ISBN 9781568811741.

GALLIER, Jean H. What's so special about Kruskal's theorem and the ordinal Γ_0? A survey of some results in proof theory. **Annals of Pure and Applied Logic**, 1991. DOI: doi:10.1016/0168-0072(91)90022-e.

HÁJEK, Petr; PUDLÁK, Pavel. **Metamathematics of First-Order Arithmetic**. Berlim: Springer, 1998. (Perspectives in Mathematical Logic). ISBN 9780136019701.

HAMANO, Masahiro; OKADA, Mitsuhiro. A Relationship Among Gentzen's Proof-Reduction, Kirby-Paris' Hydra Game and Buchholz's Hydra Game. **Mathematical Logic Quarterly**, 2006. DOI: https://doi.org/10.1002/malq.19970430113.

HARRINGTON, Leo; MORLEY, Michael; SCEDROV, Andre; SIMPSON, Stephen. **Harvey Friedman's Research on the Foundations of Mathematics**. Amsterdam: Elsevier Science Ltd, 1985. (Studies in Logic and the Foundations of Mathematics). ISBN 9780444878342.

HARRINGTON, Leo; PARIS, Jeff. A Mathematical Incompleteness in Peano Arithmetic. **Studies in Logic and the Foundations of Mathematics**, 1977. DOI: https://doi.org/10.1016/S0049-237X(08)71130-3.

HEYTING, Arend. Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie. **Berlin: Springer**, 1934. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-65617-0_9.

HILBERT, David. Über das Unendliche. **Matematische Annalen**, 1926. DOI: 10.1007/BF01206605. Disponível em: https://link.springer.com/article/10.1007/BF01206605.

HUGHES, G. E.; CRESSWELL, M. J. **A New Introduction to Modal Logic**. Londres: Routledge, 1996. ISBN 0415125995.

JECH, Thomas. **Set Theory**. Nova Iorque: Springer, 2014. ISBN 9783540440857.

KETONEN, Jussi; SOLOVAY, Robert. Rapidly growing Ramsey functions. **Annals of Mathematics**, 1981. DOI: https://doi.org/10.2307/2006985.

KIRBY, Laurie; PARIS, Jeff. Accessible Independence Results for Peano Arithmetic. **Bulletin of the London Mathematical Society**, v. 14, p. 285–293, 1982. ISSN 14692120. DOI: 10.1112/blms/14.4.285.

KLEENE, Stephen Cole. **Introduction to Metamathematics**. Nova Iorque: North-Holland, 1971. (Bibliotheca Mathematica, A Series of Monographs on Pure and Applied Mathematics). ISBN 9780444100887.

KLEENE, Stephen Cole. **Mathematical Logic**. Nova Iorque: Dover Publications, 2002. ISBN 0486425339.

KOELLNER, Peter. On the Question of Absolute Undecidability. **Philosophia Mathematica**, 2006. DOI: 10.1093/philmat/nkj009.

LAGARIAS, Jeffrey C. An Elementary Problem Equivalent to the Riemann Hypothesis. **The American Mathematical Monthly**, 2002. DOI: doi:10.2307/2695443.

MURAWSKI, Roman. **Recursive Functions and Metamathematics**. Dordrecht: Springer Netherlands, 1999. ISBN 9780792359043.

PINTER, Charles. **A Book of Set Theory**. Nova Iorque: Dover Publications, 2014. ISBN 9780486497082.

POTTER, Michael. **Set Theory and its Philosophy: A Critical Introduction**. Oxford: Oxford University Press, 2006. ISBN 9780199270415.

ROGERS, Robert. **Mathematical Logic and Formalized Theories**. Nova Iorque: North-Holland, 2015. ISBN 9781483249964.

ROGERS JR., Hartley. **Theory of Recursive Functions and Effective Computability**. Nova Iorque: The MIT Press, 1987. (McGraw-Hill Series in Higher Mathematics). ISBN 9780262680523.

SHOENFIELD, Joseph R. **Mathematical Logic**. Reading, MA: Addison-Wesley Publishing Company, 1967.

SIPSER, Michael. **Introduction to the Theory of Computation**. Boston: Thomson Course Technology, 2006. ISBN 9780534950972.

SMITH, Peter. **An Introduction to Gödel's Theorems**. [*S.l.*]: Cambridge University Press, 2013. ISBN 9781107606753.

SMORYŃSKI, Craig. Hilbert's Programme. **CWI Quartely 1**, p. 291–341, mai. 1988. DOI: 10.1090/hmath/033/10.

SMULLYAN, Raymond M. **Diagonalization and Self-Reference**. Midsomer Norton: Oxford University Press, 1994. (Oxford Science Publications). ISBN 198534507.

SMULLYAN, Raymond M. **First-Order Logic**. Nova Iorque: Dover Publications, 1995. ISBN 9780486683706.

SMULLYAN, Raymond M. **Forever undecided, a puzzle guide to Gödel**. Nova Iorque: Alfred A. Knopf, Inc., 1987. ISBN 9780394549439.

SMULLYAN, Raymond M. **Gödel's Incompleteness Theorems**. Nova Iorque: Oxford University Press, 1992. (Oxford Logic Guides). ISBN 9780195046724.

SMULLYAN, Raymond M.; FITTING, Melvin. **Set Theory and the Continuum Problem**. Oxford: Oxford University Press, 1996. ISBN 0198523955.

STILLWELL, John. **Roads to Infinity: The Mathematics of Truth and Proof**. Natick: A K Peter/CRC Press, 2010. ISBN 9781568814667.

SUPPES, Patrick. **Axiomatic Set Theory**. Nova Iorque: Dover Publications, 1972. ISBN 9780486616308.

WANG, Hao. **A logical journey: from Gödel to philosophy**. Cambridge: The MIT Press, 1997. ISBN 9780262231893,0262231891.

ZACH, Richard. **Incompleteness and Computability: An Open Introduction to Gödel's Theorems**. [*S.l.*]: Publicado de maneira independente, 2019. ISBN 9781077323391.

APÊNDICE A – NÚMEROS ORDINAIS

Desde a época da pré-escola, somos ensinados que os números naturais desempenham dois papéis: o ato de *contar* e o ato de *ordenar*. Assim, por exemplo, na frase “Tenho 27 anos de idade”, o número *cardinal* 27 expressa a *quantidade* de anos já vividos pelo autor, ao passo que em “Este é 11º mês que passo isolado em minha casa”, o número *ordinal* 11 expressa a *posição* que janeiro de 2021 ocupa numa lista de meses que o autor já passou em sua longa quarentena de Covid-19.

Tal distinção, a menos de semântica, parece um tanto quanto pedante: por exemplo, poderíamos reformular a segunda frase em termos de números cardinais: “Faz 11 meses que estou isolado em minha casa.” Ocorre que, após as descobertas de Cantor, tal distinção passou a ser, também, sintaticamente importante, pois ordinais e cardinais *infinitos* se comportam de maneiras bastante diversas (principalmente quanto a suas operações análogas).

A teoria que vamos construir neste apêndice é baseada na teoria dos conjuntos, sob a axiomática de ZFC (i.e., Zermelo-Fraenkel mais o axioma da escolha). Desse modo, embora não mais do que um entendimento basilar sobre a porção “ingênua” da mesma seja necessário, ficam subentendidos seus principais conceitos. Em particular, as definições de *relação de ordem parcial* e *ordem total*, bem como as construções teóricas advindas destas (como as ideias de pares ordenados, de elementos maximais e minimais, de supremos e ínfimos, etc.), serão de suma importância.[1] A maioria dos resultados e definições a seguir foram retirados dos textos de (JECH, 2014) e (PINTER, 2014) e (GALLIER, 1991). Por questão de escopo e espaço, omitiremos as provas.

Dado um conjunto parcialmente ordenado A, dizemos que A é *bem-ordenado* se todo subconjunto não vazio de A tiver um menor elemento. Nesse caso, a relação de ordem correspondente, $\leq$, que podemos provar ser uma ordem total, é dita uma *boa ordenação*. Por um *segmento inicial de A*, entende-se todo subconjunto na forma $\{x \in A : x \leq u\}$ para algum $u \in A$. Usando apenas esses conceitos, pode-se provar o seguinte teorema, que, por sua vez, motiva a construção dos números ordinais:

Teorema A.0.1. *Sejam A e B conjuntos bem-ordenados quaisquer. Então, uma (e apenas uma) das seguintes condições é satisfeita:*

(i) A e B são isomorfos (i.e., há uma bijeção entre A e B);

(ii) A é isomorfo a algum segmento inicial de B;

(iii) B é isomorfo a algum segmento inicial de A.

[1] Além disso, usaremos o conceito de *classe de objetos* — algo que ZFC, e tantas outras axiomáticas conjuntistas, em geral, não consideram. Assim, se $\varphi(x, y_1, ..., y_n)$ é uma fórmula, definimos $C = \{x : \varphi(x, y_1, ..., y_n)\}$, i.e., os elementos de uma classe C são todos os conjuntos para os quais determinada fórmula é válida em ZFC. Pelo Paradoxo de Russell, sabemos que C não pode ser um conjunto.

Tal resultado nos diz que conjuntos bem-ordenados são comparáveis por seus comprimentos (ou "tamanhos"). Se A e B são isomorfos, dizemos que eles têm o mesmo *tipo de ordem*. A ideia de se definir números ordinais é que eles sejam o *tipo de ordem* de um conjunto bem-ordenado. Além disso, para que tal definição esteja de acordo com o teorema A.0.1, devemos fazê-lo de modo que, se α e β representam ordinais quaisquer, então

$$\alpha \leq \beta \leftrightarrow (\alpha \in \beta \vee \alpha = \beta), \quad \text{e} \quad \alpha = \{\beta : \beta \leq \alpha\}.$$

Para tanto, defina A como sendo um *conjunto transitivo* se para qualquer $a \in A$ tal que a é um conjunto, tenha-se $a \subset A$. [Isso equivale a dizer que $A \subset P(A)$, i.e., que A contém cada um de seus elementos que são conjuntos como um subconjunto. Um exemplo de conjunto transitivo é $B = \{b, \{b\}\}$, pois $\{b\}$ (que pertence a B) é também um subconjunto de B, já que $b \in B$.] Assim, temos a seguinte

Definição A.0.1. Um conjunto transitivo é dito um *número ordinal* (ou, simplesmente, um *ordinal*) se e somente se for bem-ordenado pela relação de pertencimento, $\in$.

Como sugerido acima, usaremos letras gregas minúsculas para nos referirmos a ordinais. À *classe* de todos os ordinais (porque tal entidade é muito maior do que um mero "conjunto") atribuímos o símbolo $\mathcal{O}$. Também, definimos $\alpha \leq \beta$ se e só se $\alpha \in \beta$. Em posse disso, podemos provar vários resultados elementares.

Por exemplo, $0 = \emptyset$ é trivialmente um número ordinal, e todo ordinal é o conjunto dos ordinais menores do que ele mesmo. Dado α um ordinal qualquer, temos que $\alpha \cup \{\alpha\}$ também o é, e $\alpha \cup \{\alpha\} = \inf\{\beta : \alpha \leq \beta\}$. Assim, definimos $\alpha \cup \{\alpha\} = \alpha + 1$, i.e., o *sucessor* de α. Se $\alpha = \beta + 1$ para algum ordinal β, dizemos que que α é um *ordinal sucessor*; se α não é um ordinal sucessor, então α deve ser tal que $\alpha = \sup\{\beta : \beta \leq \alpha\} = \bigcup \alpha$, e o chamamos de um *ordinal limite*. Por definição, consideramos $\emptyset$ também como um ordinal limite, e colocamos $\sup \emptyset = 0$.

Com isso, podemos definir os números naturais em termos de ordinais: ao menor ordinal limite não-nulo, atribuímos o símbolo ω, e todo ordinal menor que ω será um *número natural* (ou, ainda, um *ordinal finito*). Sob essa ótica, temos que

$$0 = \emptyset;$$
$$1 = 0 + 1 = \emptyset \cup \{\emptyset\} = \{\emptyset\};$$
$$2 = 1 + 1 = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\};$$
$$3 = 2 + 1 = \{\emptyset, \{\emptyset\}\} \cup \{\{\emptyset, \{\emptyset\}\}\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}; ...$$

Ou seja, todo número natural, visto como um número ordinal, é simplesmente o resultado de alguma iteração da operação de união, a partir do conjunto vazio — exatamente o mesmo processo que utilizamos para definir $\mathbb{N}$ na linguagem de PA. A vantagem, aqui, é que tal definição já embute a relação de ordem que queremos.

Finalmente, definimos *A* como sendo um conjunto *finito* se existir uma função injetiva de *A* em algum $n \in \omega$. Caso contrário, *A* é dito *infinito*. Com isso e o próximo teorema, podemos facilmente definir o conceito de *sequência* de números ordinais:

Teorema A.0.2 (Indução Transfinita)**.** *Seja C uma classe de ordinais. Assumindo que*

(i) $0 \in C$*;*

(ii) se $\alpha \in C$*, então* $\alpha + 1 \in C$*; e*

(iii) se $\alpha \neq 0$ *é um ordinal limite tal que* $\beta \in C$ *para todo* $\beta \leq \alpha$*, então* $\alpha \in C$*,*

então C é a classe de todos os ordinais, i.e., $C = \mathcal{O}$*.*

Uma *sequência infinita*, portanto, será qualquer função *f* cujo domínio é ω. Uma *sequência finita* terá um domínio finito, i.e, *n* para algum $n \in \omega$. Nesse sentido, diremos que *s* é uma *sequência de comprimento n*, ou uma *n-sequência*. Já uma *sequência transfinita* terá como domínio um ordinal qualquer. [Portanto, sequências finitas e infinitas são todas transfinitas]. A notação será a mesma utilizada anteriormente: e.g., $s = \langle a_\beta : \beta \leq \alpha \rangle$, e para tal sequência também podemos dizer que seu *comprimento* é α, ou que ela é uma α*-sequência*.

A.1 ARITMÉTICA ORDINAL

Nosso objetivo, agora, é estabelecer um pouco de *aritmética* na teoria de números ordinais. Para isso, lançaremos mão de um método análogo àquele que nos permitiu construir (quase) todas as funções computáveis em PA. Chamaremos esse processo de *recursão transfinita*. A ideia é que sempre que tenhamos uma função *f* na classe das sequências transfinitas, deve existir, para todo ordinal γ, uma única γ-sequência $\langle a_\alpha : \alpha \leq \gamma \rangle$ de forma que

$$\forall \alpha \leq \gamma \, [a_\alpha = f(\langle a_\alpha : \alpha \leq \gamma \rangle)].$$

O seguinte teorema diz que isso sempre é possível:

Teorema A.1.1 (Recursão Transfinita)**.** *Seja f uma função qualquer. Se g for definida como sendo*

$$\begin{aligned} g(\alpha) = x \leftrightarrow \textit{ existe uma sequência } \langle a_\beta : \beta \leq \alpha \rangle \textit{ tal que:} \\ \textit{(i) } \forall \beta \leq \alpha [a_\beta = f(\langle a_\gamma : \gamma \leq \beta \rangle)]; \textit{ e} \\ \textit{(ii) } x = f(\langle a_\beta : \beta \leq \alpha \rangle), \end{aligned}$$

então g determina uma função em $\mathcal{O}$ *tal que para cada ordinal* α*, obtém-se*

$$g(\alpha) = f(g \restriction \alpha).$$

(Aqui, o símbolo "$\upharpoonright \alpha$" significa a restrição do domínio de g ao ordinal α.)

Chamemos qualquer sequência $\langle a_\beta : \beta \leq \alpha \rangle$ tal que se $\beta \leq \gamma$, então $a_\beta \leq a_\gamma$, de uma *sequência não-decrescente*. Juntando isso ao conceito de ordinal limite, podemos definir o que vem a ser o *limite de uma sequência transfinita*:

Definição A.1.1. Sejam $\alpha \geq 0$, $\alpha \neq 0$, um ordinal limite e $\langle a_\beta : \beta \leq \alpha \rangle$ uma sequência não-decrescente quaisquer. Então, o *limite de* a_α é tal que

$$\lim_{\beta \to \alpha} a_\beta = \sup\{a_\beta : \beta \leq \alpha\}.$$

Com isso, podemos definir adição, multiplicação e exponenciação de números ordinais. No que segue, perceba como a ideia é um paralelo perfeito com as definições apresentadas no capítulo 2 (a despeito do novo conceito de ordinal limite, que não tem paralelo na teoria de números).

Definição A.1.2 (Adição)**.** Dados α, β, γ ordinais quaisquer, temos que

(i) $\alpha + 0 = \alpha$;

(ii) $\alpha + (\beta + 1) = (\alpha + \beta) + 1$;

(iii) $\alpha + \beta = \lim_{\gamma \to \beta}(\alpha + \gamma)$, para todo ordinal limite $\beta \geq 0$.

Definição A.1.3 (Multiplicação)**.** Dados α, β, γ ordinais quaisquer, temos que

(i) $\alpha \cdot 0 = 0$;

(ii) $\alpha \cdot (\beta + 1) = \alpha \cdot \beta + \alpha$;

(iii) $\alpha \cdot \beta = \lim_{\gamma \to \beta}(\alpha \cdot \gamma)$, para todo ordinal limite $\beta \geq 0$.

Definição A.1.4 (Exponenciação)**.** Dados α, β, γ ordinais quaisquer, temos que

(i) $\alpha^0 = 1$;

(ii) $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$;

(iii) $\alpha^\beta = \lim_{\gamma \to \beta}(\alpha^\gamma)$, para todo ordinal limite $\beta \geq 0$.

Usando indução transfinita, vários resultados análogos aos da aritmética elementar podem ser obtidos. Em particular, adição e multiplicação em ordinais são operações associativas, e as interações das mesmas com a relação de ordem $\leq$ são bastante semelhantes (e.g., valem as leis de cancelamento). Além disso, existe um algoritmo de divisão totalmente análogo àquele de Euclides para a teoria de números. Diretamente desse resultado, segue um análogo à representação de números naturais numa base qualquer. Nesse caso, todo ordinal pode ser escrito, de forma única, como uma combinação linear de potências de ω:

Teorema A.1.2 (Forma Normal de Cantor)**.** *Dado $\alpha \geq 0$, existem únicos $n \in \mathbb{N}^*$, $\beta_1, ..., \beta_n$, com $\beta_i \leq \alpha$ e $\beta_j \geq \beta_\ell$ para todo $i \in \{1, ..., n\}$ e $j \leq \ell$, e $k_1, ..., k_n \in \mathbb{N}^*$ de forma que*

$$\alpha = \omega^{\beta_1} \cdot k_1 + ... + \omega^{\beta_n} \cdot k_n.$$

As semelhanças, todavia, param por aí. Veja, por exemplo, que as somas envolvendo ordinais *in*finitos geralmente não comutam. Com efeito: pela definição A.1.2, temos que

$$1 + \omega = \sup\{1 + \alpha : \alpha \leq \omega\},$$

e como todo ordinal menor do que ω é um número natural, deve-se ter que $1 + \alpha = \alpha + 1 = \beta$, para algum β que seja um ordinal sucessor. Assim, obtemos:

$$1 + \omega = \sup\{\beta : \beta \leq \omega\} = \omega.$$

Já o ordinal $\omega + 1$ é, por definição, igual a $\omega \cup \{\omega\}$, e

$$\omega \cup \{\omega\} = \{0, 1, 2, ..., \omega\} \neq \{0, 1, 2, ...\} = \omega.$$

O mesmo acontece com a multiplicação, pois, e.g.,

$$\begin{aligned} 2 \cdot \omega = \sup\{2 \cdot \alpha : \alpha \leq \omega\} = \omega \\ \neq \omega + \omega = \omega \cdot 1 + \omega = \omega \cdot (1 + 1) = \omega \cdot 2. \end{aligned}$$

Agora, tudo isso parece fazer muito sentido algebricamente — pois tais operações são definidas de maneiras análogas àquilo com que já estamos acostumados —, mas os contra-exemplos de comutatividade, acima, sugerem que nossa intuição não deve ser o caminho a se trilhar (ao menos não em situações envolvendo o infinito). No entanto, alguns casos particulares, limitados a ordinais menores ou iguais a ω, podem nos ajudar a entender o que as definições indutivas, acima, estão computando. Tome, por exemplo, a soma $\omega + \omega$. Procedendo apenas pela definição, seu resultado é o limite de $\omega + \gamma$, com γ tendendo ao ordinal ω. Em outras palavras, $\omega + \omega$ é o limite da sequência

$$\langle \omega, \omega + 1, \omega + 2, ..., \omega + n, ... \rangle.$$

Por outro lado, como ordinais são os *tipos de ordem* de algum conjunto bem-ordenado, podemos tomar dois conjuntos com tipo de ordem ω, e pensar em $\omega + \omega$ como sendo *o tipo de ordem da união desses conjuntos*.

Para tanto, precisamos encontrar A e B isomorfos a ω, com A e B disjuntos. Isso sempre pode ser feito: basta tomar, digamos, $A = \omega \times \{0\}$ e $B = \omega \times \{1\}$, ambos ordenados pela primeira entrada dos pares. Considerando as respectivas ordens em A e B como $\leq_A$ e $\leq_B$, definimos uma ordem em $A \cup B$ da seguinte forma:

(i) se $x, y \in A$, então $x \prec y$ quando $x \leq_A y$;

(ii) se $x, y \in B$, então $x \prec y$ quando $x \leq_B y$;

(iii) se $x \in A$ e $y \in B$, então $x \prec y$ é verdadeira;

(iv) se $x \in B$ e $y \in A$, então $x \prec y$ é falsa.

Com isso, pode-se provar que $\prec$ é uma boa-ordem para $A \cup B$. Logo, deve existir um ordinal α tal que $A \cup B \cong \alpha$; assim, temos que $\alpha = \omega + \omega$. Nesse caso, tomamos

$$A = \{0, 1, 2, ...\} \times \{0\} = \{\{0, 0\}, \{1, 0\}, \{2, 0\}, ...\};$$
$$B = \{0, 1, 2, ...\} \times \{1\} = \{\{0, 1\}, \{1, 1\}, \{2, 1\}, ...\},$$

de forma que $A \cup B = \{\{0, 0\}, \{1, 0\}, ..., \{0, 1\}, \{1, 1\}, ...\}$. O significado de $\omega + \omega = \omega \cdot 2$, então, se torna bastante intuitivo: *ele representa a listagem, na ordem $\prec$, de duas cópias isomorfas de ω*.

É claro que extrapolar esse tipo de análise para ordinais maiores do que ω torna-se uma tarefa árdua (tanto pela não-intuitividade quanto pela crescente falta de símbolos disponíveis para representar tais números, à medida que se opera com ordinais cada vez maiores); nesse sentido, as definições A.1.2-A.1.4 tornam tudo mais fácil. Voltemos nossa atenção, portanto, a ordinais *contáveis* — i.e., àqueles ordinais atribuídos a conjuntos bem-ordenados que são isomorfos a $\mathbb{N}$. (São exemplos destes os de maior interesse para o nosso trabalho no capítulo 4.)

A.2 ORDINAIS CONTÁVEIS

Vamos restringir nosso ambiente de estudo da classe dos ordinais, $\mathcal{O}$, à classe dos ordinais contáveis, $\mathcal{C}$. Lembrando que um subconjunto $B \subset A$ é *estritamente limitado* quando existe $a \in A$ tal que para todo $b \in B$, tem-se $b \leq a \wedge b \neq a$, a caracterização de $\mathcal{C}$ se dá pelos seguintes axiomas:

(i) $\mathcal{C}$ é bem-ordenada pela relação $\leq$ usual para ordinais;

(ii) Todo $A \in \mathcal{C}$ que for um conjunto estritamente limitado (como subconjunto de $\mathcal{C}$) deve ser um ordinal contável;

(iii) Todo $\alpha \in \mathcal{C}$ que for um ordinal contável deve ser um conjunto estritamente limitado (como subconjunto de $\mathcal{C}$).

Com isso, podemos provar que, embora $\mathcal{C}$ não seja uma classe própria (de modo que podemos chamá-la de um conjunto) $\mathcal{C}$ é *in*contável — i.e., existem incontáveis ordinais contáveis. Mais importante que isso, o teorema A.0.2, de indução transfinita, vale trivialmente para $\mathcal{C}$. Logo, podemos desenvolver a mesma aritmética de $\mathcal{O}$ para o

contexto restrito desta seção. Em particular, todo ordinal contável tem sua respectiva forma normal de Cantor.

Estendendo a noção de segmento inicial, dado $A \subset \mathcal{C}$, chamamos A de um $\mathcal{C}$*-segmento* se e somente se para quaisquer $\alpha, \beta \in \mathcal{C}$, se $\beta \in A$ e $\alpha < \beta$, então $\alpha \in A$. Com isso, definimos o importantíssimo conceito de função de ordenamento:

Definição A.2.1. Dado $B \subset \mathcal{C}$, a função $f : A \longrightarrow B$ será dita uma *função de ordenamento de B* se, e somente se,

(i) o domínio de f for um $\mathcal{C}$-segmento;

(ii) Para quaisquer $\alpha, \beta \in \mathcal{C}$, se $\alpha < \beta$, então $f(\alpha) < f(\beta)$;

(iii) A imagem de f for igual a B.

Nesse sentido, como a cláusula (ii) afirma que f deve ser estritamente crescente, o que uma tal função faz é enumerar os elementos de B em ordem crescente. Fato é que todo ordinal contável tem uma função de ordenamento respectiva, e essa função é única. Um atributo importante dessas funções é que elas são *contínuas*, no seguinte sentido: primeiramente, chamemos A de um *conjunto fechado* de ordinais quando se verificar que

$$B \subset A \rightarrow \bigcup B \in A,$$

qualquer que seja o ordinal contável B, desde que não-vazio. Daí, temos a

Definição A.2.2. Seja $f : A \longrightarrow B$ uma função de ordenamento de B. Então, f será *contínua* se e somente se A for fechado e para qualquer ordinal não-vazio $M \subset A$, for verdade que

$$f\left(\bigcup M\right) = \bigcup f(M).$$

Com isso, podemos caracterizar continuidade em termos de contradomínio fechado, i.e.,

Teorema A.2.1. *Seja $f : A \longrightarrow B$ uma função de ordenamento de B. Então, f será contínua se e somente se B for fechado.*

Se f for contínua e $A = \mathcal{C}$, diremos que f é uma *função normal*. Funções normais são importantes na construção de hierarquias (que, por sua vez, determinam ordinais contáveis gigantescos e bastante relevantes na discussão sobre incompletude concreta, tema do capítulo 4). Tais funções têm a seguinte caracterização:

Teorema A.2.2. *Seja $f : A \longrightarrow B$ uma função de ordenamento de B. Então, f será normal se e somente se B for fechado e ilimitado (i.e., não estritamente limitado).*

Para os nossos propósitos, mais importante do que o conceito de normalidade é o fato de que toda função contínua (ou normal) possui pontos fixos, i.e., ordinais α tais que se f é contínua (normal), então $\alpha = f(\alpha)$. Isso se dá, de forma construtiva, através dos seguintes resultados:

Teorema A.2.3. *Seja $f : \mathcal{C} \longrightarrow \mathcal{C}$ uma função contínua. Para cada $\alpha \in \mathcal{C}$, defina $f^0(\alpha) = \alpha$ e, para cada $n \geq 0$, $f^{n+1}(\alpha) = f(f^n(\alpha))$. Se $\alpha \leq f(\alpha)$ para qualquer $\alpha \in \mathcal{C}$, então $\bigcup_{n\geq 0} f^n(\alpha)$ é um ponto fixo de f.*

Corolário A.2.1. *Seja $f : A \longrightarrow B$ uma função normal, e tome $\alpha \in \mathcal{C}$. Então, $\bigcup_{n\geq 0} f^n(\alpha)$ é um ponto fixo de f.*

Chamemos um ordinal não nulo $\alpha \in \mathcal{C}$ de um *ordinal aditivo principal* quando para cada $\beta < \alpha$, tivermos $\beta + \alpha = \alpha$. (Observe que 1 é o menor ordinal aditivo principal, e que o próximo da lista é ω.) Ao conjunto de todos os ordinais aditivos principais, atribuímos o signo $\mathcal{A}$. Não é muito difícil demonstrar que $\mathcal{A}$ é fechado e ilimitado; portanto, pelo teorema A.2.2, a função de ordenamento de $\mathcal{A}$ é normal.

Tal função é dada por $f : \mathcal{C} \longrightarrow \mathcal{A}$, $f(\alpha) = \omega^\alpha$. (Para verificar que é esse o caso, basta checar os axiomas (i)-(iii) da definição A.2.1; a parte mais complicada é seu caráter crescente, i.e., que $\alpha < \beta \rightarrow \omega^\alpha < \omega^\beta$.) Por ser normal, f possui pontos fixos. Ou seja, existem ordinais α para os quais $\alpha = \omega^\alpha$. O menor desses ordinais é chamado de ε_0 (teremos mais oportunidades de falar dele adiante).

Introduzimos, agora, para cada ordinal α, o importante conceito de *ordinais α-críticos*, e definimos suas funções de ordenamento. Chamamos o conjunto de todos os ordinais α-críticos de $\mathrm{Cr}(\alpha)$, e as respectivas funções de ordenamento serão nomeadas g_α, com $g_\alpha : X_\alpha \longrightarrow \mathrm{Cr}(\alpha)$. Procedemos indutivamente da seguinte forma:

(1) $\mathrm{Cr}(0) = \mathcal{A}$, $X_0 = \mathcal{C}$, e para cada $\beta \in \mathcal{C}$, $g_0(\beta) = \omega^\beta$, i.e., $g_0(\beta)$ é a função de ordenamento de $\mathrm{Cr}(0) = \mathcal{A}$;

(2) $\mathrm{Cr}(\alpha + 1) = \{\eta \in X_\alpha : g_\alpha(\eta) = \eta\}$, i.e., o conjunto dos pontos fixos de g_α, e $g_{\alpha+1} : X_{\alpha+1} \longrightarrow \mathrm{Cr}(\alpha + 1)$ é a função de ordenamento de $\mathrm{Cr}(\alpha + 1)$;

(3) Para qualquer ordinal limite $\beta \in \mathcal{C}$, tem-se que

$$\mathrm{Cr}(\beta) = \left\{\eta \in \bigcap_{\alpha<\beta} X_\alpha : \forall \alpha < \beta [g_\alpha(\eta) = \eta]\right\},$$

i.e., o conjunto dos pontos fixos comuns a todos os $\mathrm{Cr}(\alpha)$, com $\alpha < \beta$; por sua vez, $g_\beta : X_\beta \longrightarrow \mathrm{Cr}(\beta)$ é a função de ordenamento de $\mathrm{Cr}(\beta)$.

Como indicado pelas cláusulas acima, os elementos de $\mathrm{Cr}(\alpha)$ são os pontos fixos das funções de ordenamento g_β, para todo $\beta < \alpha$. Usando argumentos de indução transfinita, podemos provar que cada $\mathrm{Cr}(\alpha)$ é fechado e ilimitado; mais ainda, sempre

se tem $X_\alpha = \mathcal{C}$. Por conseguinte, cada função de ordenamento g_α tem seus valores avaliados em todos os ordinais, e g_α é sempre normal. Logo, toda g_α tem pontos fixos.

Agora, pelo fato de o domínio de g_α sempre ser $\mathcal{C}$, podemos construir uma função binária $g : \mathcal{C} \times \mathcal{C} \longrightarrow \mathcal{C}$ tal que $g(\alpha, \beta) = g_\alpha(\beta)$. Disso, segue imediatamente que todo ordinal $g(\alpha, \beta)$ é um ordinal aditivo principal, que $g(0, \beta) = \omega^\beta$, e que se $\alpha < \beta$, então $\mathrm{Cr}(\beta) \subset \mathrm{Cr}(\alpha)$.

Uma questão de maior interesse (em nosso contexto) é encontrar ordinais α tais que $\alpha \in \mathrm{Cr}(\alpha)$, i.e. ordinais α que sejam α-críticos. Indo diretamente pela definição, isso significaria que α é um ponto fixo comum a todas as funções normais g_γ, com $\gamma < \alpha$. Ilustrativamente, se ε_0 fosse um ordinal ε_0-crítico, deveríamos ter $\varepsilon_0 = g_\gamma(\varepsilon_0)$, qualquer que fosse $\gamma < \varepsilon_0$. Mas, como já apontamos, ε_0 é o limite da sequência dada por $f(\alpha) = \omega^\alpha$, logo ele não pode ser ponto fixo de funções de ordenamento cujos ordinais correspondentes estão todos abaixo dele.

Isso demonstra a dificuldade de se encontrar ordinais com tal propriedade. No entanto, o último resultado acima,

$$\alpha < \beta \rightarrow \mathrm{Cr}(\beta) \subset \mathrm{Cr}(\alpha),$$

nos mostra que a tarefa é, ao menos, possível: para que seja $\alpha \in \mathrm{Cr}(\alpha)$, precisamos encontrar ordinais α e β tais que (i) $\alpha < \beta$ e (ii) $\alpha \in \mathrm{Cr}(\beta)$. O fato de ε_0 não satisfazer essa implicação sugere que precisamos de ordinais (ainda) maiores, o que deve demandar ferramentas mais complexas do que as que vimos até aqui (sem considerar o fato de que cada vez mais símbolos são necessários para representar tais números, e que, em algum momento, as computações se tornam ininteligíveis). De qualquer maneira, a princípio, esses ordinais *podem* existir — desde que encontremos ordinais que satisfaçam as condições (i) e (ii) acima. Também a princípio, nada garante que eles realmente existam, mas a verdade é que há uma quantidade incontável deles. Ao menor de todos estes, atribuímos o símbolo Γ_0.

Na tentativa de captar o tamanho absurdo desse ordinal, primeiramente vamos analisar alguns menores. Por definição, em Cr(1) moram todos os ordinais α tais que $\omega^\alpha = \alpha$. A eles, damos o nome de *ordinais epsilon*, e o menor destes se chama ε_0. Usando a definição de limite de sequências, pode-se provar que ε_0 é a menor cota superior de

$$\langle \omega, \omega^\omega, \omega^{\omega^\omega}, ..., \omega^{\omega^{\cdot^{\cdot^{\cdot^\omega}}}}, ... \rangle.$$

Seu tamanho é, inegavelmente, muito impressionante. Mas basta subirmos um degrau nessa hierarquia que encontramos ordinais mais impressionantes ainda. Em Cr(2), temos pontos fixos da função ϵ que ordena os elementos de Cr(1), i.e., os α tais que $\epsilon(\alpha) = \alpha$. Como anteriormente, prova-se que o menor desses ordinais é o limite da sequência

$$\langle \varepsilon_0, \varepsilon_1, ...\varepsilon_\omega, ..., \varepsilon_{\varepsilon_0}, ..., \varepsilon_{\varepsilon_1}, ..., \varepsilon_{\varepsilon_{\varepsilon_0}}, ... \rangle.$$

Para se ter uma ideia de sua magnitude, o segundo termo da sequência, ε_1, é o limite da sequência

$$\langle \varepsilon_0 + 1, \omega^{\varepsilon_0+1}, \omega^{\omega^{\varepsilon_0+1}}, ... \rangle.$$

Ocorre que Γ_0 é *muito* maior do que esses exemplos — na verdade, *incontavelmente* maior, no sentido de que as iterações necessárias do processo acima somam uma quantidade não-enumerável de passos. Mas voltemos às construções envolvendo as funções $g(\alpha, \beta)$, pois elas são a chave para entender o significado (por enquanto, muito difuso) de Γ_0. Primeiramente, pode-se provar o seguinte resultado:

Teorema A.2.4. *Dado γ um ordinal aditivo principal, existem únicos $\alpha, \beta \in \mathcal{C}$ tais que $\alpha \leq \gamma$, $\beta < \gamma$ e $g(\alpha, \beta) = \gamma$.*

Como, na hipótese, temos $\alpha \leq \gamma$, com certeza existe um ordinal γ tal que $g(\gamma, \beta) = \gamma$. Isso acontece se e somente se $\gamma \in \mathrm{Cr}(\gamma)$. Nesse sentido, chamamos cada um desses ordinais de um *ordinal fortemente crítico*. O seguinte teorema caracteriza tais ordinais:

Teorema A.2.5. *Dado $\alpha \in \mathcal{C}$, tem-se que α é fortemente crítico se e somente se $g(\alpha, 0) = \alpha$.*

Assim, podemos definir uma função $h : \mathcal{C} \longrightarrow \mathcal{C}$ tal que $h(\alpha) = g(\alpha, 0)$. Isso nos permite construir o conjunto $G = \{g(\alpha, 0) : \alpha \in \mathcal{C}\}$, e, provando que tal conjunto é fechado e ilimitado, concluímos que sua função de ordenamento é uma função normal. Em particular, isso prova que existem infinitos ordinais fortemente críticos.

Chamando a função normal de G pelo nome Γ, temos que $\Gamma(0) = \Gamma_0$, i.e., o menor ordinal fortemente crítico (cuja existência já havíamos previsto). A seguinte proposição evidencia uma característica bastante peculiar de tais ordinais:

Teorema A.2.6. *Dados $\alpha, \beta \in \mathcal{C}$ quaisquer, se $\alpha < \Gamma_0$ e $\beta < \Gamma_0$, então,*

$$\alpha + \beta < \Gamma_0, \quad e \quad g(\alpha, \beta) < \Gamma_0.$$

Isso nos diz que Γ_0 — e, por conseguinte, qualquer ordinal fortemente crítico — não pode ser obtido de ordinais menores que ele através da operação elementar +, nem através das funções g_α. Assim, em certo sentido, podemos dizer que os ordinais obtidos pela função Γ são análogos a ordinais *in*contáveis.

Mais especificamente, a definição de Γ_0 (i.e., "o *menor* ordinal α tal que $\alpha = g(\alpha, 0)$") pressupõe o universo $\mathcal{O}$ inteiro, e não apenas $\mathcal{C}$. Isso configura, portanto, uma definição circular, ou *impredicativa*. [Lembre-se de nossa discussão sobre impredicatividade no primeiro capítulo. Segundo os intuicionistas, era esse o fenômeno

responsável por todos os paradoxos dos fundamentos da Matemática. No entanto, como já apontamos, nem todo conceito circular gera contradições; a definição de Γ_0, até onde se sabe, não é conflitante com a teoria dos números ordinais. Mesmo assim, seu *status* impredicativo, por pressupor a totalidade de $\mathcal{O}$, também pressupõe ordinais incontáveis, e isso parece atestar quão gigantesco ele é.]

Outro fato que evidencia a impredicatividade de Γ_0 é que, na contramão do teorema acima, *todos os ordinais menores que* Γ_0 podem ser obtidos de ordinais menores através da adição e das funções g_α. Isso se lê no seguinte resultado, que segue imediatamente da Forma Normal de Cantor e do fato que para todo ordinal aditivo principal γ, com $\gamma \leq \alpha$, $\gamma < \beta$ e $\gamma = g(\alpha, \beta)$, tem-se que $\gamma < \alpha$ se e só se γ *não é* um ordinal crítico (vide (GALLIER, 1991), p. 234):

Teorema A.2.7. *Dado qualquer ordinal* η*, com* $0 < \eta < \Gamma_0$*, existem únicos ordinais* $\alpha_1, ..., \alpha_n$*,* $\beta_1, ..., \beta_n$*, com* $n \in \mathbb{N}^*$*, tais que para cada* $i \in \{1, .., n\}$*,* $\alpha_i < g(\alpha_i, \beta_i) < \eta$*,* $\beta_i < g(\alpha_i, \beta_i) < \eta$*, e se verifica que*

(i) $\eta = g(\alpha_1, \beta_1) + ... + g(\alpha_n, \beta_n)$*; e*

(ii) $g(\alpha_1, \beta_1) \geq ... \geq g(\alpha_n, \beta_n)$.

Desse teorema, podemos derivar (via indução em n na decomposição acima, no item (i).) uma caracterização recursiva de todos os ordinais menores que Γ_0. Nesse sentido, há três possibilidades para um ordinal $\gamma < \Gamma_0$: ou (i) $\gamma = 0$; ou (ii) $\gamma = \beta + \alpha$, com $\alpha < \gamma$, $\beta < \gamma$ e $\alpha \leq \beta$; ou (iii) $\gamma = g(\alpha, \beta)$, com $\alpha < \gamma$ e $\beta < \gamma$. Em última instância, isso corrobora a afirmação, acima, quanto à representação de ordinais menores que Γ_0 através de $0, +$ e g.

Tal representação, no entanto, tem algumas desvantagens. Uma delas é o fato de que o mesmo ordinal pode ser representado por notações distintas, embora sua decomposição, pelo teorema A.2.7, seja única. Por exemplo, temos que $\varepsilon_0 = g(0, \varepsilon_0) = g(1, 0)$. O ideal seria encontrar uma função g' tal que ambos α e β sejam estritamente menores que $g'(\alpha, \beta)$. Isso pode ser obtido se restringirmos nossa atenção a ordinais α-críticos maximais.

Definição A.2.3. Um ordinal $\gamma \in \mathcal{O}$ é dito α*-crítico maximal* quando $\gamma \in \mathrm{Cr}(\alpha)$, mas para qualquer $\beta \geq \alpha$, tem-se que $\gamma \notin \mathrm{Cr}(\beta)$.

A função g'_α, então, é definida como sendo a função ordenadora do conjunto dos ordinais α-críticos maximais. Da mesma forma que antes, colocamos $g'(\alpha, \beta) = g'_\alpha(\beta)$, para todo ordinal β. Com isso, obtêm-se propriedades análogas às de g; em particular, para todo ordinal aditivo principal γ, existem $\alpha < \gamma$ e $\beta < \gamma$ tais que $\gamma = g'(\alpha, \beta)$. Mais importante que isso é o fato que para quaisquer $\alpha < \Gamma_0$ e $\beta < \Gamma_0$,

$$\alpha < g'(\alpha, \beta), \quad \text{e} \quad \beta < g'(\alpha, \beta).$$

Isso estabelece o desejado, pois da propriedade de ordinais aditivos acima mencionada segue uma nova versão do teorema *A*.2.7, dessa vez com respeito a g':

Teorema A.2.8. *Dado qualquer ordinal* η*, com* $0 < \eta < \Gamma_0$*, existem únicos ordinais* $\alpha_1, ..., \alpha_n$*,* $\beta_1, ..., \beta_n$*, com* $n \in \mathbb{N}^*$*, tais que para cada* $i \in \{1, .., n\}$*,* $\alpha_i < g'(\alpha_i, \beta_i) < \eta$*,* $\beta_i < g'(\alpha_i, \beta_i) < \eta$*, e se verifica que*

(i) $\eta = g'(\alpha_1, \beta_1) + ... + g'(\alpha_n, \beta_n)$*; e*

(ii) $g'(\alpha_1, \beta_1) \geq ... \geq g'(\alpha_n, \beta_n)$.

Como apontado, esse resultado permite criar um sistema de notações em que diferentes notações representam ordinais distintos. Mas não abordaremos, neste trabalho, o que vem a ser um tal sistema. Nosso maior interesse está na relação do ordinal Γ_0 com o Teorema de Kruskal, e a essência dessa relação está na representação recursiva dos ordinais menores que Γ_0. Deixamos esses comentários para o capítulo 4, que é o local destinado à discussão de indecidibilidade via ordinais contáveis.