LIMITAÇÕES DE SISTEMAS FORMAIS E A INTELIGÊNCIA ARTIFICIAL

Wolney Borde

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM HISTÓRIA DAS CIÊNCIAS E DAS TÉCNICAS E EPISTEMOLOGIA.

Aprovada por:

Prof. Cláudio T. Bornstein, Dr. Rer.

Prof. Ricardo Silva Kubrusly, Ph. D.

Prof. Carlos Antônio de Moura, Ph. D.

RIO DE JANEIRO, RJ – BRASIL

ABRIL DE 2005

BORDE, WOLNEY

Limitações de Sistemas Formais e a Inteligência Artificial [Rio de Janeiro] 2005

V, 95 p. 29,7 cm (COPPE/UFRJ, M. Sc., História das Ciências e das Técnicas e Epistemologia, 2005)

Tese – Universidade do Rio de Janeiro, COPPE

1. Limitações de Sistemas Formais

2. Inteligência Artificial

I. COPPE/UFRJ II. Título (série)

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M. Sc.)

LIMITAÇÕES DE SISTEMAS FORMAIS E A INTELIGÊNCIA ARTIFICIAL

Wolney Borde

Abril/2005

Orientador: Cláudio T. Bornstein

Programa: História das Ciências e das Técnicas e Epistemologia

Este trabalho apresenta um relato histórico da descoberta das limitações intrínsecas de sistemas formais em relação à incompleteza e à prova de consistência. Discute também como estas limitações são usadas em argumentações sobre a possibilidade de uma inteligência artificial similar à inteligência humana. As limitações de sistemas formais foram demonstradas por Kurt GÖDEL em 1931 e relacionam-se à questão da inteligência artificial devido a uma identificação entre os sistemas formais e a máquina de TURING, que representa o modelo básico de funcionamento dos computadores digitais.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M. Sc.)

LIMITATIONS OF FORMAL SYSTEMS AND ARTIFICIAL INTELLIGENCE

Wolney Borde

April/2005

Advisor: Cláudio T. Bornstein

Department: History of Science and of Technics and Epistemology

This work presents a historical account of the discovery of intrinsic limitations of formal systems. These limitations are related to incompleteness and to the proof of consistency. This work discuss also how these limitations are used in argumentations about the possibility of an artificial intelligence similar to the human intelligence. The limitations of formal systems were proved by Kurt GÖDEL in 1931 and are related to the issue of artificial intelligence due to an identification between formal systems and the TURING machine. The TURING machine represents the basic model of functioning for digital computers.

# SUMÁRIO


**INTRODUÇÃO** .................................................................... 1

**I - Lógica, Paradoxos e o Formalismo** ........................................ 6

1.1 – Desenvolvimento da antiguidade até o século XIX ........... 6
1.2 – A renovação da lógica no século XIX ........................... 11
1.3 – O *Grundlagen* de Frege .................................................... 12
1.4 – Russell, seu paradoxo e os *Principia Mathematica* .......... 15
1.5 – Hilbert e o formalismo .................................................... 19

**II - Os Resultados de Gödel e a Máquina de Turing** .................... 27

2.1 – O primeiro resultado ........................................................ 32
2.2 – O segundo resultado ........................................................ 37
2.3 – A máquina de Turing ...................................................... 41

**III - A Argumentação de Lucas** ...................................................... 48

3.1 – As geometrias não-euclideanas .......................................... 48
3.2 – A discussão de Lucas ...................................................... 53
3.3 – Aplicação da proposta de Lucas às geometrias não-euclideanas .................................................................. 54
3.4 – Outros pontos colocados por Lucas .................................. 56
3.5 – A contra-argumentação de Hofstadter ............................... 57
3.6 – Respostas de Lucas .......................................................... 60

**IV - O Argumento de Penrose e Contra-argumentações** ............. 62

4.1 – Criticas e contra-argumentações ....................................... 75
4.2 – A argumentação de Redhead ............................................ 78

**CONCLUSÃO** ........................................................................ 83

**BIBLIOGRAFIA** ...................................................................... 87

# INTRODUÇÃO

O advento dos computadores vem causar uma mudança de direção no modo como o homem inventa ferramentas e máquinas para ampliar seu poder de atuação no mundo. Isto porque as primeiras ferramentas e os primeiros tipos de máquinas servem para ampliar a habilidade manual, estender os sentidos ou multiplicar a força que se pode exercer sobre os objetos. Ou seja, estes artefatos ampliam o que o homem é capaz de fazer com o seu corpo. Mas, no caso dos computadores, as habilidades envolvidas são aquelas que o homem realiza com seu cérebro. O advento de propostas envolvendo a inteligência artificial torna esta mudança de direção ainda mais nítida, pois o sucesso destas propostas implica em máquinas que realmente "pensam".

Como todo processo de mudança, o surgimento da inteligência artificial suscitou uma série de questionamentos, como por exemplo: podemos considerar processos mentais como produto da execução de um algoritmo? Este questionamento certamente traz embutidos muitos outros, como, por exemplo, se todos os algoritmos são obrigatoriamente baseados numa lógica binária, como os que conhecemos, ou se haveria outro tipo de algoritmo.

RUSSELL e NORVIG (1995) situam o nascimento da IA no ano de 1943, com um trabalho de Warren McCULLOCH e de Walter PITTS. Neste trabalho estes propuseram um modelo de neurônios artificiais no qual cada neurônio funcionava de modo binário, ou seja, podia estar "ligado" ou "desligado". O estado de cada neurônio era determinado pela estimulação de um número suficiente de neurônios vizinhos. Para este trabalho, McCULLOCH e PITTS reuniram conhecimentos da psicologia, da fisiologia dos neurônios, da lógica de RUSSELL e WHITEHEAD, e da teoria da computação que TURING começava a criar.

Procurando um enfoque histórico mais abrangente, optou-se, no presente estudo, a começar o relato desta história numa época bem mais remota. É o que encontramos no conteúdo do capítulo I, que apresenta um relato sucinto da evolução da lógica e de alguns problemas e questões que marcaram seu desenvolvimento. Este capítulo termina relatando os acontecimentos do período situado na virada do século XIX para o século XX. Neste período, os matemáticos - mais precisamente os seguidores da escola formalista - acreditavam que seria possível ter um método para decidir, para qualquer proposição aritmética, se esta é verdadeira ou falsa. Ou seja, através de sistemas formais que representassem a aritmética chegaria-se a ter um procedimento de decisão (*decision procedure*) para dizer se uma proposição qualquer sobre números inteiros é um teorema ou não. DELONG (1971 p.132) define procedimento de decisão como "*qualquer método finito eficaz pelo qual pode-se determinar se uma fórmula arbitrária de um sistema formal é um teorema ou não*"[1]. O autor ainda esclarece que este "método finito eficaz" tem a mesma função de um algoritmo.

O estudo dos sistemas formais e das manipulações simbólicas feitas nas sentenças destes sistemas foi promissor. Tanto que produziu, nos anos 20, alguns procedimentos de decisão pelos quais classes inteiras de teoremas podiam ser verificadas (GÖDEL, 1992 p. 3). De fato, em 1930, PRESBURGER (idem) publica um procedimento de decisão aplicável a um sistema que representa a aritmética com a operação de adição, mas sem a de multiplicação, provando assim que toda proposição, nesta aritmética "mutilada", é decidível.

Os aspectos de completeza[2] e consistência de um sistema formal são conceitos-chave para o que se pretende discutir neste trabalho. CHAITIN (1990, 1993 e 2002) define um sistema consistente como aquele no qual não é possível provar falsos teoremas, ou como aquele que não permite demonstrar uma determinada proposição e também sua negação. RADU (2003 p.374) conta que o problema da consistência da geometria é resolvido mostrando-se que o mesmo pode ser reduzido ao da consistência da aritmética, cuja demonstração será um dos objetivos do programa formalista.

A importância destes aspectos nos leva a questões da fundamentação da matemática e ao programa formalista de HILBERT (1862-1943) que é comentado no último item do capítulo I.

---

[1] Any effective finite method by which it can be determined whether or not an arbitrary formula of a formal system is a theorem

[2] Alguns autores usam os termos "completude"/"incompletude" e outros os termos "completeza"/"incompleteza". Neste trabalho optou-se pela segunda forma por estar de acordo com o dicionário Aurélio da língua portuguesa

Mais tarde, os resultados a que GÖDEL (1906-1978) chegou, e que são conhecidos como "teorema de GÖDEL" vieram frustrar esta pretensão. Mas apesar de estes resultados terem ficado famosos no singular, isto é, como um único teorema, GÖDEL prova dois teoremas neste trabalho, que é apresentado na Academia de Ciências de Viena em 1930 e publicado no *Monatshefte für Mathematik und Physik* em 1931 (GÖDEL, 1992 p.1) sob o título de *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme 1* (Sobre proposições formalmente não-decidíveis do Principia Mathematica e sistemas similares 1). Estes resultados são apresentados no capítulo II.

Os resultados de GÖDEL geram uma série de reflexões e uma nova postura na lógica, principalmente na escola formalista. Incitam, também, reflexões e debates filosóficos que perduram até hoje. Além disso, devido à similaridade entre sistemas formais e máquinas de TURING (GÖDEL, 2001a p. 194-195), os resultados de GÖDEL tornaram-se importantes como argumentos contra a possibilidade de uma inteligência artificial "forte". Curiosamente, os mesmos resultados são usados por alguns autores de modo inverso, ou seja, como argumento a favor da inteligência artificial "forte". O termo "forte" usado aqui segue uma classificação feita por PENROSE (1996 p.12) para os níveis em que sistemas computacionais conseguiriam imitar a inteligência humana, sendo o termo "forte" utilizado para designar o mais alto nível, isto é, uma imitação completa da mente humana por um sistema computacional. Como a máquina de TURING representa o modelo básico de funcionamento dos computadores digitais, estes sofrem as limitações expostas pelos resultados de GÖDEL.
Uma descrição informal da máquina de TURING e suas limitações são apresentados também no capítulo II.

A discussão em torno da possibilidade de uma "IA forte" intensificou-se desde que LUCAS publicou em 1961 o artigo "*Minds, machines and Gödel*" (LUCAS, 1964), gerando críticas ferozes. Tanto que em palestra na "*Turing Conference*" em 1990 LUCAS (2005) diz que deve ter tocado em algum "nervo exposto" ao argumentar contra a possibilidade da "IA forte". A argumentação de LUCAS é discutida no capítulo III deste trabalho. Entre outras considerações, LUCAS comenta uma tentativa proposta por HEMPEL e ROGERS para tornar um sistema formal completo ao fazê-lo capaz de aceitar axiomas “provisórios”. Como veremos, esta tentativa é infrutífera, o que mostra o próprio LUCAS.

Ainda no capítulo III, exemplificando e estendendo a argumentação de LUCAS, foi colocada uma argumentação tendo como base a questão das geometrias não-euclideanans e sua formalização. O objetivo é apresentar alguns argumentos acerca da dificuldade de simular a criatividade humana admitindo-se os “axiomas provisórios”. Usando a história das geometrias não-euclideanas veremos que a alteração no significado de um ou mais axiomas geram mundos inteiramente novos, e, conforme o segundo resultado de GÖDEL, um sistema formal fica impossibilitado de gerar tais mundos sem colocar em risco sua consistência. A aplicabilidade das geometrias euclideana e não-euclideanas às ciências naturais, ou a importância dos métodos da análise matemática no desenvolvimento e aceitação das mesmas, por importante que sejam, não têm papel relevante para o presente estudo. Estamos aqui interessados na relação entre a alteração ou substituição de axiomas na geometria e o comportamento dos sistemas formais em que a mesma pode ser caracterizada.

Defendendo tese oposta à de LUCAS, HOFSTADTER publica “*Gödel, Escher e Bach*” em 1979, uma das mais inspiradas defesas da IA (HOFSTADTER, 1980). Nesta obra, HOFSTADTER critica diretamente os argumentos de LUCAS. As colocações de HOFSTADTER e a contra-argumentação de LUCAS foram incluídas também no capítulo III.

Dez anos mais tarde, PENROSE publica “A mente nova do rei” (*The emperor’s new mind*), obra que, assim como a de HOFSTADTER, consegue ser um sucesso editorial. PENROSE retoma a argumentação de LUCAS no sentido de usar os resultados de GÖDEL contra a possibilidade de uma “IA forte”. Mas, se a argumentação de PENROSE supera a de LUCAS em detalhamento e pretensão, também a supera no número e na intensidade das criticas e contra-argumentações. Tanto que tais criticas levaram PENROSE a mais duas publicações: “*Shadows of the mind*” (PENROSE, 1996b) e “O grande, o pequeno e a mente humana” (PENROSE, 1998). Além disso, a revista eletrônica PSYCHE promove um simpósio (PENROSE, 1996a) somente para discutir as argumentações, réplicas e tréplicas relativas às idéias de

PENROSE. O capítulo IV desta pesquisa apresenta a argumentação de PENROSE e de alguns de seus comentadores.

Este trabalho faz um relato histórico dos fatos que levaram GÖDEL ao seu famoso trabalho de 1931 e da relação entre seus resultados e a teoria da computação. Quanto aos argumentos pró e contra a “IA forte”, verificou-se que uma apresentação exaustiva é impraticável, devido ao grande número de argumentos e contra-argumentos. Assim, algumas linhas de argumentação foram eleitas para discussão neste trabalho. O critério de seleção procurou balancear importância histórica, coerência e criatividade.

Adotando-se como premissa a visão analítica e racionalista que norteia o desenvolvimento da lógica clássica, sitemas formais e computadores digitais, cabe um questionamento. A partir desta perspectiva, para copiar algo seria preciso primeiramente conhecer este algo. Como portanto copiar completamente a mente humana se, pelo menos no momento, não conhecemos completamente sua abrangência e potencialidades?

# I - Lógica, Paradoxos e o Formalismo

Um tema recorrente na filosofia e na psicologia tem sido a questão da representação do conhecimento e das formas corretas de raciocínio. A inteligência artificial – daqui por diante referida como AI – introduz uma nova meta que consiste em **construir** entidades inteligentes (RUSSELL & NORVIG, 1995 p.3). Esta nova meta requer que o processo de computação seja visto como um algoritmo formal (idem, p. 11), o que torna o estudo da lógica simbólica e dos sistemas formais – e suas limitações – importantes nas discussões expostas neste trabalho. Este capítulo expõe um breve relato do desenvolvimento da lógica até o surgimento do programa formalista na virada do século XIX para o XX.

## 1.1 – Desenvolvimento da antiguidade até o século XIX

Ao tocar no tema "lógica" convém lembrar importância que a revolução racional grega teve para a compreensão do modo de pensar da sociedade atual, especialmente para o pensamento ocidental. Dentro deste tema, o pensamento de Aristóteles (384-322 A.C.) surge, como referência quase que obrigatória. De fato, segundo Goldfarb (ARISTÓTELES, 2001 p 9) o entendimento, mesmo que mínimo, da história da filosofia ou da ciência não pode se furtar ao denso toque do **Corpus aristotelicum.**

Os trabalhos de Aristóteles sobre lógica foram agrupados no "Organon", que quer dizer instrumento, e que compreende as "*Categorias*", "*De interpretatione*", as "*Analíticas*" anterior e posterior e o "*De sophisticis elenchis*". Ressalta-se como importante na lógica aristotélica a questão do ser e do não-ser. Daí se originam os três princípios (ou leis) do pensamento:

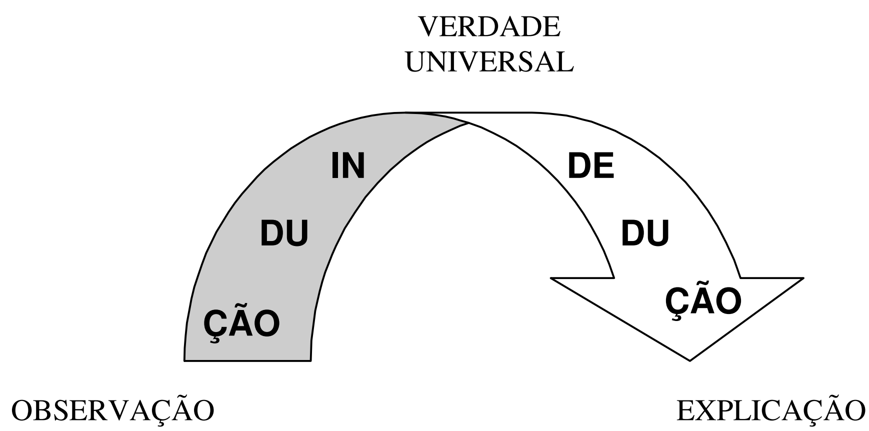
- O princípio da identidade: segundo o qual as características básicas de cada coisa não mudam. **A** é **A** ;
- O princípio da contradição, para o qual uma coisa não pode ao mesmo tempo ter e não ter uma característica. **A** não pode ser **B** e não **B**;

- O princípio do terceiro excluído: não é possível existir um intermediário entre enunciados contraditórios.

  Para um determinado sujeito, é preciso necessariamente, ou afirmar, ou negar um predicado, qualquer que ele seja. **A** é ou **B** ou não **B**.

Para ARISTÓTELES é da observação do mundo natural que podemos chegar à verdade sobre a natureza através da **indução**. Esta teria uma fundamentação lógica de modo a permitir ir do particular observado ao universal verdadeiro, ou do contingente ao necessário, à essência das coisas. Este é o ramo ascendente do **arco do saber**, que veio a ser polêmico ao longo da história da teoria do conhecimento científico, provocando a crítica de HUME (1711-1776) quanto à falta de fundamentação lógica da indução (ROSA, 2002).

De posse de verdades universais obtidas pela indução o arco do saber se completa com o uso da dedução, para a qual ARISTÓTELES desenvolveu a lógica dos silogismos. Assim outras verdades sobre a natureza podem ser deduzidas.




A parte mais frutífera para o estudo da lógica é a “descida” do arco, ou seja, o processo de dedução. O processo de indução de universais a partir de particulares representado neste esquema é, obviamente, o processo de **indução empírica** e não deve ser confundido com a indução matemática. A indução empírica não tem até hoje uma fundamentação lógica rigorosa, tendo sido alvo da crítica de HUME (1978). Esta crítica gerou contribuições importantes, como a resposta de KANT (1724-1804), que dizia justamente que a crítica de Hume o tinha despertado de seu "*sono dogmático*". Para KANT o estágio dogmático é o primeiro passo nas "*coisas da razão*" ou sua infância, e investiga os objetos. O segundo estágio é o cético e testemunha "*a prudência do juízo avisado pela experiência*", e o terceiro seria o da crítica da razão, que submete a exame a própria razão. (KANT, 1997 pp.608-615).

Apesar das investigações que a crítica de HUME suscitou, o problema da fundamentação lógica da indução continua sem solução. CARGILE (1998) comenta várias tentativas de resolução do problema da indução, mostrando que nenhuma transforma resultados indutivos em verdades necessárias. EISENSTADT e SIMON (1997, p. 375) colocam a procura por uma lei válida de indução no mesmo nível de atenção com que os lógicos modernos se dedicam à questão da consistência. Para eles, estas duas empreitadas lógicas têm motivação no desejo de se ter um "mundo" consistente e seguro, imune às incertezas do mundo real que habitamos. No entanto, a indução empírica é usada não só nas ciências naturais mas também em algumas argumentações lógicas, como a feita por REDHEAD no capítulo IV deste trabalho.

Voltando ao **arco do saber**, ARISTÓTELES pretende que a parte **dedutiva** de seu método, possa fornecer certeza e necessidade às conclusões sobre o particular, provenientes de verdades universais. Para tanto, ARISTÓTELES desenvolve todo um estudo de silogismos, ou seja, regras de lógica que permitem que, a partir de premissas, obtenha-se conclusões. Desenvolve também uma série de restrições para definir o que pode ser uma premissa.. Sua teoria de silogismos influenciou todo o pensamento ocidental (ROSA, 2002). Como exemplo de silogismo, vejamos o tipo que Aristóteles considerava o mais importante para demonstrações na ciência (LOSEE, 1977 p. 8):

| | | | |
|---|---|---|---|
| Todos os | **M** | são | **P** |
| Todos os | **S** | são | **M** |
| Então todos os | **S** | são | **P** |

A lógica continuou se desenvolvendo com os filósofos **estóicos**, que, juntamente com ARISTÓTELES, Platão e os atomistas pré-socráticos, constituem o ramo fundamental da filosofia grega sobre o conhecimento (ROSA, 2002). Atribui - se aos estóicos, a introdução dos **silogismos hipotéticos** - em contraste com os silogismos categóricos de Aristóteles.

Entre estes, o *Modus Tollens* está fortemente relacionado à verificação, ou, mais especificamente, à refutação de teorias científicas. Este caracteriza-se por:

$$[\text{Se } P \Rightarrow Q] \text{ e } [\text{Não } Q] \Rightarrow [\text{não } P]$$

Assim, o Modus Tollens fornece a base lógica para o processo de refutação de uma teoria, pois, realizando-se o experimento e não se verificando o que esta predisse, a teoria fica refutada. No entanto, caso **Q** seja verdadeiro, não se pode provar a teoria: **Q** verdadeiro não implica em **P** verdadeiro. Assim vemos que - na concepção moderna de ciência - a teoria nunca é confirmada definitivamente sendo sempre provisória e suscetível de ser refutada.

Para KANT, ARISTÓTELES foi o fundador da lógica, e esta já estaria completa e acabada (KANT, 1997, p. 15). Para ele, o fato de até a sua época a lógica não ter progredido era indicação de que estaria perfeita e acabada (Idem).

De fato, o próximo avanço notável no campo de lógica acontece com LEIBNIZ (1646-1716). Para KNEALE e KNEALE (1971), ele merece ser colocado ao lado dos maiores lógicos, embora seu trabalho em lógica tenha tido pouca influência por quase 200 anos. Apesar de não ser um purista aristotélico, LEIBNIZ declara seu entusiasmo com a doutrina dos silogismos de Aristóteles na sua obra "*Nouveaux Essais*" (KNEALE e KNEALE, 1971 p.322).

LEIBNIZ aplicou o rigor de um raciocínio formal em um esforço para compreender o mundo. Para ele, um esquema lógico suficientemente sofisticado poderia servir como um guia infalível para se entender a estrutura da realidade. Esta era a meta de sua "característica universal" (LEIBNIZ, 1951). Este aspecto da obra de LEIBNIZ vai chamar a atenção de GÖDEL quando este entra na sua "fase filosófica", segundo WANG (1981, p.658). WANG conta que GÖDEL passa a dedicar mais atenção a temas filosóficos após 1950, quando também procura relações entre a filosofia de KANT e a teoria da relatividade.

Para LEIBNIZ, uma vez que fossem estabelecidos números para a maioria dos conceitos, a humanidade possuiria um novo instrumento que ampliaria a capacidade da mente muito mais do que os instrumentos ópticos aumentaram a capacidade dos olhos, e suplantaria o telescópio e o microscópio na mesma proporção que a razão é superior à visão (LEIBNIZ, 1951 p.22-23).

Ao constatarmos que LEIBNIZ idealizava uma linguagem que colocaria todo o processo de raciocínio em um tipo de cálculo, vemos que ele é um precursor da lógica simbólica: LEIBNIZ falava na possibilidade de codificar-se idéias gerais de um modo tão detalhado como se fala de números na aritmética ou de retas na geometria (LEIBNIZ, op. cit. e HERSH,1997 p.124-125). Para CHAITIN (2004), LEIBNIZ tinha as duas idéias básicas do que chamamos hoje de "ciência da informação", pois:

- percebendo o caráter mecânico dos cálculos aritméticos, confeccionou uma das primeiras máquinas de calcular; e
- foi um dos primeiros a valorizar a aritmética binária e o sistema binário de numeração, ou seja, o sistema que representa números através de combinações de apenas dois símbolos: o símbolo "0" e o símbolo "1".

Mas para entendermos a proposta de LEIBNIZ é necessário ligar vários pontos de seu pensamento pois o mesmo procura sintetizar diferentes áreas do conhecimento: teleologia com mecânica, substância com energia, mente e matéria, cálculo infinitesimal com microbiologia, etc. KNEALE e KNEALE (op. cit.) listam o que é minimamente necessário para estudar a lógica de LEIBNIZ:

1) o seu respeito pela lógica tradicional;
2) a sua noção de *ars combinatoria*, ou teoria geral dos arranjos;
3) seu plano para a construção de uma linguagem ideal;
4) seu esquema para a coordenação dos conhecimentos numa enciclopédia;
5) sua esperança em relação a uma ciência geral do método.

Alguns destes interesses eram comuns na época, mas o que chama atenção no trabalho de LEIBNIZ é a sua preocupação em relacioná-los, deixando evidente o seu desejo de síntese. Embora fosse um estudo enriquecedor pela quantidade de tópicos envolvidos, fica fora do escopo deste trabalho um aprofundamento da lógica de LEIBNIZ.

## 1.2 - A renovação da Lógica no século XIX

O século XIX vê surgir o que se chama "lógica matemática", na qual a lógica passa a usar uma linguagem artificial em que a função semântica das palavras e sinais é muito reduzida, chegando mesmo a não haver qualquer função semântica no extremo do formalismo de Hilbert.

No início do século XIX GEORGE BOOLE (1815-1864) fundou uma unificação da lógica com a álgebra, publicando, em 1847, uma obra curta chamada "*The Mathematical Analysis of Logic*" e, mais tarde, em 1854, "*An Investigation of the Laws of Thought*" (BOOLE, 1958), sua obra mais conhecida, na qual amplia e esclarece as idéias da primeira obra, e inclui aplicações à probabilidade. Nestas obras, BOOLE estabelece ao mesmo tempo a lógica formal e um novo tipo de álgebra, chamada álgebra de BOOLE.

Mas, talvez mais importante que sua "lógica algébrica", é a concepção que BOOLE tem da própria matemática: a nova visão da matemática que ele funda liberta-a do domínio restrito dos números e das medidas. A matemática deixa de estar limitada a questões de número e grandeza. Aqui, pela primeira vez, mostra-se que a característica essencial da matemática não é seu **conteúdo**, mas sim sua **forma** (BOYER, 1974 p.428). RUSSELL e NORVIG (1995, p.11) destacam que a lógica tinha o aspecto filosófico mais preponderante do que o aspecto matemático até a introdução da linguagem formal de BOOLE, quando a situação se inverte.

Assim temos a concepção atual de matemática: a área do conhecimento que estuda as relações e propriedades de tópicos internamente consistentes, apresentados com um alfabeto de símbolos e com um conjunto de regras precisas de operação sobre estes símbolos.

Interessante notar que na álgebra de BOOLE está incluído o raciocínio silogístico, como o próprio BOOLE coloca. Além disso, ele avalia seu método como mais simples e, ao mesmo tempo, mais abrangente (BOOLE, 1958 pp. 10-11). Ele mesmo exemplifica (op. cit. pp. 230-231) como operações de sua lógica chegam às mesmas conclusões de um silogismo, dadas as mesmas premissas.

Dando continuação ao trabalho de BOOLE podemos citar J. VENN (KNEALE e KNEALE, pp.420-423) De MORGAN (op. cit. p.427-434), C. S. PEIRCE (idem). KNEALE e KNEALE (op. cit. p.421) relatam também que JEVONS foi o primeiro a perceber que os métodos de BOOLE podem ser reduzidos a regras que se aplicam mecanicamente. Com este conhecimento, JEVONS construiu uma máquina lógica em 1869 que foi apresentada à Royal Society no ano seguinte.

## 1.3 - O *Grundlagen* de FREGE

O próximo passo no desenvolvimento da lógica é dado por GOTTLOB FREGE (1848-1925), que avançou muito em relação à questão do rigor formal. Após o trabalho de FREGE a linha de demarcação entre matemática e lógica se torna muito menos nítida. Outro ponto no estudo de FREGE é a tentativa de mostrar que é possível definir **número** sem qualquer referência à noções externas ao seu cálculo lógico.

A meta de FREGE é diferente da meta de BOOLE, que era a de colocar a lógica como parte da matemática. FREGE empenhou-se em mostrar que a aritmética e a lógica são idênticas (KNEALE e KNEALE, 1971 p.435), justamente porque a noção de número poderia ser obtida sem qualquer noção externa à sua lógica. Mas FREGE percebeu que para realizar a tarefa de demonstrar a identidade entre aritmética e lógica seriam necessários aperfeiçoamentos. KNEALE e KNEALE (op. cit. pp. 435-436) citam os dois requisitos colocados por FREGE neste sentido:

- Organizar o conteúdo tradicional e as contribuições de LEIBNIZ e de BOOLE de forma a tornar clara a estrutura da ciência e a grande variedade de formas proposicionais;
- Tudo que for requerido para as provas dos teoremas deve ser listado inicialmente e explicitamente e as regras de dedução devem ser reduzidas a um

pequeno número de procedimentos-padrão para não haver risco de erros nas demonstrações.

Além disso, para FREGE o uso da linguagem natural era um obstáculo ao pensamento lógico (FREGE, 1970 pp.5-6), tanto que ele considerava uma das tarefas de filosofia "*acabar com o domínio da palavra sobre a mente humana*" (idem, p.7). Com estes objetivos em mente ele escreve, em 1879, sua "*Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*". O termo *Begriffsschrift* é interpretado por HEIJENOORT como uma linguagem voltada ao conteúdo conceitual, escrita com símbolos especiais para o "pensamento puro" (op. cit. p. 1 – introdução). Para FREGE, sua ideografia poderia tornar-se uma ferramenta útil para os filósofos, pois apontaria claramente os enganos ligados aos conceitos e às relações entre os mesmos (op. cit., p.7).

Mas os filósofos não viram sua obra com tanto interesse. Muitos a consideraram incompreensível por causa do seu conteúdo altamente simbólico, conforme conta VILKKO (1998, p.413). Este conta que FREGE esperava mais reconhecimento do que recebeu na época. Conta também que os matemáticos relutaram em avaliar a obra por esta lidar com considerações filosóficas como "conceito", "relação" ou "juízo". Dando continuidade ao seu projeto, FREGE escreve, em 1884, "*Die Grundlagen der Arithmetik*" (os fundamentos da aritmética). Nesta obra, expõe mais informalmente suas idéias, critica as idéias da época sobre a aritmética e apresenta sua definição de **número cardinal** - que será mais tarde reformulada, devido aos paradoxos. Esta definição de número viria, para FREGE, introduzir precisão e, ao mesmo tempo, retirar qualquer subjetivismo da idéia de número. HERSH (1997, p. 142) mostra que para eliminar este subjetivismo FREGE se preocupa, antes de criar sua definição de número na *Grundlagen*, em atacar definições anteriores. Estas são: a de que números são idéias nas mentes das pessoas (psicologismo); o historicismo, segundo o qual os conjuntos numéricos evoluem paralelamente à história humana; ou o empiricismo, defendido por Mill, que postula que os números são apreendidos do mundo físico.

A definição de número cardinal de FREGE baseia-se na teoria dos conjuntos de BOOLE e nas idéias de Cantor sobre conjuntos infinitos. Cantor mostra que dois conjuntos infinitos têm a mesma potência, ou "grandeza", se seus elementos podem ser colocados em correspondência biunívoca. FREGE usa a idéia desta correspondência para a igualdade de inteiros (de conjuntos finitos). BOYER explica que dois conjuntos finitos têm o mesmo número cardinal se os elementos de cada um podem ser postos em correspondência biunívoca com os do outro. Se, então, começarmos com um conjunto inicial, como o conjunto de dedos de uma mão humana normal, e formarmos o conjunto muito maior de todos os conjuntos cujos elementos podem ser postos em correspondência biunívoca com os elementos do conjunto original, então esse conjunto de todos tais conjuntos constituiria um número cardinal, nesse caso o número cinco (BOYER, 1974 p.436). Assim temos um exemplo de que, no trabalho de FREGE pode-se formar conjuntos de conjuntos. E é esta possibilidade que vai ser a causa do paradoxo de RUSSELL, conforme veremos no item seguinte deste trabalho.

De uma forma geral, o número cardinal de uma classe, finita ou infinita, é a classe de todas as classes cujos elementos podem ser colocados em correspondência biunívoca com os elementos da classe dada.

Mais tarde, FREGE amplia sua idéias no "*Grundgesetze der Arithmetik*" (leis básicas da aritmética) em dois volumes (1893 e 1903), mas, ainda no "*Begriffsschrift*" ele introduz a teoria da quantificação (FREGE, 1970, pp.24-27). Na concepção e notação atuais, esta teoria inclui os quantificadores lógicos "para todo" - hoje escrito como $\forall$ e "existe pelo menos um" - hoje escrito como $\exists$[1]. Os quantificadores são generalizações dos operadores de conjunção (hoje escrito como "$\wedge$" ou "&") e de disjunção (hoje escrito como "$\vee$"). Uma proposição $\forall x\ \Phi(x)$ pode ser interpretada como a conjunção de todas as possíveis proposições da forma $\Phi(x)$, ou seja, $\Phi(x_1) \wedge \Phi(x_2) \wedge \Phi(x_3) \wedge \ldots$. Uma proposição $\exists x\ \Phi(x)$ pode ser interpretada como a disjunção de todas as possíveis proposições da forma $\Phi(x)$, ou seja, $\Phi(x_1) \vee \Phi(x_2) \vee \Phi(x_3) \vee \ldots$. (KNEEBONE, 2001 p.63).

Os quantificadores atuam sobre sujeitos, para indicar que um determinado predicado, acima denotado por "$\Phi$", se aplica a todos os sujeitos – no caso do $\forall$ – ou para indicar que existe pelo menos um sujeito ao qual determinado predicado se aplica –

[1] Para HERSH (op. cit.), os quantificadores foram inventados independentemente por O. H. Mitchell, um aluno de C. S. Peirce

no caso do ∃. HILBERT e ACKERMANN mostram que o significado do termo **predicado** é o mesmo do significado usual em filosofia, ou seja, o que permite que um sujeito possa ser caracterizado de forma particular (HILBERT e ACKERMANN,1950 pp.44-45).

Para alguns autores, a introdução dos quantificadores é considerada o nascimento da lógica moderna. Acredita-se que a lógica com quantificadores, também chamada de “cálculo de predicados”, possa exprimir qualquer desenvolvimento matemático numa forma estrita e formal.

## 1.4 - RUSSELL, Seu Paradoxo e os *Principia Mathematica*

Bertrand RUSSELL (1872-1970) procurava acima de tudo substituir a fé religiosa pela certeza científica e, para ele, esta certeza se encontraria mais facilmente na matemática do que em qualquer outra ciência.

FREGE tinha transformado numa regra geral sua quinta lei básica. Esta lei corresponde ao axioma da compreensão na teoria dos conjuntos, e diz que uma condição particular determina um conjunto ou classe. Esta regra geral de FREGE diz que para toda condição exprimível no sistema lógico existe uma classe correspondente. Por exemplo, a condição “ser vermelho” determina o conjunto de objetos vermelhos. A princípio, esta lei não provoca questionamentos, pois como poderia haver um predicado sem que exista a classe de coisas que o satisfaçam?. HERSH (1997, p.148) resume que FREGE considerava “conjunto” ou “classe” como equivalente à “propriedade”. A cada propriedade corresponde o conjunto de coisas tendo aquela propriedade. Por exemplo, à propriedade “ser número inteiro e positivo” corresponde o conjunto dos números naturais. A qualquer conjunto corresponde a propriedade de ser membro do mesmo.[2]

Assim FREGE seguiu seu trabalho, no segundo volume do seu *Grundgesetze* até que RUSSELL, em 1901, acha uma classe (a classe que contém todas as classes que não

[2] FREGE regarded “set” or “class” as equivalent to “property”. To any property corresponds the set of things having that property. To any set corresponds the property of membership in it.

são elementos de si próprias) que mostra que a afirmação "toda condição exprimível determina uma classe" leva a uma contradição. Embora RUSSELL tenha descoberto o paradoxo em junho de 1901, ele só o comunica a FREGE um ano após (HEIJENOORT, 1977 p.124), e o publica nos *Principles of mathematics*, em 1903.

A maioria das classes que podemos conceber é do tipo das que não contém a si próprias, como, por exemplo, a classe dos automóveis (que certamente não é um automóvel), Mas alguma classes são também elementos de si mesmas, por exemplo, a classe de tudo que está impresso nesta página também está impresso nesta página, e a classe de todas as idéias é obviamente uma idéia.

RUSSELL vai se concentrar na classe que contém todas as classes que não são elementos de si próprias, que chamaremos de Z. E a pergunta que RUSSELL coloca é: Z é ou não elemento de si própria? Se dissermos que Z é elemento de Z, então, pela sua própria definição, ela não o é. Por outro lado, se dissermos que Z não é elemento de Z, a sua própria definição decreta que ela pertence a si própria. Assim, RUSSELL verifica que cada possibilidade leva ao seu oposto lógico.

RUSSELL expõe este raciocínio a FREGE numa carta escrita em 19 de junho de 1902. Nesta, ele expressa a idéia do paradoxo não em termos de classes, como exposto acima, mas em termos de um predicado: seja *w* o predicado que é o de não ser predicado de si próprio. Pode *w* ser predicado de si próprio? de cada possível resposta segue-se seu oposto[3] (RUSSELL, 1977 p.125). FREGE imediatamente responde a carta de RUSSELL e mostra o embaraço que sentiu: "*sua descoberta da contradição causou-me a maior surpresa e, quase diria, consternação, pois ela abalou as bases sobre as quais eu tinha a intenção de construir a aritmética*"[4] (FREGE, 1977 p.127). Menciona também sua quinta lei – a que considera "conjunto" equivalente à "propriedade" – que é justamente a que permite o paradoxo: "*com a perda de minha Regra V, não apenas a fundamentação da minha aritmética, mas também a única possível fundamentação da*

[3] Let *w* be the predicate: to be the predicate that cannot be predicated of itself. Can *w* be predicated of itself? From each answer its opposite follows.

[4] Your discovery of the contradiction caused me the greatest surprise and, I would almost say, consternation, since it has shaken the basis on which I intended to build arithmetic

*aritmética, parece desaparecer*"[5] (idem). Ele ainda comenta que, apesar de ser constrangedora à primeira vista, a descoberta de RUSSELL poderá acabar resultando em um grande aperfeiçoamento da lógica.

Assim, como FREGE reconhece, o paradoxo revela mais do que um descuido. Revela uma falha na lógica, um golpe no sistema. Se o axioma da compreensão é contraditório, o que poderá substituí-lo? Como definir, agora, o que determina um conjunto? GUILLEN comenta que "*talvez ninguém pudesse então dizer qual a envergadura da falha ou o que seria necessário para a eliminar, mas foi o bastante para que FREGE reconhecesse que ela viciava todo o seu esforço de dez anos*" (GUILLEN, 1987 p. 23). FREGE chega a inserir um *post scriptum* no segundo volume do seu *Grundgesetze*, falando sobre a frustração de descobrir tal falha no momento em que julgava o trabalho já concluído (FREGE, 1977 p.127 e GUILLEN, 1987 p.23).

Além do devastador paradoxo de RUSSELL, que desestrutura toda a construção de FREGE, tinham surgido também outros paradoxos na teoria dos conjuntos, como o paradoxo de Burali-Forti e o de Berry. Para HERSH (1997, p.148), o paradoxo de RUSSELL e outras "antinomias" mostraram que a lógica intuitiva apresenta mais riscos do que a matemática clássica, pois levava a contradições de uma forma que nunca acontecia na aritmética ou na geometria. Esta foi a "crise nos fundamentos", a questão central nas famosas controvérsias do primeiro quarto do século XX.

Esta situação ameaçava os objetivos do"logicismo", escola da qual pertenciam, além de FREGE e RUSSELL, A. N. WHITEHEAD (1861-1947), e cuja intenção era provar que a matemática e a lógica são similares. Para contornar o problema, RUSSELL tenta reformular a teoria dos conjuntos criando sua **Teoria dos Tipos**, que não levanta os paradoxos, mas que traz outros problemas. Assim, RUSSELL e WHITEHEAD escreveram o *Principia Mathematica*, que tenta reestruturar o programa logicista com a teoria dos tipos. Para este trabalho, que acaba tendo quase 2.000 páginas distribuídas em três volumes publicados entre 1910 e 1913, RUSSELL e WHITEHEAD optaram pela notação de PEANO (que é elogiada por eles no prefácio), mais moderna que a de FREGE. Os paradoxos são evitados ao se proibir que qualquer item possa ser definido

[5] with the loss of my Rule V, not only the foundations of my arithmetic, but also the sole possible foundations of arithmetic, seem to vanish

por referência a uma totalidade que contenha o item que se quer definir, ou seja, criando, por decreto, uma estratificação que proíbe fazer declarações que envolvam ao mesmo tempo elementos de uma classe e a classe em si.

Mas, para expulsar os paradoxos, a teoria dos conjuntos teve de ser remendada de modo demasiado complicado, passando a incluir novos axiomas como o axioma da reducibilidade e o axioma do infinito da teoria dos tipos, que postula que existe um conjunto infinito. No entanto é difícil sustentar que este axioma consiste em um princípio de lógica pura. Para KNEALE e KNEALE (1971, p.669) há algo profundamente insatisfatório sobre este axioma. Sua introdução é tão estranha que, para eles, equivale a abandonar o projeto de FREGE de colocar a aritmética como um desenvolvimento da lógica. HIBERT (1977, p.473) diz que a validade dos axiomas do infinito e da reducibilidade é duvidosa. E GÖDEL acha a solução sugerida por RUSSELL e WHITEHEAD, de proibir que uma proposição não possa dizer nada sobre si mesma, é "*demasiado drástica*", pois pode-se construir proposições que falam a respeito de si próprias e que fazem sentido (GÖDEL, 1979 p.339).

KNEEBONE (2001, p.191) concorda com os autores citados no parágrafo anterior. Para ele, a natureza controversa da teoria dos tipos e a inclusão arbitraria do axioma da reducibilidade constituem uma imperfeição óbvia na proposta de RUSSELL e WHITEHEAD. No entanto, para ele a questão que teria causado o desapontamento maior com o programa logicista foi de outra ordem: os paradoxos conhecidos poderiam realmente ser eliminados com a teoria dos tipos, mas nada garantia que novos paradoxos pudessem surgir no futuro. De fato, a eliminação de defeitos conhecidos não impede que ocorram outros.

Assim vemos que a proposta logicista chega, de uma forma ou de outra, a um impasse, não conseguindo nem construir a fundamentação da matemática pela lógica, nem abolir os paradoxos de forma permanente. Melhor do que eliminar paradoxos seria ter uma garantia de que o sistema é consistente. Esta seria a fundamentação lógica definitiva da matemática (KNEEBONE, 2001, p.192). Assim, surge outra tentativa, tão ambiciosa quanto o logicismo e que vai perseguir a meta de provar a consistência de um sistema que represente a matemática. Este é o tema do próximo item.

## 1.5 - Hilbert e o Formalismo

Enquanto RUSSELL defendia que a matemática era redutível à lógica, HILBERT acreditava que a matemática era uma atividade autônoma. Dessa forma, o formalismo tem uma tendência mais matemática que filosófica e reflete a tendência para a abstração total que se tornava dominante no início do século XX (KNEEBONE, 2001 p.201). A tendência geral em matemática no final do século XIX era na direção de se usar cada vez mais o método axiomático. Conseqüentemente, HILBERT pensava a matemática nestes termos, o que fez com que o programa formalista fosse baseado no desenvolvimento do princípio axiomático (KNEEBONE, 2001, p.201).

Pode-se dizer que David Hilbert (1862-1943) funda o programa formalista no verão de 1900. No *Segundo Congresso Internacional de Matemáticos*, em Paris, ele lista os 23 problemas que, para ele eram os mais instigantes para o futuro da matemática. Para ele, a resolução destes problemas serviriam para "*levantar o véu sob o qual se esconde o futuro, para lançar um olhar nos avanços que aguardam nossa disciplina e nos segredos do seu desenvolvimento nos séculos vindouros*"[6] (KNEEBONE, 2001 p.322). Alguns destes problemas são:

- a hipótese do contínuo;
- a consistência da aritmética (que, conforme mostrará o teorema de GÖDEL, não é possível de prova dentro dos preceitos formalistas);
- problemas a respeito de números primos;
- um procedimento de decisão para a resolução de equações diofantinas.

O último item da lista acima é o décimo problema proposto por HILBERT e requer, para sua solução, que exista um procedimento que nos permita dizer se uma determinada expressão é ou não uma fórmula-bem-formada – uma fórmula que ou é axioma ou é demonstrável pelos axiomas e regras do sistema (HATCHER, 1968 p.12). A este procedimento deu-se o nome de procedimento de decisão. HILBERT impôs que este deveria ser um "procedimento mecânico" e seria usado para decidir se uma prova está correta. No entanto, ele nunca deixou claro o que deveria ser este "procedimento

[6] lift the veil under which the future lies concealed, in order to cast a glance at the advances that await our discipline and into the secrets of its development in the centuries that lie ahead.

mecânico" (CHAITIN, 2002 p.167). Quem vai definir este conceito é TURING (1912-1954), em 1936.

A lista com os 23 problemas de HILBERT ficou famosa, mas THIELE e WOS (2002) contam que há um vigésimo quarto problema, descoberto na década passada, ao serem revisados os arquivos de Hilbert. O problema refere-se a critérios de simplicidade de demonstrações e à forma de se reconhecer se uma dada demonstrações é a mais simples.

Embora, segundo BOYER (1974), cerca da metade dos problemas propostos por Hilbert ainda não tenham sido resolvidos, a sua iniciativa gerou avanços na teoria da prova e inaugurou um novo conceito, que passou a chamar-se "metamatemática". O ponto de vista metamatemático distingue teoria e metateoria e será usado na prova de GÖDEL. Neste caso, a teoria é o sistema usado por RUSSELL e WHITEHEAD no *Principia Mathematica* com os axiomas da aritmética de PEANO e a metateoria é a linguagem na qual GÖDEL discute seu argumento, dando nomes às fórmulas e analisando suas relações (ver cap. seguinte).

As preocupações de Hilbert envolviam a questão das geometrias não-euclideanas e os paradoxos da teoria dos conjuntos, que surgiam das operações com conjuntos infinitos. Segundo PECKAUS (2002 p. 159-162) estas operações eram consideradas livres de contradições até o final do século XIX. Mas logo a teoria dos conjuntos começou a mostrar contradições.

No seu discurso "*On the infinite*", proferido em Münster em 4 de junho de 1925, HILBERT (1977 p.375) compara o que acontece na teoria dos conjuntos com o que tinha ocorrido no início do desenvolvimento do cálculo infinitesimal. No início, a teoria do cálculo ainda não tinha a fundamentação rigorosa atual e as contradições pipocavam, gerando ataques à nova teoria. Justamente por isso, HILBERT presta uma homenagem ao trabalho de WEIERSTRASS na eliminação das controvérsias sobre o cálculo infinitesimal. Ainda no mesmo discurso, HILBERT continua, dizendo que a reação aos paradoxos da teoria dos conjuntos foi "*tão violenta que as noções mais comuns e os modos de inferência mais simples e importantes da matemática estavam ameaçados e* *seu uso estava para ser proibido*" (HILBERT, 1977 p.375). Onde mais encontraríamos confiabilidade e verdade se até o pensamento matemático falha? (idem).

Tendo em vista a preocupação exposta acima, podemos formular a idéia básica do formalismo da seguinte maneira: se temos problemas com o raciocínio lógico ou matemático, que parece estar certo, mas que no fundo apresenta paradoxos, então vamos criar uma linguagem artificial e um conjunto de regras, desta vez excluindo a possibilidade de interpretações dúbias nas sentenças. O formalismo deixa de lado, assim, a **semântica** e vai se preocupar apenas com a **sintaxe**. GUILLEN (1987 p.26) conta que os formalistas atribuíam as deficiências que os paradoxos expunham não à lógica em si, mas ao conteúdo semântico que expressava esta lógica. Em particular, a palavra "todo" originaria muitos deles, pois afirmações como "todas as regras têm exceção" são inócuas ou paradoxais, dependendo de se interpretamos a palavra "todas" como incluindo ou excluindo a afirmação da qual faz parte. Para os formalistas, incertezas deste tipo são mais semânticas do que lógicas, e poderiam ser expurgadas simplesmente limpando a lógica da sua coloração semântica.

STRATHERN (1997 p.23) conta que WITTGENSTEIN (1889-1951) havia identificado filosofia e lógica e que, para ele, "*desconfiar da gramática é o primeiro requisito para filosofar*". De fato, WITTGENSTEIN percebe as dificuldades provenientes da semântica quando, na sua primeira fase, discute com Bertrand RUSSELL (1872-1970) os problemas da lógica e como livrá-la dos paradoxos. STRATHERN (1997 p.21) relata o relacionamento de WITTGENSTEIN com RUSSELL quando o primeiro faz observações sobre o simbolismo, como por exemplo que "A" quer dizer o mesmo que a letra "A" (WITTGENSTEIN, 1994 p.151). Ou seja, numa proposição lógica, a letra "A" não deve ser considerada como nada mais que um símbolo. Esta assertiva é o aforismo 3.203 do seu "*Tractatus Logico-Philosophicus*". RUSSELL entende perfeitamente o que WITTGENSTEIN diz: a fim de evitar-se os paradoxos, as proposições devem ser mostradas de forma simbólica.

O propósito do programa formalista foi resumido pelo próprio HILBERT, em duas ocasiões. No discurso citado anteriormente (1925), ele coloca que a meta de sua teoria "*é fornecer ao método matemático a confiabilidade definitiva*" (HILBERT, 1977 p.370). E em outro discurso – *The foundations of Mathematics* – proferido em 1927 para o Seminário Matemático de Hamburgo, HILBERT diz:

> eu gostaria de eliminar de uma vez por todas as questões relativas à fundamentação da matemática transformando toda proposição matemática numa fórmula que pode ser concretamente exibida e estritamente derivada, colocando, assim, as definições e inferências matemáticas de maneira que sejam inatacáveis e ainda forneçam uma descrição adequada da ciência. (HILBERT, 1977 p.464)

Além do programa de HILBERT, voltado exclusivamente para o problema da fundamentação de matemática, a abordagem formalista oferece vantagens de um modo mais geral. HEYLIGHEN (1999) expõe as vantagens da formalização em contraposição à influência que o contexto exerce sobre as expressões não formais. Quando mais formal uma expressão, mais ela é independente do contexto. Para ele, as vantagens do formalismo são (HEYLIGHEN, 1999 p.32-33):

- armazenamento por um longo período de tempo. Sendo o tempo um componente do contexto, quanto maior o período de tempo que quisermos que o conhecimento seja válido, mais formal terá de ser a forma de armazená-lo. Esta seria uma das razões pelas quais a linguagem escrita é mais formal que a linguagem falada;
- a capacidade de comunicação universal. Quando maior a variedade de pessoas com as quais queremos nos comunicar, maior a variedade de contextos. Como a expressão formal é independente do contexto, ela é compreensível por todos;
- a facilidade em testar o conhecimento. Como o contexto onde será feito o teste nunca será exatamente igual ao contexto em que a proposição foi feita, quanto mais formal a expressão mais apurado será o teste do conhecimento

Para HEYLIGHEN, a combinação dos três fatores acima torna mais fácil a acumulação e a melhoria do conhecimento. A testabilidade permite selecionar as melhores descrições; o armazenamento mantém as descrições assim escolhidas; e a universalidade permite que conhecimento descoberto por pessoas diferentes em contextos diferentes possa ser compartilhado. HEYLIGHEN (op. cit. pp. 34-42) aponta também limitações no formalismo, dentre as quais as reveladas pelos resultados de GÖDEL e que são expostas no capítulo seguinte deste trabalho.

O formalismo mantém a importância do método axiomático, ou seja, a partir de alguns axiomas - ou pontos de partida - independentes entre si, construir teoremas a partir da aplicação das regras da lógica sobre os axiomas e os teoremas já provados - do mesmo modo como Euclides. Mas com uma diferença importante: aqui as regras da lógica para derivação dos teoremas são substituídas por **regras de produção**, que são nada mais do que modos permitidos de manipular símbolos, ou melhor, cadeias de símbolos, para gerar outras cadeias de símbolos válidas para o sistema. As cadeias válidas são chamadas **fórmulas bem formadas** (fbf), e mesmo o nome "axioma" deveria ser substituído por **fórmula inicial**, pois os termos "axioma", "regra de inferência", e "teorema" são relativos ao sistema dedutivo que usamos na interpretação do sistema formal.

Desta forma, com regras absolutamente precisas, seria possível obter o tão desejado rigor absoluto, e uma prova formulada num sistema destes, chamado **sistema formal**, seria absolutamente clara e sem falhas.

O programa de Hilbert pode ser resumido em três etapas (HERSH, 1997):

Primeira:

Criar uma linguagem formal e regras formais de maneira que cada prova clássica pudesse ser substituída por uma derivação formal a partir de fórmulas iniciais por intermédio de passos mecânicos de manipulação de símbolos. As fórmulas iniciais representam os axiomas e os passos mecânicos são estipulados pelas regras de transformação de uma fórmula em outra fórmula do sistema formal. Uma derivação formal é uma seqüência finita de aplicações destes passos mecânicos à uma fórmula ou conjunto de fórmulas de forma a obter, na última aplicação da seqüência, o resultado - uma cadeia de caracteres - correspondente a um "teorema" do sistema axiomático.

Feito isto, os axiomas da matemática poderiam ser tratados como cadeias de caracteres sem significado. Os teoremas seriam outras cadeias sem significado. A derivação de axiomas em teoremas (a prova) seria apenas uma seqüência de rearranjos nos símbolos;

Segunda:

Desenvolver uma teoria combinatória sobre estas manipulações que geram "provas". As regras de inferência agora são regras de manipulação de *strings*, ou seja, cadeias de caracteres;

Terceira:

As seqüências possíveis de caracteres são sempre finitas, e podem ser estudadas por análise combinatória, ou seja, pelas possibilidades de combinações e arranjos entre os caracteres permitidas pelas regras de transformação.

HATCHER (1968, p.12) define sistema formal como

- um conjunto A, chamado alfabeto, cujos elementos são chamados símbolos;
- um conjunto M(A), que é o conjunto de todas seqüências finitas dos elementos de A;
- um conjunto S, cujos elementos são as fórmulas-bem-formadas. S é um subconjunto de M(A)
- um conjunto de axiomas, chamado P, que é subconjunto de S
- um conjunto de regras de inferência, chamado R. Estas regras são definidas como relações finitas sobre S. Uma relação é um sub conjunto de $S^n$, sendo que $S^n$ é o conjunto de todas as n-uplas de elementos de S.

A exigência de que as cadeias de caracteres sejam sempre finitas vem da imposição de que as discussões, proposições e definições possam ser inspecionadas concretamente. Esta é postura finitista, imposta por HILBERT Ou, como explica KNEEBONE (2001, p.205), o que é essencial ao raciocínio finitista é que:

- as entidades mencionadas sejam produzidas e não meramente postuladas;
- nenhum modo de definição ou processo de cálculo seja admitido a não ser que possamos garantir que termine em um número finito de passos; e
- um limite superior para este número de passos possa ser estipulado antecipadamente.

A demonstração do teorema da incompleteza de GÖDEL, que é comentada no próximo capítulo, é central nas argumentações discutidas neste trabalho. Esta demonstração mostra a preocupação do mesmo com os requisitos finitistas. Para chegar à definição de “fórmula demonstrável”, GÖDEL cria 45 definições (GÖDEL, 1992 pp.49-55) que seguem exatamente as três exigências listadas acima. A última definição, que é justamente a de “fórmula demonstrável” é a única exceção, fato que GÖDEL (op. cit. p.55) deixa bem claro.

Já HOFSTADTER (1980, pp, p. 229-230) faz uma analogia dos propósitos do programa de Hilbert com o processo usado para ligar dois navios com uma corda espessa através de uma roldana. Primeiramente uma flecha de metal é passada pelo orifício da roldana, puxando uma corda fina. Esta corda fina é ligada, na outra extremidade, a uma corda mais espessa. Quando a conexão já está formada pela primeira corda, a corda espessa – atada à primeira - pode ser puxada pelo orifício e a ligação entre os dois navios é estabelecida por esta última.

HOFSTADTER compara a corda fina a um sistema formal mais simples e a corda espessa ao sistema formal, mais complexo, cuja consistência quer se demonstrar. Para ele, o formalismo tinha a expectativa de provar a consistência do sistema mais complexo através do sistema mais simples. Esta analogia é baseada na suposição de que, para mostrar que é impossível chegar-se a uma contradição, fatos sobre a aritmética não seriam necessários. Ou seja, se todas as propriedades dos números inteiros não fossem necessárias, então poderia se provar a consistência da aritmética através de um sistema mais simples. Para HOFSTADTER o objetivo era provar a consistência da formalização da aritmética através do conjunto restrito de métodos de raciocínio “finitistas” – que, na analogia, correspondem à corda fina. Continuando na sua analogia, HOFSTADTER conta que o que GÖDEL mostrou foi que, no caso da aritmética, não se pude usar uma corda fina para puxar a corda espessa.

Outra questão imposta pelo formalismo é a da interpretação de um sistema formal. Como as sentenças, por si mesmas, não tem significado, não são falsas nem verdadeiras. Mas podemos atribuir valores lógicos (falso, verdadeiro) a estas sentenças através de uma interpretação do sistema formal (RUSSELL e NORVIG, 1995 p.163). Os valores lógicos falso e verdadeiro são atribuições **externas** ao sistema. Uma sentença é verdadeira sob uma interpretação particular se o estado das coisas exteriormente ao sistema formal verifica o que a mesma diz. Esta atribuição de valores é feita pelo que RUSSELL e NORVIG chamam de **semântica**. Assim, podemos destacar dois tipos de classificação de sentenças de um sistema formal:

- uma interna, que classifica as sentenças em fórmulas-bem-formadas ou não. Como subconjuntos da classe de fórmulas-bem-formadas, as fórmulas podem ser classificadas em demonstráveis ou não-demonstráveis. Esta classificação tem relação com os axiomas e regras de inferência do sistema;
- uma externa, que classifica as sentenças em falsas ou verdadeiras, de acordo com uma interpretação do sistema.

Podemos ver que a interpretação de um sistema formal é parte da solução de um problema que GÖDEL (2001c, p. 249) destaca que o formalismo teve de enfrentar. Este é o problema de demonstrar a utilidade das suas manipulações puramente simbólicas. E tais manipulações só se mostram úteis à medida que pode-se mostrar que cada fórmula deduzida pelas regras do sistema representa uma sentença verdadeira da aritmética.

Além disso, supõe-se que a estrutura das proposições do sistema formal sejam de tal forma que as interpretações destas proposições sejam frases declarativas da linguagem natural. Se A é uma proposição do sistema formal e uma certa frase é uma interpretação de A, então diz-se que A é a “expressão no sistema formal” dessa frase ou de qualquer frase equivalente à mesma (ROSSER, 1979 p.388).

# II – Os resultados de GÖDEL e a Máquina de TURING

Os resultados a que GÖDEL (1906-1978) chega em 1931 provam que as metas do formalismo nunca poderão ser atingidas. WANG (1981) conta que GÖDEL começou estudando física em Viena, mas seu interesse na precisão levou-o da física para a matemática e desta para a lógica matemática (WANG, 1981 p.653).

A motivação de GÖDEL para pesquisar as propriedades de sistemas formais surgiu de um questionamento sobre a meta de HILBERT de provar a consistência da análise pelos métodos finitistas (WANG, 1981 p.654). GÖDEL achava esta meta muito ambiciosa. Para ele, seria mais sensato primeiramente provar a consistência da aritmética pelos métodos finitistas e só depois provar a consistência da análise através da aritmética. Mas o plano de GÖDEL não funcionou, pois para provar a consistência da análise ele precisaria usar o conceito de verdade na aritmética. Este conceito seria necessário para verificar certos axiomas da análise. Mas, ao tentar trabalhar com ele, GÖDEL logo deparou-se com paradoxos, percebendo, desta forma, que o conceito de verdade na aritmética não pode ser definido dentro dela mesma (idem).

Para sua demonstração, GÖDEL cria um sistema formal ***P*** dentro do formalismo proposto por HILBERT, ou seja, um conjunto de símbolos, fórmulas iniciais (equivalentes aos axiomas) e de regras segundo as quais cada nova fórmula pode ser obtida de outra pela manipulação mecânica dos símbolos, de acordo com as regras de produção.

GÖDEL exige para este sistema uma condição mais forte que a mera consistência, ou consistência simples. Ele exige que o sistema seja *w*-consistente. A consistência simples impõe que não seja possível demonstrar no sistema as proposições **A** e **~A.** A condição para um sistema ser *w*-consistente é mais rígida: se conseguirmos provar a proposição **~** $\forall$**n (P(n))** – ou seja, "não é caso que para todo **n** o predicado **P** se aplica a **n**", então não é possível demonstrar todas as proposições **P(0), P(1), P(2), P(3)...**

PENROSE (1996 p.91) explica esta condição ao dizer que em um sistema que não é *w*-consistente temos a situação anômala na qual, para alguns predicados P, todas as P(0), P(1), P(2), P(3), ... poderiam ser provadas; e mesmo assim a proposição que diz que não todos entre eles é verdade é demonstrável também! Certamente nenhum sistema formal digno de confiança poderia admitir este tipo de coisas. Se os teoremas demonstráveis em um dado sistema formal correspondem à verdade em alguma interpretação, então este sistema é certamente *w*-consistente.

A possibilidade de um sistema ser consistente e, ao mesmo tempo não ser *w*-consistente parece implausível, mas TARSKI foi o primeiro a dar um exemplo de sistema simplesmente consistente e ao mesmo tempo *w*-inconsistente, no artigo intitulado "*Einige Betrachtungen über die Begriffe der w-Widerspruchsfreiheit und der w-Vollständigkeit*"[1] (Observações sobre os conceitos de *w*-consistência e de *w*-completeza), segundo ROSSER (1979, p. 389).

É claro que se um sistema não é simplesmente consistente, também não é *w*-consistente, pois sendo inconsistente, pode-se provar qualquer proposição, inclusive todas as **P(0), P(1), P(2), P(3), ...** e também a proposição **~ ∀n (P(n))**. Daí, pelo *modus tollens* (ver cap. I deste trabalho), conclui-se que *w*-consistência implica consistência simples.

A *w*-consistência é uma propriedade muito mais restrita que a consistência simples. Mas o lógico americano J. Barkley ROSSER (1936) mostrou que os resultados de GÖDEL se estendem a sistemas formais **apenas** consistentes. Ou seja, aplicam-se a um conjunto muito maior de sistemas formais. Isto porque é óbvio que o conjunto de sistemas *w*-consistentes é um subconjunto próprio do conjunto de sistemas apenas consistentes. ROSSER chega assim a "*uma versão mais forte do teorema VI de Gödel*" (ROSSER, 1936 p.87) cuja demonstração é publicada sob o título "*Extensions of some Theorems of Gödel and Church*" (Extensões dos teoremas de Gödel e Church).

Antes de dar uma definição para **prova**, ou **demonstração** (*Beweisfigur*), GÖDEL lista as regras de inferência que vai utilizar. São somente duas:

---

[1] Monatsheft für Mathematik und Physik, vol. 10 (1933), pp. 97-112.

**i )** se $\boldsymbol{a}$ é um axioma ou uma fórmula que já foi demonstrada, então a generalização da fórmula, ou seja, $\forall$**x (*a*)** também é válida;

**ii )** a segunda regra é conhecida como *modus ponens*.
ela nos diz exatamente que de $\boldsymbol{b} \Rightarrow \boldsymbol{c}$ e de $\boldsymbol{b}$ podemos inferir $\boldsymbol{c}$. Esta expressão latina é também chamada de regra da separação (*law of detachment*) e de *modus ponendus ponens* (SUPPES, 1966, p.32). HILBERT e ACKERMANN (1950, p.70) ainda se referem à mesma por regra da implicação (*rule of implication*).

GÖDEL define demonstração de uma fórmula ***x*** como uma seqüência finita de fórmulas, da qual cada fórmula é um axioma ou uma conseqüência imediata de duas fórmulas anteriores; e ***x*** é a última fórmula da seqüência (GÖDEL, 1992 p.55). Esta definição de GÖDEL para **prova** está de acordo com prescrição de HILBERT (1977, p.381-382) para a formalização de uma prova matemática.

GÖDEL também define o conjunto de conseqüências (*Folgerungsmenge)* de um conjunto de fórmulas ***k***: é o menor conjunto que contém todas as fórmulas de ***k***, todos os axiomas, e também todas as fórmulas obtidas das anteriores por meio da aplicação das regras de inferência (GÖDEL,1992 p.57).

Além do conceito de sistema formal, GÖDEL cria um método inteiramente novo, chamado "aritmetização", que consiste em associar números, buinivocamente, às sentenças dentro do seu sistema ***P***.

Ele associa números:

- Aos símbolos;
- Às seqüências de símbolos (fórmulas);
- Às seqüências de seqüências de símbolos (provas).

Quanto aos símbolos, GÖDEL faz a seguinte correspondência:

| | | | | | |
|---|---|---|---|---|---|
| **0** [zero] | ---- 1 | **V** [ou] | ----- 7 | **)** --------- 13 | |
| **f** [sucessor] | ---- 3 | $\forall$ [para todo] | ---- 9 | $x_n$ -------$17^n$ | n = 1,2,3....... |
| ~ [negação] | ----- 5 | **(** | ------------ 11 | $y_n$ -------$19^n$ | n = 1,2,3........ |

Quanto às seqüências de símbolos (fórmulas) GÖDEL cria uma regra de formação dos números chamados “números de GÖDEL”, que são formados através de um produto, da seguinte forma:

Para cada série de K elementos, cujos elementos têm números de GÖDEL

$n_1, n_2, n_3, \ldots\ldots n_k$ é construído o produto:

$2^{n1}.3^{n2}.5^{n3}\ldots\ldots\ldots\ldots P_k^{nk}$ , um produto cujos fatores primos são os primeiros K números primos, com o primeiro, segundo,........k-ésimo número primo ocorrendo respectivamente $n_1, n_2, n_3, \ldots\ldots n_k$ vezes no produto.

Seja como exemplo, a expressão $\forall x_1[x_2(x_1) \ \vee \ \sim x_2(x_1)]$
que quer dizer:
para todo elemento $x_1$ , tem-se que a propriedade $x_2$ aplica-se a $x_1$ ou que a propriedade $x_2$ não se aplica a $x_1$.
A esta expressão corresponde o produto:

$$2^9.3^{17}.5^{11}.7^{17^2}.11^{11}.13^{17}.17^{13}.19^7.23^5.29^{17^2}.31^{11}.37^{17}.41^{13}.43^{13}$$

Através deste exemplo, podemos ver como uma característica numérica pode corresponder a uma observação metamatemática. No caso, à observação metamatemática “a fórmula tem um quantificador universal” equivale a característica “o número de GÖDEL correspondente é divisível por $2^9$ , mas não é divisível por nenhuma outra potência de 2 maior que esta” (KNEALE e KNEALE,1971 p.715-716).

Para as seqüências de seqüências de símbolos (que são justamente as provas), vale a regra acima, sendo cada $n_i$ o número de GÖDEL de uma linha da seqüência.

O motivo pelo qual GÖDEL escolhe esta regra de formação dos números é para garantir uma correspondência um-para-um entre cada número de GÖDEL e cada símbolo, seqüência, ou seqüência de seqüência. Isto é garantido pelo **teorema**

**fundamental da aritmética**, que diz que qualquer inteiro positivo a partir de 1 pode ser expresso como um produto de números primos. Esta expressão é única exceto pela ordem na qual os primos ocorrem.[2]

BRAITHWAITE, que faz a introdução do artigo em GÖDEL (1992), comenta as relações entre os números de GÖDEL e as fórmulas e demonstrações no sistema ***P***. Ele explica que a regra de aritmetização de GÖDEL assegura que para cada classe de *strings*, ou seja, cadeias de caracteres, corresponde uma única classe de números de GÖDEL e vice-versa. Assegura também que a cada relação ***R*** entre *strings* corresponde uma única relação ***R'*** entre números de GÖDEL, e vice-versa. Por exemplo, a declaração metamatemática de que uma seqüência de fórmulas ***s*** é uma 'prova' da fórmula ***f*** é verdadeira se e somente se uma determinada relação aritmética ocorre entre os números de GÖDEL de ***s*** e de ***f*** . No caso, esta relação corresponde à relação "ser uma *prova* de"[3] (GÖDEL, 1992 p. 9). Este exemplo corresponde à definição 45 (GÖDEL, 1992, p.55), uma dentre as 46 definições que GÖDEL usa para construir sua sentença.

Assim, GÖDEL procede a sua demonstração tendo os axiomas do *Principia Mathematica* de RUSSELL e WHITEHEAD e da aritmética de PEANO como teoria a ser estudada e usa os números de GÖDEL como elementos da metalinguagem com que a argumentação é feita. GRATTAN-GUINESS (1979, p.296) ressalta que GÖDEL enfatiza o caráter metamatemático do seu argumento. De fato, na seção 1 do seu artigo, que apresenta um resumo de sua prova, GÖDEL (1992, pp.37-41) deixa claro o que é o sistema formal por ele utilizado e o que são conceitos e considerações metamatemáticas. Estas são "percebidas do exterior"[4] (GÖDEL, 1992 p. 38).

Para GRATTAN-GUINESS (op. cit.), esta ênfase é feita porque, apesar da proposta que HILBERT tinha feito, desde os anos 20, de separar teoria de metateoria, até a prova de GÖDEL os lógicos ainda tendiam a confundir estes dois níveis. Para ele,

---

[2] Cf. *Concise Dictionary of Mathematics* (1996)
[3] GÖDEL's rule of arithmetization ensures that to every class of strings there corresponds a unique class of GÖDEL numbers, and *vice versa*. And that to any relation *R* between strings there corresponds a unique relation *R'* between GÖDEL numbers, and *vice versa*.(…) For example, the metamathematical statement that the series *s* of formulae is a 'proof' of the formula *f* is true if and only if a certain arithmetical relation holds between the GÖDEL numbers of *s* and of *f* which corresponds to the relation: being a 'proof' of.
[4] looked at from the outside

a prova de GÖDEL tem um mérito adicional: só a partir desta os lógicos passaram a ver o quanto cuidadosos precisam ser em separar linguagem de metalinguagem.

Os itens 2.1 e 2.2 abaixo discorrem sobre os dois resultados a que GÖDEL chega no seu trabalho apresentado na Academia de Ciências de Viena em 1930 e publicado no *Monatshefte für Mathematik und Physik* em 1931 (GÖDEL, 1992 p.1) sob o título de *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme 1* (Sobre proposições formalmente não-decidíveis do *Principia Mathematica* e sistemas similares 1). Ainda neste ano, em 15 de setembro, GÖDEL dá uma palestra sobre seus resultados no encontro de outono da *Deutsche Mathematiker-Vereinigung*, em Bad Elster (GRATTAN-GUINNESS, 1979 p.295).

Ele prova que um sistema formal abrangente o bastante para conter a aritmética com as operações de adição e multiplicação será tal que:

a) sempre terá proposições dentro do próprio sistema que não podem ser provadas, ou seja, um sistema deste tipo será sempre incompleto. Este é o teorema VI, relativo à questão da incompleteza e que é apresentado com algum detalhe no item 2.1;
b) se o sistema for consistente, a sua consistência não poderá nunca ser demonstrada dentro do próprio. Este é o teorema XI, do qual trata o item 2.2.

## 2.1 – O primeiro resultado

O item (a) acima é conhecido como “teorema de GÖDEL” e, mais precisamente, ele diz que o sistema será **ou** incompleto **ou** inconsistente. Se ele gerar apenas verdades, não gerará nunca todas as verdades (incompleteza). Se ele gerar toda a verdade, gerará também inverdades (inconsistência).

Como um sistema inconsistente não tem nenhuma utilidade, temos que optar por sistemas incompletos. A inutilidade de um sistema inconsistente vem do fato de que ele inclui, ao mesmo tempo, uma afirmação e a negação da mesma, ou seja, inclui como teoremas tanto **A** como **não-A** (CHAITIN, 1990, 1993 e 2002) . É um sistema no qual se prova qualquer coisa, onde tudo é verdade. DAGHLIAN (1995, p. 68) apresenta uma demonstração sucinta do fato de que a partir de uma contradição pode-se deduzir qualquer proposição. Para esta demonstração ele usa as seguintes regras de inferência (op. cit. pp. 56-57):

- simplificação: da conjunção **p ∧ q** podemos inferir tanto **p** quanto **q**.
  podemos conferir que esta regra é correta conferindo a última coluna da tabela-verdade de **p ∧ q** e vendo que **p ∧ q** só é verdadeiro quando **p** e **q** também o são:

| p | q | p ∧ q |
|---|---|---|
| F | F | F |
| F | V | F |
| V | F | F |
| **V** | **V** | **V** |

- adição: de uma proposição **p** podemos inferir a disjunção **p ∨ q**.
  podemos conferir que esta regra é correta conferindo na tabela-verdade que, se **p** é verdadeiro, então **p ∨ q** também o é, independente do valor-verdade de **q**:

| p | q | p ∨ q |
|---|---|---|
| F | F | F |
| F | V | V |
| **V** | **F** | **V** |
| **V** | **V** | **V** |

- silogismo disjuntivo: da disjunção **p ∨ q** e da negação de **p**, ou seja, **~p**, podemos inferir **q**.
  podemos conferir que esta regra é correta conferindo na tabela-verdade que, se **p ∨ q** é verdadeiro e se **~p** também é verdadeiro, ou seja, **p** é falso, então **q** é verdadeiro:

| p | q | $p \vee q$ |
|---|---|---|
| F | F | F |
| **F** | **V** | **V** |
| V | F | V |
| V | V | V |

Com as três regras listadas acima podemos demonstrar que de uma contradição podemos chegar a qualquer proposição. A demonstração parte da contradição $\mathbf{p} \wedge \mathbf{\sim p}$ como premissa para provar uma proposição $\alpha$ qualquer:

1. $\mathbf{p} \wedge \mathbf{\sim p}$ premissa
2. **p** simplificação de (1)
3. **~p** simplificação de (1)
4. $\mathbf{p} \vee \alpha$ adição de (2)
5. $\alpha$ silogismo disjuntivo de (3) e (4)

Lembrando que, sendo $\alpha$ uma proposição qualquer, então, a demonstração acima mostra que de $\mathbf{p} \wedge \mathbf{\sim p}$ prova-se qualquer proposição, ou seja, tudo. E, se tudo é verdade, então nada é verdade, e não pode-se provar coisa alguma, pois não há contraponto.

GÖDEL prova, na proposição VI, que há proposições aritméticas que são indecidíveis, isto é, não se pode provar sua veracidade e nem sua falsidade. Esta proposição é formalizada através da construção, no sistema formal ***P***, de uma fórmula que expressa sua própria impossibilidade em ser demonstrada. Mais especificamente, nem esta fórmula e nem a sua negação são demonstráveis (GÖDEL, 1992 p.57-59, Proposition VI).

GÖDEL segue considerando uma hipótese do tipo : "Dentro deste sistema, não é possível provar que esta hipótese é verdadeira", o que conduz a uma contradição:

- Se provarmos que ela é verdadeira, então provamos que "não é possível provar que esta hipótese é verdadeira", ou seja, que ela é falsa;
- Se provarmos que ela é falsa, então provamos que é possível provar que ela é verdadeira, ou seja, que é verdadeira.

Mas esta hipótese precisa ser formalizada para que se torne uma sentença do sistema ***P*** , o que exige que GÖDEL crie predicados e contorne problemas da linguagem natural. DELONG (1971 pp.160-164) sugere os seguintes passos para a obtenção da sentença de GÖDEL através da formalização da hipótese acima:

**1º.** os conceitos de verdadeiro/falso são semânticos, e portanto a expressão "Esta hipótese não é verdadeira" deve ser convertida para "Esta sentença não é demonstrável";

**2º.** a característica de demonstrabilidade não é absoluta. Uma sentença que não é demonstrável em um sistema pode ser demonstrável em outro. Por isto, deve-se especificar que se trata do sistema ***P***. A sentença deve, então, ser da forma " Esta sentença não é demonstrável em ***P***;

**3º.** restrições quanto à semântica já foram comentados neste trabalho (ver capítulo anterior). TARSKI (1969, p.68) também alerta para o fato de que pronomes demonstrativos e advérbios podem causar interpretações dúbias de sentenças e, portanto, não devem ser permitidos em uma linguagem formal.

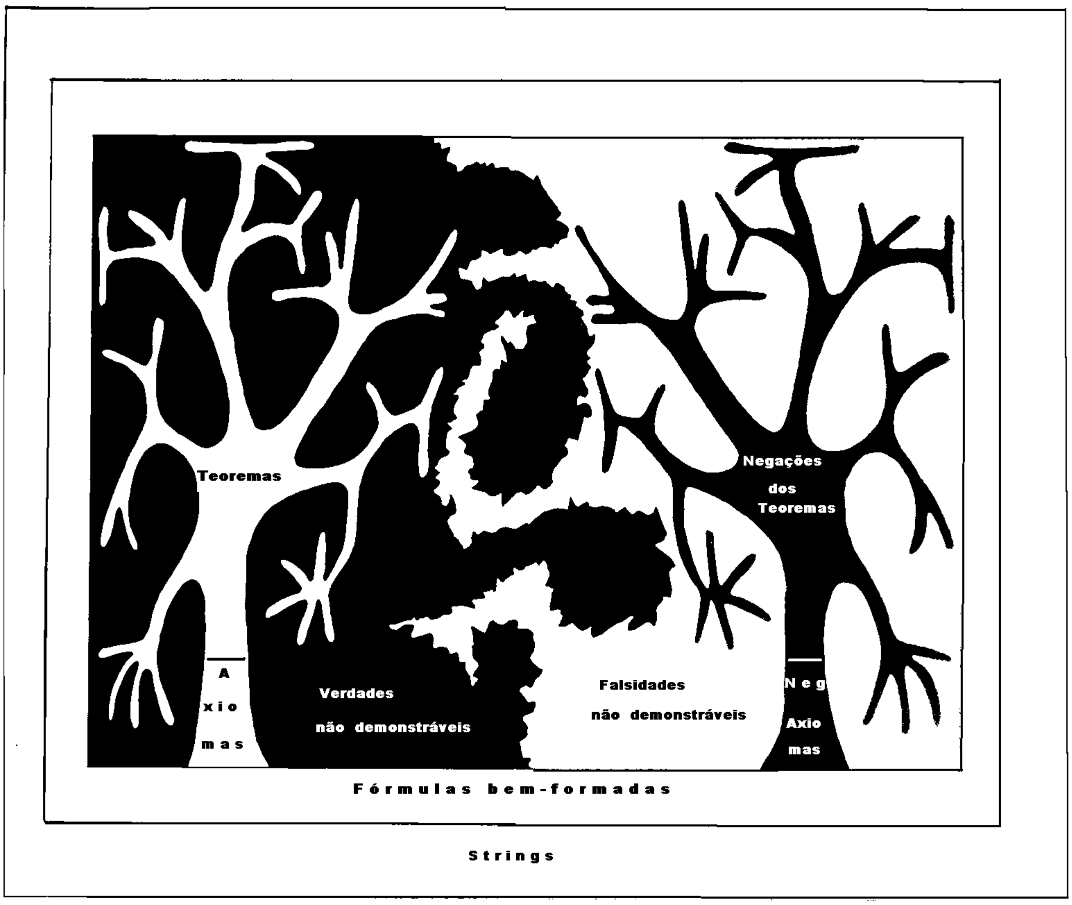
Logo, para eliminar a palavra "esta" foi necessário criar uma descrição unicamente verificada pela sentença em particular que se quer formalizar. Isto vai exigir que GÖDEL defina mais dois predicados, que fazem parte das 46 definições dadas por ele antes da demonstração propriamente dita.

**4º.** GÖDEL ainda precisa criar definições para o conceito de substituição de variáveis, para o conceito de generalização (o operador $\forall$), para o conceito de negação, entre outras.

O que normalmente é chamado de "teorema de GÖDEL" é a sua proposição VI, que prova que há proposições aritméticas que são indecidíveis, isto é, não se pode provar sua veracidade e nem sua falsidade. GÖDEL chega à formalização deste fato ao confeccionar uma sentença que chamaremos de **G**, em concordância com a nomenclatura usada pela maioria dos autores da área. Esta expressão expressa sua própria impossibilidade em ser demonstrada. Mais especificamente, nem **G** e nem a negação de **G** são demonstráveis (Op. cit. p.57-59, Proposition VI ).

GÖDEL mostra que a proposição, que é verdadeira de um modo ou de outro, não é demonstrável dentro do sistema, se este sistema for consistente - mais precisamente, *w*-consistente, como vimos acima. Ou seja, todo sistema formal como este, ou mais complexo, sendo consistente, será sempre incompleto: haverá verdades não demonstráveis pelo sistema.

A interpretação metamatemática deste resultado é que, mesmo dentro da aritmética, teremos sempre proposições que não podem ser provadas (e nem negadas, o que seria o equivalente a provar sua negação). Estas proposições são atualmente chamadas de indecidíveis. HOFSTADTER, no seu livro "*Gödel, Escher and Bach*" criou uma ilustração para esta interpretação (HOFSTADTER (1979) p. 71):




A moldura mais externa representa o conjunto de *strings*, ou seja, de seqüências de caracteres do sistema. Um subconjunto deste é o formado pelas fórmulas-bem-formadas (*well-formed formulas*) que são os axiomas e as fórmulas cujas seqüências de símbolos satisfazem o formato imposto para fórmulas do sistema.

A moldura mais interna representa a situação explicitada pelo primeiro resultado de GÖDEL. Como os sistemas inconsistentes não têm nenhuma utilidade, os sistemas formais úteis revelam-se sempre incompletos, o que vai gerar os chamados "indecidíveis", representados na ilustração de HOFSTADTER como as áreas intermediárias que circundam as duas "árvores" representantes das certezas. A árvore branca representa um tipo de certeza: os teoremas, tendo como base os axiomas do sistema. A árvore preta representa outro tipo de certeza: a negação dos teoremas e dos axiomas, e é exatamente simétrica à árvore branca.

À medida que se prova um novo teorema, forma-se um novo galho na árvore branca e, conseqüentemente, e simetricamente, na preta. Como exemplo, temos o famoso "último teorema de FERMAT", que, por três séculos ficou na condição de candidato a indecidível, até que, em 1994, foi provado por Andrew WILES (SINGH, 1998). Consiste na seguinte proposição:

**$x^n + y^n = z^n$ não tem solução inteira para $n > 2$, com x,y,z e n inteiros**

Com a prova deste teorema, a proposição que o representa passou a ser um novo ramo na árvore branca, gerando também um novo ramo – simétrico – na árvore preta. No entanto, a área intermediária sempre existirá. Outro exemplo de candidato a indecidível é a proposição de GOLDBACH:

**Todo e qualquer número par é composto pela soma de dois números primos**

GOLDBACH afirmou isto em 1742, e consultou EULER sobre o assunto, mas nem este nem ninguém mais conseguiu provar a proposição.

## 2.2 – O segundo resultado

O próprio GÖDEL pode prefaciar este item com seu comentário sobre o primeiro resultado: "*a análise precisa desta situação curiosa conduz a resultados surpreendentes sobre as provas de consistência para sistemas formais*" (GÖDEL,1970 p.90).

Este é o segundo resultado do mesmo artigo de GÖDEL, e que consiste na prova da proposição XI: a fórmula que estabelece a consistência do sistema ***P*** não é demonstrável em ***P*** (GÖDEL, 1992 p.70-72-Proposition XI e KRAJEWSKI, 2004 pp. 310, 317 e 319). Embora as demonstrações apliquem-se especificamente ao sistema ***P***, GÖDEL indica como demonstrações similares se aplicariam a qualquer sistema formal capaz de expressar a aritmética.

Esta proposição dá resposta justamente ao problema de se determinar a **consistência** de um sistema formal, tão valorizada pelos formalistas: *"(...) em particular, a consistência de P não é demonstrável em P, sendo assumido que P é consistente (se não fosse, claro, qualquer proposição seria demonstrável)*"[5] (GÖDEL,1992 p.70). Na opinião de BRAITHWAITE este fato foi, para HILBERT, uma notícia pior do que o teorema da incompleteza propriamente dito, pois uma das metas principais de HILBERT e de sua escola era estabelecer a **consistência** de um cálculo capaz de ser interpretado como expressando a aritmética, e assim provar a consistência do sistema dedutivo da aritmética. Para os formalistas, o segundo grande teorema contido neste artigo foi um choque ainda maior que o teorema da incompleteza. Pois este segundo teorema prova a indecidibilidade dentro de ***P*** de uma fórmula que expressa a **consistência** de ***P***. Assim, fica estabelecido que a **consistência** de ***P***, se ***P*** é consistente, não pode ser estabelecida por uma prova interna à ***P***[6] (GÖDEL,1992 p. 23).

Realmente, o problema da demonstração da consistência, a partir daqui, ficou mais complexo, pois, se o sistema que exprime a parte aritmética dos *Principia* de Russell e WHITEHEAD for consistente, então, para provar este fato, precisaremos de um sistema ainda mais abrangente. Quer dizer, se ele representasse uma formalização de toda a matemática, seria preciso um sistema ainda maior que a totalidade da matemática

[5] (...) in particular, the consistency of P is unprovable in P, it being assumed that P is consistent (if not, of course, every statement is provable).

[6] A principal aim of HILBERT and his school had been to establish the 'consistency' of a calculus capable of being interpreted as expressing arithmetic, and thus to prove the consistency of a deductive system of arithmetic. To them the second great theorem contained in this paper was even more a shock than the 'Unprovability' Theorem. For this second theorem proves the undecidability within P of a formula expressing the 'consistency' of P, thus showing that the 'consistency' of P, if P is consistent, cannot be established by a 'proof' within P.

existente para provar a sua consistência. E, transitivamente, ao encontrar tal sistema seria necessário outro ainda mais abrangente para provar a sua própria consistência.

Na prova desta proposição, GÖDEL (1992 pp.70-72) parte do fato de que para provar que um sistema formal é consistente, temos que mostrar que não se pode demonstrar nenhum par de fórmulas **A** e **~A**. No entanto, é mais fácil fazer a demonstração indiretamente, mostrando que alguma fórmula em particular não é demonstrável no sistema. KNEEBONE (2001, pp. 210-211) coloca isto na forma de um teorema:

Teorema 1: *Um sistema formal F é consistente se e somente se existe uma fórmula do sistema que não é demonstrável*.

Como o teorema acima atesta uma equivalência ($\Leftrightarrow$), a prova deste deve ser feita em duas etapas:

a) ($\Rightarrow$) Se F é consistente, para uma escolha arbitrária de uma fórmula **A**, é impossível que ambos **A** e **~A** sejam demonstráveis. Daí segue que ou **A** ou **~A** é a fórmula não-demonstrável do sistema;
b) ($\Leftarrow$) Se F é inconsistente, então há um par de fórmulas **A** e **~A** que são ambas demonstráveis. Mas da conjunção **A** e **~A** podemos derivar qualquer fórmula, como mostra a demonstração do item **2.1** deste capítulo.

GÖDEL, dentro do seu formalismo, define consistência desta maneira, ou seja, define que o sistema é consistente se existe uma fórmula que não pode ser demonstrada no mesmo (GÖDEL, 1992 p.70 n 63).

Desenvolvendo o raciocínio, GÖDEL chega ao resultado que diz que

CON $\Rightarrow$ **G** é demonstrável em ***P*** **(1)**, onde

- ***P*** é o sistema formal usado por GÖDEL na sua demonstração;
- CON é a fórmula proposicional que expressa que ***P*** é consistente;
- **G** é a sentença de GÖDEL.

GÖDEL, neste ponto, pondera que, se CON – a fórmula que expressa a consistência de ***P*** – fosse demonstrável em ***P***, a sentença **G** seria, também, demonstrável em ***P***. Ele afirma isto baseado no *modus ponens*. Por este motivo, seguiria, por **(1)** que ***P*** não é consistente, já que, pelo teorema VI (item anterior) se a sentença **G** fosse demonstrável, ***P*** seria inconsistente. Chega-se, assim, a uma contradição (GÖDEL, 1992 p,71), o que nega a hipótese de que a consistência de ***P*** possa ser demonstrada.

Os resultados da demonstração de GÖDEL constituem-se numa das mais importantes contribuições em toda a história da lógica e a importância dos mesmos atinge áreas além da matemática. Eles mostram que sempre haverá verdades aritméticas que não podem ser formalmente deduzidas, o que implica que uma abordagem axiomática não expressará todo o domínio das verdades aritméticas. NAGEL & NEWMAN (1958, p.99) inferem a partir deste fato que a maneira de um matemático demonstrar teoremas não coincide exatamente com a aplicação de regras de produção de cadeias de caracteres num sistema formal. Mais ainda: os resultados de GÖDEL mostrariam limites na tentativa de construção de uma máquina com inteligência equiparável à humana, para o caso de a máquina ser redutível a um sistema formal.

GÖDEL, em carta para ZERMELO[7], datada de 12 de outubro de 1931 comenta (pp. 7-8) que a partir do seu resultado sabemos que a classe de fórmulas demonstráveis está estritamente contida na classe de fórmulas verdadeiras, já que seu teorema mostra uma fórmula verdadeira mas não demonstrável GRATTAN-GUINESS (1979).

Refletindo também sobre estes resultados, TARSKI (1969) salienta a diferença que atualmente temos de fazer entre os conceitos de demonstrabilidade e de verdade. Para ele, os triunfos iniciais do programa formalista foram logo seguidos de uma queda. Isto porque, embora a prova ainda seja o único método que temos para averiguar verdades na matemática, os resultados mostram que em nenhum domínio da mesma a noção de demonstrabilidade pode ser um substituto perfeito da noção de verdade (TARSKI, 1969 p.77). Ao considerar que uma teoria pode ser ampliada acrescentando-se novos axiomas e/ou novas regras de derivação, mas que sempre haverá sentenças

[7] Esta carta é parte da correspondência que GÖDEL troca com Zermelo após um ciclo de palestras em que os dois participam em setembro de 1931 no *Deutsche Mathematiker-Vereinigung*, em Bad Elster. Esta correspondência é reproduzida no original por GRATTAN-GUINESS (1979).

verdadeiras não demonstráveis, TARSKI coloca a verdade como um limite ideal que nunca poderá ser alcançado, mas do qual podemos tentar nos aproximar gradualmente. TARSKI sugere que o mesmo deve estar acontecendo na área do conhecimento empírico, embora por razões bem diversas (idem).

KNEALE & KNEALE (1971 p.735) pensam como TARSKI, dizendo que a verdade em matemática não pode ser igualada à demonstrabilidade, não mais do que a verdade em estudos empíricos pode ser igualada com verificabilidade, e que qualquer doutrina filosófica que sugere o contrário deve estar equivocada.

## 2.3 – A Máquina de TURING

Analogias entre o teorema de GÖDEL e limitações do computador podem ser feitas examinando-se restrições da máquina de TURING. Alan TURING (1912-1954). Em 1935 TURING tinha tomado conhecimento dos resultados de GÖDEL e do décimo problema de HILBERT. Tendo trabalhado neste problema, apresentou seu artigo de 1936 com o título de "*On computable numbers, with an application to the Entsscheidungsproblem*" (Sobre números computáveis, com uma aplicação ao problema do procedimento de decisão)

Neste artigo, TURING criou um modelo de computador genérico chamado "Máquina de TURING" e concebe uma definição para o conceito de "procedimento mecânico" que HILBERT exigia para seu problema do procedimento de decisão. Como CHAITIN (2002 p.167) lembra, HILBERT não chegou a dar uma definição precisa de "procedimento mecânico". GÖDEL atribui a TURING o fato de termos agora "*uma definição precisa e incontestavelmente adequada de conceito geral de sistema formal*" (GÖDEL, 1979 p.355). Para GÖDEL a obra de TURING fornece uma análise do conceito de "processo mecânico", ou "algoritmo" e mostra que este conceito é equivalente ao de uma "máquina de TURING" (idem). Ele continua, dizendo que um sistema formal pode ser simplesmente definido como sendo qualquer processo mecânico de produzir fórmulas, sendo estas chamadas fórmulas demonstráveis (idem).

Para SOARE (1996, p.291) o artigo de TURING é importante porque:

**1)** criou a idéia de um agente humano que faria as operações de maneira idealizada. Este agente reuniu as concepções intuitivas de "funções produzidas por um procedimento mecânico;

**2)** formalizou um aparato muito simples – a máquina de TURING – e provou a equivalência entre este aparato e o agente humano idealizado;

**3)** provou que o décimo problema de HILBERT não tem solução;

**4)** propôs uma máquina universal, que teria a capacidade de imitar o funcionamento de qualquer outra.

Para CLELAND (2004, p.211) a similaridade entre a concepção de computabilidade de TURING e a concepção formalista da matemática é óbvia. Isto porque os símbolos da máquina de TURING são estruturas sem significado que são manipuladas somente com respeito à sua forma.

A máquina de TURING é definida de modo formal por LOECKX (1972, pp.27-28) e também por LEWIS e PAPADIMITRIOU (2000, p.174). Estes a apresentam informalmente como "*um controle finito, uma fita e uma cabeça que pode ser utilizada para ler ou gravar na fita*" (op. cit. p. 172). Mas a explicação de BOOLOS (1974, pp. 20-22) é mais detalhada. Ele começa ao dizer que a máquina de TURING utiliza uma fita dividida em quadrados e de tamanho indefinido em ambas as direções. TURING imaginou uma fita ideal de tamanho infinito para não limitar a computação ao tamanho da memória disponível. Cada quadrado da fita pode estar em branco ou conter exatamente um de um número finito de símbolos $S_1$,$S_2$,.....,$S_n$. LOECKX (1972, p.11) chama o conjunto destes símbolos de **vocabulário** e ressalta que obrigatoriamente deve haver um símbolo fora deste conjunto que também pode preencher um quadrado da fita. É o símbolo que representa o "branco" ou o "vazio" na fita.

A cada estágio da computação, a cabeça está sobre um quadrado da fita e a máquina de TURING é capaz de detectar qual símbolo está neste quadrado. A máquina pode apagar um símbolo do quadrado e escrever outro símbolo no mesmo. Também pode movimentar a fita nos dois sentidos, movendo um quadrado para a esquerda ou para a direita.

A máquina de TURING tem um modo de funcionamento de forma que a cada estágio da computação ela está em um determinado estado interno, que corresponde a estar executando uma determinada instrução. O número de estados e, conseqüentemente, o de instruções, é sempre finito. As instruções têm uma forma condicional, pois são da forma: se o estado atual é **A** e o símbolo lido é **B** então faça o conjunto de operações **C**. As ações que uma máquina de TURING executa podem ser divididas em atos internos e atos externos. Os atos internos consistem nas mudanças de estado. Supondo que o vocabulário da uma máquina de TURING tem **n** símbolos – $S_1$,$S_2$,.....,$S_n$ – há exatamente **n** + 4 atos externos que ela pode fazer BOOLOS (1974, p. 22):

1) parar a computação;
2) mover a fita um quadrado para a direita;
3) mover a fita um quadrado para a esquerda;
4) escrever o símbolo que representa o "branco" ou "vazio" no quadrado que está sendo lido;
5) apagar o símbolo que estiver escrito e escrever S1 no quadrado que está sendo lido;
6) apagar o símbolo que estiver escrito e escrever S2 no quadrado que está sendo lido;

.
.
.

**n** + 4) apagar o símbolo que estiver escrito e escrever $S_n$ no quadrado que está sendo lido.

Assim, dependendo de qual estado esteja e de qual símbolo foi lido, a máquina executará uma destas **n** + 4 operações externas. A menos que tenha parado – no caso de ter executado a instrução 1 – a máquina de TURING executará também um ato interno, que é determinar qual será a próxima instrução a executar. Garante-se, desta forma, que, a cada etapa, o estado da máquina e o símbolo lido determinam exatamente a próxima ação e o próximo estado.

A relação entre estados e símbolos pode ser mais claramente compreendida pelo que PUTNAM (1964) chama de "tabela da máquina". Esta tabela é construída da seguinte maneira: cada linha corresponde a um símbolo do vocabulário e ainda há uma linha extra para o símbolo de "branco" – ou "vazio"; cada coluna a um estado da máquina; e cada elemento da tabela é uma instrução, como por exemplo:

- **$S_1$EA** é a instrução que diz: escreva o símbolo $S_1$ no quadrado que está sendo lido, vá para o quadrado imediatamente à esquerda do atual e mude para o estado A;
- **$S_5$DC** é a instrução que diz: escreva o símbolo $S_5$ no quadrado que está sendo lido, vá para o quadrado imediatamente à direita do atual e mude para o estado C;
- **$S_5$CC** é a instrução que diz: escreva o símbolo $S_5$ no quadrado que está sendo lido, permaneça sobre o quadrado atual e mude para o estado C.

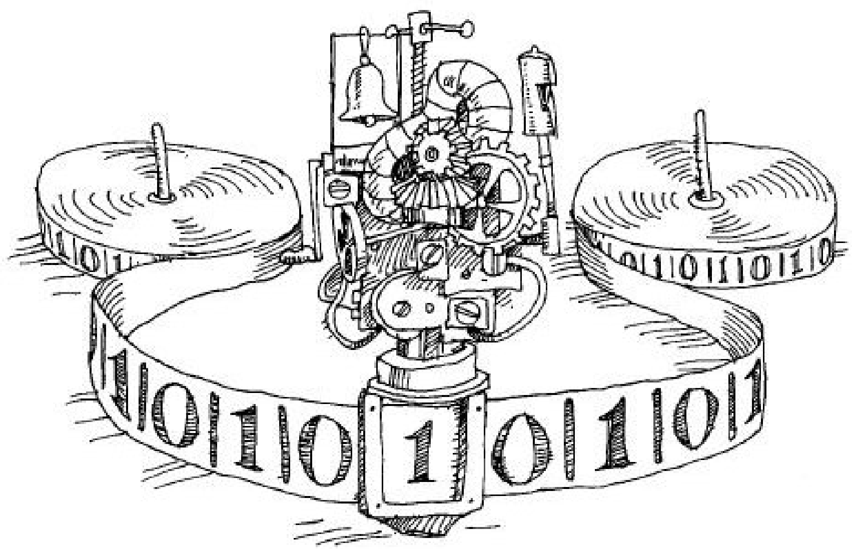
Como vemos nos exemplos, cada instrução dita que a cabeça de leitura/impressão mova-se um quadrado para a esquerda (**E**), um quadrado para a direita (**D**), ou que fique sobre o mesmo quadrado (**C**). Uma tabela deste tipo tem o formato abaixo:

| | **A** | **B** | **C** | **D** | **. . .** |
|---|---|---|---|---|---|
| **$S_1$** | **$S_4$EC** | **$S_5$DC** | **$S_2$CB** | **$S_2$CC** | **. . .** |
| **$S_2$** | **$S_3$DK** | **$S_5$CB** | **$S_4$DH** | **$S_2$CB** | **. . .** |
| **$S_3$** | **$S_1$CC** | **$S_2$EG** | **$S_1$CK** | **$S_4$EF** | **. . .** |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| **"vazio"** | **$S_1$DK** | **$S_3$EF** | **$S_2$DI** | **$S_5$CB** | **. . .** |

A máquina-exemplo descrita pela tabela acima começaria a computação no estado **A** e, caso o símbolo no primeiro quadrado fosse **$S_1$**, ela apagaria este símbolo e escreveria no mesmo quadrado o símbolo **$S_4$**. Após isso, posicionaria a cabeça sobre o quadrado imediatamente à esquerda e então mudaria para o estado **C**. Caso encontrasse no primeiro o símbolo **$S_2$**, ela apagaria este símbolo e escreveria no mesmo quadrado o símbolo **$S_3$**. Após isso, posicionaria a cabeça sobre o quadrado imediatamente à direita e então mudaria para o estado **K**. Assim, a tabela descreve o comportamento da máquina de TURING para cada estado (coluna) e para cada símbolo do vocabulário, incluindo o símbolo do “branco”.

Qualquer máquina que pode ser descrita por uma tabela como a do exemplo acima é uma máquina de TURING (PUTNAM, 1964 p. 76).

CHAITIN (2002, p. 168) nos fornece um esquema bem-humorado de uma máquina de TURING:



A máquina de TURING é um computador para uso geral e, embora tenha sido idealizada no início do séc. XX, todos os computadores digitais podem ser reduzidos, para a análise a seguir, a uma máquina de TURING. Como é um máquina de uso geral, ela poderia fazer o que qualquer outra máquina faz.

Mas,neste ponto, TURING mostra que existem limites para o que esta máquina geral pode fazer. Há um problema que nenhuma máquina de TURING pode resolver. É o problema da parada, ou seja, o de prever se uma dada máquina de TURING chegará a um resultado e então parará. Se impusermos um limite de tempo, este problema se torna muito fácil. Basta deixar o programa funcionar durante este limite e verificar se ele pára ou não. Mas no caso geral, sem imposição de um limite para o tempo de funcionamento, o problema não tem solução (CHAITIN, 2002 P. 168).

A prova de que o problema da parada é insolúvel é feita por uma demonstração por contradição (do latim *reductio ad absurdum*). Neste tipo de prova deriva-se uma contradição de uma hipótese e daí se conclui que a hipótese é falsa (DELONG, 1971 p.7-9). Com este objetivo, vamos supor (CHAITIN, 2002 p. 168) que temos um programa que verifica se qualquer outro programa pára. Esta é a hipótese que queremos mostrar que é falsa. Vamos chamá-lo de ***V***, já que ele verifica algo sobre outros programas. Este programa forneceria sempre um de dois tipos de resultado, mutuamente excludentes:

- o resultado "sim" se o programa que estiver sendo posto em teste parasse; ou
- o resultado "não" se o programa em teste continuasse para sempre.

Seguindo o raciocínio, vamos imaginar outro programa, chamado **P,** que usa o resultado dado por ***V*** para avaliar um programa qualquer. Vamos supor também que colocamos em **P** instruções tais que ele entra em *loop*[8] caso o programa que está sendo avaliado termine (pare); e também colocamos instruções tais que ele pare quando recebe de ***V*** a informação de que o programa em teste não pára. É claro que o programa **P** tem a informação que diz se programa em teste pára ou não, pois ele usa o programa verificador ***V***.

Agora vamos supor que alimentamos **P** com uma cópia de suas próprias instruções. Agora ele próprio é o programa em teste. O que acontece? Se ele pára, como foi programado para entrar em *loop* se o programa em teste pára, ele entrará em *loop* e não parará, o que é uma contradição. Mas, se ele não pára, o programa ***V*** informará isto a **P** e, então, ele não entrará em loop, parando quando obtiver esta informação de ***V***, o que também é uma contradição. Assim, como a partir da nossa hipótese chegamos a uma contradição, podemos concluir que a mesma é falsa, ou seja, não é possível termos um programa que verifica se outro programa pára.

Podemos fazer, como PENROSE (1997, p.64), uma ligação entre um indecidível aritmético e o problema da parada de TURING. Seja a **conjectura de GOLDBACH**, que diz que todo e qualquer número par maior do que dois pode ser escrito como a soma

[8] chama-se *loop* uma seqüência fechada de instruções na qual a última instrução leva à primeira, fazendo com que a máquina fique repetindo a seqüência, teóricamente, para sempre.

de dois números primos. Ela foi afirmada por GOLDBACH em 1742 mas não foi demonstrada até hoje.

Para decidirmos se determinado número natural é primo, precisamos apenas testar sua divisibilidade por números menores que ele próprio, uma questão de cálculo finito. Podemos também pensar numa máquina de TURING que, para cada número par, fizesse a seguinte verificação: testar todos os modos possíveis de escrevê-los como soma de dois primos (ex.: 10 = 3 + 7 ou 5 + 5). Usando este exemplo, PENROSE (op. cit.) mostra que se fosse possível resolver o problema da parada, então também estaríamos "decidindo" se a conjectura de GOLDBACH é falsa ou verdadeira. Pois, nossa máquina só deve parar quando chegar a um número par para o qual nenhum dos pares em que esse número se divide consiste em dois primos. Nesse caso, teríamos um exemplo contra a conjectura de GOLDBACH, ou seja, um número par maior do que *dois* que **não pode** ser escrito como a soma de dois primos. Mas TURING prova que não há como dizer se a máquina pára ou não. Ou, de um modo geral, que não existe nenhum algoritmo que decida a questão da parada das máquinas de TURING.

# III – A argumentação de LUCAS

Em "*Minds, machines and Gödel*", publicado em 1961, LUCAS (1964) recorre ao teorema da incompleteza – o que chamamos de primeiro resultado de GÖDEL no capítulo anterior – para argumentar contra a corrente filosófica que ele chama de "*mechanism*". Esta corrente, que aqui será traduzida como **mecanicismo** postula que mentes podem ser explicadas como máquinas (LUCAS, 1964 p.43). Como ele mesmo diz, "*o teorema de GÖDEL aplica-se à máquinas porque é da essência de ser uma máquina que ela seja uma concretização de um sistema formal*" (idem, p.44). No seu artigo, LUCAS cita sugestões de HEMPEL e ROGERS para tornar um computador livre da restrição que GÖDEL provou ser inerente aos sistemas formais. O próprio LUCAS contra-argumenta, mostrando que as sugestões não procedem. A contra-argumentação de LUCAS usa reflexões sobre o primeiro resultado de GÖDEL.

Aqui será usado também o segundo resultado de GÖDEL para corroborar a apresentação de LUCAS. E como exemplo e exercício sobre alteração em sistemas de axiomas, será usada o modo com que as geometrias não-euclideanas surgiram.

## 3.1 – As Geometrias Não-Euclideanas

A questão das geometrias não-euclideanas é introduzida aqui devido à sua importância na argumentação central deste capítulo, que expõe o fato de que mesmo se um sistema formal fosse capaz de aceitar novos "axiomas"[1], conforme a sugestão de LUCAS, ainda assim não poderia reproduzir a criatividade. Isto porque, como veremos a seguir, o advento destas geometrias está relacionado à alteração do significado de um ou mais axiomas, o que equivale a sua substituição por outros, ao invés simplesmente da adição de novos axiomas.

[1] As aspas foram introduzidas para manter a exatidão: sistemas formais não possuem **axiomas**, e sim **fórmulas iniciais**, sendo o termo "axioma" relativo ao sistema dedutivo que é uma interpretação do sistema formal.

MARTINS (1995 p.67) vê as geometrias não-euclideanas como o formalismo matemático básico utilizado atualmente na teoria da relatividade geral e faz um estudo sobre tentativas de aplicar à física alguns resultados destas geometrias, já no século XIX. No entanto, deixa claro que o surgimento das mesmas partiu de estudos puramente formais (idem, p.73). GRAY (1979 pp. 237, 239, 243, 249, 252 e 253) ressalta que as questões envolvidas nas geometrias não-euclideanas não se restringem apenas a considerações formais da fundamentação da geometria, ou à exigência crescente de rigor no método axiomático. Para ele, o passo crucial neste desenvolvimento é a introdução das técnicas de análise, culminando naquelas da geometria diferencial.

EUCLIDES (séc. III a. C.), através da lógica dedutiva de Aristóteles, percebe o elo entre o conhecimento geométrico empírico e a geometria como estudo sistemático, com o estabelecimento de definições, postulados e axiomas (EUCLID, 1956 p.122). Este fato sugere que EUCLIDES compreendeu que *"a estruturação do conhecimento geométrico deveria começar por uma assepsia na linguagem, com o esclarecimento das noções utilizadas de modo intuitivo"* (MACHADO, 1991 p.138). O mérito de EUCLIDES não é tanto o de ter descoberto soluções para problemas, mas sim a maneira como sistematizou o conhecimento matemático grego da época. WANG (1954 p.241) considera indubitável que esta é a primeira vez em que ocorreu uma tal sistematização de conhecimentos até então isolados. Assim, EUCLIDES escreve "Os Elementos", uma coleção de livros, que parte de cinco postulados (EUCLID, 1956 pp.154-155):

1) Dados dois pontos quaisquer, existe um segmento de reta que os liga;
2) Uma linha pode ser estendida em ambas as direções indefinidamente;
3) Dados um ponto e uma distância, pode-se desenhar um círculo no qual todos os pontos têm esta distância do primeiro;
4) Todos os ângulos retos são iguais entre si;
5) Se, dados um par de retas, uma terceira se coloca sobre elas e se esta terceira faz ângulos interiores do mesmo lado somarem menos do que dois ângulos retos, então as duas primeiras, se prolongadas indefinidamente, se encontrarão neste lado em que a soma dos ângulos é menor que dois ângulos retos.

Além das cinco proposições acima, que EUCLIDES denomina "postulados", ele usa cinco "noções do senso comum" (idem, p. 155):

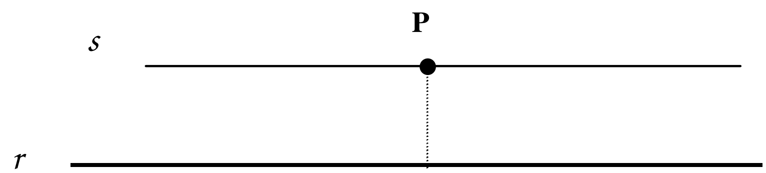
1) Duas coisas iguais a uma terceira são iguais entre si;
2) Se parcelas iguais forem adicionadas a quantidades iguais os resultados serão iguais entre si;
3) Se parcelas iguais forem subtraídas de quantidades iguais os resultados serão iguais entre si;
4) Coisas que coincidem uma com a outra são iguais;
5) O todo é maior que as partes.

Devemos ter em mente que a distinção que EUCLIDES fazia, entre "postulados" e "noções do senso comum", não se aplica atualmente, pois considera-se todas estas proposições iniciais numa mesma categoria, e considera-se os termos *axioma* e *postulado* equivalentes. EUCLIDES usa ainda 23 definições, criando, assim, um sistema axiomático dedutivo para o conhecimento geométrico. Segundo REICHENBACH, *"a geometria tornou-se o protótipo de uma ciência demonstrável, o primeiro exemplo de um ideal de rigor científico"* (REICHENBACH, 1958 p.1).

EUCLIDES inaugura, assim, o chamado método axiomático que é usado até hoje nas demonstrações matemáticas, tendo servido como guia e modelo de rigor e disciplina para gerações de matemáticos. O método caracteriza-se justamente por permitir que demonstrações de um grande número de teoremas seja provado a partir de um número reduzido de axiomas, ou "verdades primeiras". Um sistema axiomático deve, então começar com uma lista dos axiomas que serão usados para as demonstrações. As regras usadas pelas demonstrações são dadas pela lógica.

Desde a publicação dos "Elementos" de EUCLIDES houve questionamentos sobre o seu quinto postulado (o postulado das paralelas), questionamentos estes que se intensificaram no final do século XVII, quando um grande número de matemáticos apostou na idéia que ele seria um teorema, ou seja, poderia ser demonstrado a partir dos demais. Como em outros estudos, não será usada aqui a forma original do quinto postulado, mas sim a versão de Playfair, que é equivalente (GREENBERG, 1994 p.18-19) e que simplifica o texto de EUCLIDES:

Por um ponto **P** fora de uma reta ***r*** só se pode traçar uma única paralela a essa reta.




A questão fundamental era: os postulados de EUCLIDES representavam proposições não redutíveis a outras mais elementares ou existiria uma possibilidade de reduzi-los ainda mais, até afirmações que não necessitam demonstração? (REICHENBACH, 1958). Sobre os quatro primeiros postulados, não houve questionamentos. A questão da dúvida sobre ser demonstrável ou não referia-se em particular ao quinto, conhecido como o postulado das paralelas, pois a afirmação de que duas linhas retas não se cruzam em uma distância finita soava complexa demais, ou seja, poderia ser um teorema. Na época de EUCLIDES os postulados eram afirmações auto-evidentes, ou seja, de forte apelo intuitivo. O quinto postulado, por ser de formulação mais complexa, aparentemente não se encaixaria nesta categoria. Possivelmente esta foi a razão das várias tentativas feitas por matemáticos no sentido de deduzir o quinto postulado a partir dos demais.

Entre os muitos matemáticos que tentaram, em vão, demonstrar o postulado das paralelas a partir dos outro nove, destaca-se Girolamo Saccheri (1667-1733), que tentou tal demonstração pelo "método indireto" de prova, ou prova por absurdo: para provar A, assume-se que A é falso e procede-se para mostrar que esta suposição leva a uma contradição. Se chegarmos realmente a uma contradição, mostramos, indiretamente, que A é verdadeiro.

Como a versão de Playfair diz que existe apenas uma paralela a ***r*** que passa por um ponto fora de ***r***, negar isto implica em uma das alternativas abaixo:

a) há mais de uma paralela;
b) não há nenhuma paralela.

Se estas duas suposições levassem a uma contradição, então o quinto postulado seria verdadeiro, pois sua negativa seria falsa. O método de Saccheri, além de correto, levou-o a resultados que poderiam tê-lo tornado o pai das geometrias não-euclideanas, mas a época em que vivia não permitia a interpretação destes resultados e ele acabou considerando-os contraditórios. No entanto, estes resultados só são contraditórios do ponto de vista do nosso senso comum. Dentro do sistema de axiomas proposto, não há contradição lógica alguma.

Foi necessário um século para que alguém tivesse a visão suficiente para perceber a conclusão correta. Carl Friedrich Gauss (1777-1855) acaba percebendo que o conjunto de postulados escolhidos para uma determinada área de estudo é, de certa forma, arbitrário. Muda-se o conjunto de postulados, acrescentando-se ou retirando-se um ou mais, ou altera-se o enunciado de algum, e surge uma nova estrutura. O fato de a estrutura estar ou não de acordo com nosso senso comum, como vemos ou sentimos o mundo, é irrelevante, pois o que importa é a consistência lógica. Dois outros matemáticos, sem renome na época, chegaram às mesmas conclusões: O húngaro Janos Bolyai (1802-1860) e o russo Nicolai Lobachevsky (1793-1856).

A nova geometria de Gauss, Bolyai e Lobachevsky, chamada *geometria hiperbólica*, mantém os outros 9 postulados e substitui o quinto pela afirmativa de que, por um ponto fora de uma reta , passam pelo menos duas retas paralelas à mesma. Pouco mais tarde, Riemann (1826-1866) substitui o quinto postulado pela segunda alternativa de negá-lo, ou seja, por um ponto fora de uma reta, não passa nenhuma paralela à mesma e também faz uma alteração na definição de ponto e no segundo postulado de EUCLIDES. Riemann foi o primeiro a substituir o segundo postulado (que diz que uma reta pode ser estendida indefinidamente nas duas direções). Ele o substitui por um que diz que uma "reta" não é delimitada.

GUILLEN (1987, p.119) reflete sobre a questão da verdade, quando conta o aparecimento das geometrias não euclideanas. Na época, a geometria de EUCLIDES era tida como uma bíblia e, portanto, o surgimento da novidade foi tratado como uma heresia; aos matemáticos não parecia possível que a geometria euclideana fosse apenas **uma de muitas** geometrias, e não **a** geometria do universo. Tampouco parecia aceitável que uma geometria matematicamente válida fosse baseada num postulado que não é uma verdade evidente. Parecia absurdo que Bolyai, Gauss e Lobatchevski pudessem inventar um postulado que nada tinha que ver com a realidade nem com o senso comum e derivar dele um sistema matemático logicamente válido. Forçoso era concluir ser a matemática, ela também, uma invenção. Estes fatos levaram a uma substituição da procura da verdade absoluta pela exigência de consistência de um sistema axiomático.

## 3.2 – A discussão de Lucas

LUCAS (1964) cita o argumento de HEMPEL e ROGERS (op. cit, p. 50) sobre o que estes propõem ser um bom modelo de mente. Para eles, um modelo de mente tem que permitir que algumas afirmativas não demonstradas sejam aceitas. Por exemplo, permitindo que o computador fosse programado para incluir, de forma provisória, algumas proposições ainda não provadas, colocando-as na sua lista de axiomas. Caso fosse verificado que uma determinada proposição incluída desta forma leva a alguma contradição, esta proposição seria retirada da lista de axiomas e colocaria-se a sua negação nesta mesma lista.

Assim, esta máquina seria capaz de incluir fórmulas verdadeiras que não são geradas de seus axiomas pelas regras do sistema, libertando-se da limitação ditada pelo teorema de GÖDEL.

O próprio LUCAS (1964, p. 51) contra-argumenta que uma máquina seguindo este esquema teria que ter um **critério** para selecionar quais proposições poderiam ser incluídas na sua lista de axiomas. Caso contrário esta máquina poderia incluir uma determinada proposição e também sua negação, tornando-se assim inconsistente.

Este critério também teria que evitar que fosse aceita, por exemplo, a negação da fórmula de GÖDEL. Já que esta fórmula é verdadeira, sua negação, pelo princípio do terceiro excluído, é falsa.

A crítica que LUCAS faz ao argumento de HEMPEL e ROGERS é que estes critérios teriam de ser inseridos na forma de regras pelas quais o sistema testaria sua aceitação. Sendo assim, justamente por estar totalmente regido por regras de novo, o teorema de GÖDEL se aplicaria sobre o sistema "enriquecido". De fato, como o próprio GÖDEL ressalta, proposições indecidíveis podem tornar-se decidíveis em sistemas mais completos, mas o seu teorema mostra que outras proposições indecidíveis seriam encontradas nestes novos sistemas (GÖDEL,2001b p.237 e GRATTAN-GUINESS, 1979, p.296-297).

## 3.3 – Aplicação da proposta de Lucas às geometrias não-euclideanas

Como vimos anteriormente, a geometria hiperbólica surge ao se substituir o quinto postulado de EUCLIDES por uma das maneiras de negá-lo, ou seja, postulando que por um ponto fora de uma reta passam pelo menos duas paralelas à esta.

Esta nova geometria surge como resultante de um lado, da frustração causada pelos sucessivos fracassos em provar o quinto postulado a partir dos outros nove e, de outro lado, de uma abertura mental em aceitar um enunciado contrário ao que diz nosso senso comum. A questão que se coloca é se é possível inserir tal processo de "abertura" nas regras de um sistema formal. Podemos ver que este processo é de natureza diferente do comentado por LUCAS no item anterior, pois não se trata de permitir que o sistema aceite provisoriamente novos axiomas, mas sim de permitir que ele **altere** um determinado axioma.

É evidente que tal alteração não pode ser aleatória: o sistema não poderia alterar um axioma de tal forma que a combinação do axioma alterado com os demais o tornasse inconsistente. Isto obrigaria que, a cada alteração, fosse feita uma verificação de consistência. No entanto, o segundo resultado de GÖDEL nos diz justamente que, se um sistema é consistente, sua consistência não pode ser demonstrada no mesmo. Assim, nosso sistema estaria sem saída pois, para determinar a sua própria consistência com o axioma alterado necessitaria de um outro sistema mais abrangente, ou de alguma interferência humana, para provar sua consistência.

A geometria elíptica de Riemann surge através de um raciocínio mais sofisticado. Nesta, além da substituição do quinto postulado pelo segundo modo de negá-lo - por um ponto ***P*** fora de uma reta não passa nenhuma paralela à esta, Riemann modifica também o segundo postulado e a definição de ponto. Embora para EUCLIDES as definições, os postulados e os axiomas representassem pontos de partida de naturezas diferentes, hoje considera-se os três tipos de enunciados como axiomas do sistema dedutivo. Isto fica claro na definição de proposição de um sistema dedutivo dada por ERNEST (1998, p.6): Cada proposição é um axioma retirado de um conjunto de axiomas previamente estipulado, ou é derivado, de uma ou mais proposições anteriores na seqüência, através de uma regra de inferência. O termo **conjunto de axiomas** deve aqui ser entendido de uma maneira ampla, de modo que inclua todas as proposições sem demonstração inseridas numa prova, incluindo axiomas, postulados, e definições.[2]

É claro que dentro de um sistema formal, que possui apenas *símbolos*, *fórmulas iniciais*, e *regras de produção*, as três categorias de EUCLIDES ficam inclusas nas *fórmulas iniciais*.

Assim como LUCAS prescreve a necessidade da inclusão de regras para avaliar e permitir a inserção de novas fórmulas, vemos, neste outro modo de alterar o sistema, que teremos de ter os mesmos cuidados. Se quisermos substituir uma fórmula inicial por outra também é necessário inserir um conjunto de regras para dizer qual ou quais fórmulas iniciais seriam modificadas e outras regras ainda para ditar quais as modificações, qual a seqüência de modificações, etc. O sistema assim estaria novamente regido por um conjunto de regras, somente que agora tratar-se-ia de um conjunto de regras ampliado. Desta forma, o teorema de GÖDEL se aplicaria ao mesmo, o que mostra que continuaria incompleto. Além disso, pode-se ir além da objeção de LUCAS se lembramos que tanto a inclusão como a alteração de um único axioma pode tornar o sistema inconsistente e, de acordo com o segundo resultado de GÖDEL, o sistema não terá como avaliar sua consistência, pois precisaria de verificação externa.

Estas considerações, deixam de fora a questão da independência dos axiomas, outro requisito de um sistema formal. Segundo TARSKI (1995, p.131) , um sistema é

[2] Each statement is an axiom drawn from a previously stipulated set of axioms, or is derived by a rule of inference from one or more statements occurring earlier in the sequence. The term ***set of axioms*** should be understood broadly, to include whatever statements are admitted into a proof without demonstration, including axioms, postulates, and definitions.
(grifos do autor)

classificado como independente quando não contém nenhum axioma supérfluo, ou seja, não contém nenhum axioma que pode ser derivado dos restantes. Neste caso, o nome axioma é incorreto, pois representa um teorema, já que é deduzido dos verdadeiros axiomas.

É claro que a escolha do que é "verdadeiro axioma", ou do conjunto de "verdadeiros axiomas" e do que é "teorema" é facultativa, pois nós temos um alto grau de liberdade na seleção dos termos primitivos e dos axiomas. E não é apenas por motivos teóricos que decidimos selecionar um determinado sistema de axiomas, em detrimento de qualquer outro possível sistema eqüipolente; outros fatores atuam nesta escolha – práticos, didáticos, e mesmo estéticos. Algumas vezes é uma questão de escolher os axiomas mais simples possíveis (TARSKI, 1995 pp.130-131).

## 3.4 – Outros pontos colocados por LUCAS

LUCAS coloca, usando também o teorema de GÖDEL, que embora haja máquinas que geram verdades aritméticas até de um modo mais eficiente que um humano, em um determinado ponto elas falham: no fato de que "*para cada máquina há uma verdade que ela não pode produzir como verdadeira, mas que uma mente pode*" (LUCAS, 1964 p.47)[3]. Este fato mostra que uma máquina não pode ser um modelo adequado para a mente. A "máquina" que LUCAS usa em sua argumentação é, como vimos, um computador executando um algoritmo que representa um sistema formal como foi definido em seções anteriores. Ele esclarece, como um corolário, que nós poderemos construir uma máquina que realize **qualquer processo** que imite a mente, mas que não podemos construir uma máquina que realize **todos os processos** que imitem a mente.

Prevendo a objeção dos que diriam que é uma contradição dizer que uma máquina pode simular um determinado processo que imite a mente, qualquer que seja este processo, mas que não pode simular **todos os processos**, ele se antecipa mostrando que "*não há nenhuma contradição entre o fato de que para qualquer número natural*

[3] for every machine there is a truth which it cannot produce as being true, but which a mind can.

*pode ser produzido um número maior e o fato de que não se pode produzir um número maior que todos os números*" (Op. cit. p.47-48)[4].

Uma segunda tentativa de fugir da limitação que GÖDEL provou seria: para garantir que o sistema pode produzir a fórmula de GÖDEL (justamente a que foi provada que não pode ser obtida pela aplicação das regras do sistema a partir dos seus axiomas), por que não construir o sistema de tal modo que tenha todas os outros axiomas e regras que já tinha, mas **acrescentando** uma regra que realiza o procedimento de GÖDEL? Assim a fórmula de GÖDEL poderia ser construída pelo sistema, pois seria um de seus axiomas.

A saída também não funciona, pois o novo sistema "reforçado" faria coisas que o antigo não faz, mas estaria sujeito da mesma forma a ter, pelo procedimento de GÖDEL, a sua própria fórmula não-demonstrável dentro do sistema. Como LUCAS resume: "*pois mesmo que um conjunto infinito de axiomas fosse adicionado, eles teriam que ser especificados por alguma regra finita ou especificação, e esta regra ou especificação extra poderia então ser percebida por uma mente que estudasse o sistema enriquecido*"[5] (LUCAS, 1964 p.49 e GÖDEL, 2001b p.237). Esta mente, então, poderia aplicar o procedimento de GÖDEL ao sistema e obter a fórmula não-demonstrável no mesmo.

## 3.5 – A contra-argumentação de HOFSTADTER

A questão do número infinito de axiomas comentada acima é justamente o alvo da crítica que HOFSTADTER (1980, pp. 472-479) faz a LUCAS. HOFSTADTER contrapõe que se os sistemas são limitados em relação a perceberem a veracidade da sentença de GÖDEL, também o são os humanos quando os sistemas se tornam mais e mais sofisticados. Este aumento em sofisticação ocorre ao se acrescentarem aos sistemas as suas respectivas sentenças de GÖDEL, obtendo-se assim sistemas mais

---

[4] there is no contradiction between the fact that for any natural number there can be produced a greater number and the fact that a number cannot be produced grater than every number

[5] for even if an infinite set of axioms were added, they would have to be specified by some finite rule or specification, and this further rule or specification could then be taken into account by a mind considering the enlarged formal system.

complexos. HOFSTADTER argumenta que, a partir de certo nível de complexidade do sistema formal, também os humanos perderiam a capacidade de determinar a sentença de GÖDEL respectiva ao sistema. Ele pondera que argumentar teoricamente a possibilidade de executar o processo para se conseguir a sentença de GÖDEL[6] não implica necessariamente que se saiba como fazê-lo efetivamente em todos os casos particulares.

Para HOFSTADTER, à medida que os sistemas formais – quer dizer, os programas – aumentam em complexidade, a capacidade de um humano de compor a sentença de GÖDEL começará, em algum ponto, a diminuir. Ele aponta, também, que não temos um modo algoritmico que descreva como fazer este processo. E, se não podemos mostrar explicitamente como aplicar o método de GÖDEL em todos os casos, então, para cada um de nós surgirá, eventualmente, um sistema tão complexo que simplesmente não saberemos como chegar à sua sentença de GÖDEL (HOFSTADTER, 1980 p. 475).

HOFSTADTER reconhece que este limite – para o qual não conseguiríamos compor a sentença de GÖDEL - está pouco definido, mas faz uma analogia com os pesos que uma pessoa pode levantar. Em alguns dias uma pessoa pode levantar um peso de 50 quilos, por exemplo. Outros dias a mesma pessoa não conseguirá tal feito. Todavia, em nenhum dia esta pessoa terá a capacidade de levantar um objeto de 250 toneladas. Da mesma maneira, a capacidade de cada indivíduo para achar a sentença de GÖDEL é vaga, mas haverá sistemas cuja complexidade estará além de sua capacidade de realizar o mesmo (idem).

Outro ponto em que HOFSTADTER (idem) toca baseia-se no teorema demonstrado por CHURCH e KLEENE sobre a estrutura dos "ordinais transfinitos". Os "ordinais transfinitos" são parte da teoria dos conjuntos infinitos de CANTOR, que está fora do escopo deste trabalho. Mas, para entendermos a argumentação de HOFSTADTER, podemos usar a explicação resumida de HILBERT (1977a, pp. 374-375) sobre estes ordinais. HILBERT explica que, quando contamos 1, 2, 3, ..., podemos

[6] HOFSTADTER usa o termo "*Gödelize*" para denotar o processo pelo qual se chega à sentença de GÖDEL.. Este termo. Ele credita a criação do termo a LUCAS e o considera "*um dos seus termos divertidamente irreverentes*"

ver os objetos assim denumerados como um conjunto infinito completo nesta ordem definida. Explica também que CANTOR denota esta ordem por $\omega$ e que continua a sua numeração em $\omega + 1$, $\omega + 2$,... até $\omega + \omega$ ou até mesmo $\omega .2$. Assim obtemos a seguinte lista:

$1, 2, 3, ...;$

$\omega, \omega + 1, \omega + 2, ...;$

$\omega.2, \omega.2 +1, \omega.2 +2, ...;$

$\omega.3, \omega.3 +1, \omega.3 +2, ...;$

$\omega^2, \omega^2 + 1, ...;$

$\omega^2 + \omega, \omega^2 + \omega.2, \omega^2 + \omega.3,...;$

$\omega^2.2, ...;$

$\omega^2.2 + \omega, ...;$

$\omega^3, ...;$

$\omega^4, ...;$

$\omega^\omega, ...;$

Estes são os primeiros números transfinitos de CANTOR, que são obtidos através de uma transnumeração (*Hinüberzählen*) além do infinito ordinariamente enumerável (HILBERT, 1977a, p. 375).

Voltando a HOFSTADTER, ele cita o teorema de CHURCH e KLEENE que diz que não existe um sistema de notação que dá um nome a cada número ordinal. Eles mostraram que não se pode fazer um programa para produzir todos estes nomes. Pois sempre é necessário criar um novo nome quando queremos encapsular todos os números ordinais até certo ponto já nomeados em um novo conjunto. Este conjunto seria usado para definir o novo padrão de ordinais, e precisaria um nome distinto (HOFSTADTER, 1980 p. 476). HOFSTADTER argumenta, baseado neste teorema, que a mente humana também falharia em pensar nomes para todos os novos ordinais e, assim, igualariam-se aos sistemas formais, no sentido de que ambos, mentes e sistemas formais, são incompletos.

Outros aspectos da argumentação de HOFSTADTER abordam várias questões relativas à computação, mas não são bem fundamentadas por ele, como por exemplo:

- a diferença entre *hardware* e *software* no sentido de que somente o primeiro seria uma representação concreta de um sistema formal, e não o segundo (op. cit. p.577);
- a consideração segundo a qual um programa composto de vários níveis de *software* não estaria sujeito à estrutura rígida de "axiomas + regras" dos sistemas formais (op. cit. p.578);
- a proposição "($\mathbf{P} \wedge \sim\mathbf{P}) \rightarrow \mathbf{Q}$", relativa à consistência. HOFSTADTER (idem) considera que processos logicos que levam à proposição citada – entre outras – emergiriam como conseqüência da "inteligência geral" dos programas de IA, ao passo que para LUCAS este teorema seria uma característica inevitável do raciocínio mecânico. No entanto, HOFSTADTER não mostra como esta "emergência" de proposições ocorreria num sistema computacional.

## 3.6 – Respostas de LUCAS

Em palestra na "*Turing Conference*" em 1990, LUCAS responde a estas criticas e comenta que deve ter tocado em algum "nervo exposto" ao avaliar a quantidade e diversidade das mesmas (LUCAS, 2005). Ele reconhece que há uma diferença grande entre o nivel fundamental e o nível em que agem linguagens modernas, como o PROLOG. No entanto, continua ele, isto não invalida o argumento de forma alguma. também não o invalidam a diferença entre linguagem em código de máquina – aquela que atua diretamente sobre o *hardware* – e as linguagens chamadas de alto nível, entre as quais a já citada PROLOG. Para que um programa em linguagem de alto nível possa ser executado pelo *hardware* é necessário que suas intruções sejam traduzidas para o código de máquina, o que é feito por outro programa chamado compilador.

LUCAS (idem) continua defendendo seu ponto de vista ao lembrar que um compilador é inteiramente determinístico e que qualquer seqüência de operações especificada em código de máquina pode ser unicamente especificado na linguagem de alto nível e vice-versa. Também, para poder ser considerada uma representação da mente, a máquina deve ser capaz de realizar as operações da aritmética simples. E então, neste nível, o teorema de GÖDEL já se aplica à mesma.

Quanto à argumentação de que as mentes igualariam-se em incompleteza aos sistemas formais devido à impossibilidade de dar nomes aos ordinais transfinitos LUCAS (2005) responde que o argumento é justamente a favor da sua postura em diferenciar mentes e máquinas. LUCAS cita ainda GÖDEL dizendo que este este rejeitava a tese mecanicista por considerar nossa habilidade em pensar sempre novas definições e axiomas mais fortes para a teoria dos conjuntos.

# IV – O argumento de PENROSE e contra-argumentações

Com o lançamento de *A Mente Nova do Rei*, em 1989 (PENROSE, 1997), iniciou-se um ciclo de debates e polêmicas até hoje inacabado. O livro conseguiu ser um êxito editorial e apresenta um ataque à possibilidade de uma Inteligência Artificial idêntica à inteligência humana, tese esta que doravante será referida pelo termo “IA forte”. Esta obra gera uma seqüência de análises, contra-argumentações e tréplicas de PENROSE. Tanto que ele mesmo publica, cinco anos mais tarde, *Shadows of the Mind* (PENROSE, 1996b). Este, em grande parte, é uma resposta detalhada às críticas levantadas contra as idéias lançadas no primeiro livro. Ainda em 1996, PENROSE continua sua defesa com *O grande, o pequeno e a mente humana* (PENROSE, 1998). Este último é baseado nas *Tanner Lectures*, realizadas em Cambridge em 1995 e é composto de três partes: primeiramente PENROSE expõe de maneira sucinta sua argumentação; após esta exposição seguem-se três ensaios críticos de Abner Shimony, Nancy Cartwright e Stephen Hawking. No final, PENROSE toma novamente a palavra e comenta os ensaios anteriores.

LONGAIR ressalta a originalidade da visão de PENROSE pelo fato de esta propor a reunião de aspectos da física, da matemática, da biologia e da filosofia numa nova, e ainda indefinida teoria dos processos fundamentais (PENROSE, 1998, p.9-10). A argumentação de PENROSE segue a linha de LUCAS e de SEARLE no ataque à possibilidade de uma I.A. "forte". O argumento do “quarto chinês” de SEARLE (1980) tornou-se um dos mais famosos argumentos contra a possibilidade de uma “IA forte”.

SEARLE (idem) propõe um experimento de pensamento que envolve uma conversação em chinês, da seguinte forma: suponhamos que SEARLE está fechado em um compartimento com uma abertura para entrada de perguntas em chinês e uma saída para as respostas, também em chinês. Ele deixa claro que não entende uma palavra sequer de chinês, mas tem um grande e completo livro de regras que dizem que caracteres chineses ele deve concatenar para gerar uma resposta para cada determinada pergunta. O livro tem regras tão completas que a pessoa que está do lado de fora – e que faz as perguntas e lê as respostas – fica convencida que SEARLE entende chinês. Uma das perguntas pode até ser “você fala chinês?” e a resposta gerada pelas regras que SEARLE usa, seria, neste caso, “sim”. No entanto, SEARLE continua sem ter a menor noção do conteúdo da conversa, pois apenas segue regras. Dessa forma faz-se uma analogia com um programa – que seria o livro de regras – e um computador executando uma tarefa. Mesmo que a tarefa reproduzisse o comportamento de um ser humano, o computador continuaria sem ter noção do que faz.

PENROSE (1997, pp.17-23 e 1996, pp.40-41) levanta as idéias de SEARLE contra a identificação entre estado mental e algoritmo e cita o argumento do “quarto chinês”. PENROSE reconhece que a argumentação não é totalmente conclusiva. Em particular, discorda da alegação de SEARLE de que a superioridade dos cérebros humanos é devida à sua composição biológica, ou seja, de que “objetos biológicos” possam ter intencionalidade e semântica ao passo que objetos eletrônicos não o possam. Mas, mesmo assim, atribui uma “*força considerável*” aos argumentos de SEARLE.

No entanto, a argumentação de PENROSE difere daquelas de LUCAS e de SEARLE no tocante à teoria que PENROSE desenvolve para explicar **mente** e **consciência**. Esta teoria postula que a consciência ocorre através de processos quânticos que ocorreriam em estruturas intra-neuronais chamadas “microtúbulos”[1] (PENROSE, 1997, capítulos 8,9, e 10 e PENROSE, 1996b, capítulos 5, 6, 7 e 8). Assim, pode-se dizer que sua argumentação tem duas linhas básicas:

**1)** aquela que usa os resultados de GÖDEL para defender uma diferença insuperável entre mentes humanas e computadores digitais. Esta linha de pensamento já foi desenvolvida anteriormente por LUCAS, SEARLE e outros;

**2)** aquela que tenta explicar o funcionamento da mente através dos processos quânticos (PENROSE, 1996a).

A teoria quântica surgiu, segundo EINSTEIN (1996), da necessidade de uma explicação teórica para medições nos comportamentos de moléculas. À medida que estas medições se tornavam mais acuradas, a explicação dada pela mecânica clássica gerava paradoxos. Uma explicação coerente do comportamento das moléculas levava a

[1] *microtubules* (PENROSE, 1994 p.358)

assumir que a energia de um sistema mecânico varia em valores discretos, isto é, dando "saltos" de um valor a outro (EINSTEIN, 1996 p.85). Ainda segundo EINSTEIN, não é possível descrever corretamente a transição entre estes valores usando-se as leis causais comuns. Para esta descrição são necessárias leis estatísticas. EINSTEIN (op. cit. p. 86) credita a formulação destas leis estatísticas à HEISENBERG, DIRAC, BROGLIE e SCHRÖDINGER.

PENROSE (1996a) reconhece que a segunda linha de argumentação só faz sentido dada a veracidade da primeira. Isto porque os processos quânticos com os quais ele explica o funcionamento da mente seriam responsáveis justamente pela parte não-computável dos processos mentais. Caso os processos computáveis não tivessem a limitação apontada pelos resultados de GÖDEL e da máquina de TURING, não haveria por que explicar o funcionamento da mente por outro tipo de processo.

Defendendo estas idéias, PENROSE conseguiu um grande número de defensores e discordantes em ambas as linhas de argumentação, gerando, a cada publicação, uma série de comentários e críticas. Este trabalho limita-se a discutir os argumentos relativos à primeira linha de argumentação, ou seja, a linha que parte dos resultados de GÖDEL para defender uma diferença insuperável entre mentes humanas e computadores digitais.

Como PENROSE baseia-se intensivamente em mostrar que a consciência não é reprodutível por um sistema computacional, seria de se esperar que ele apresentasse definições precisas destes termos, mas não é o caso. Ele diz que não acredita que uma tentativa de definição completa fosse necessariamente útil (PENROSE, 1996b, p.37) e, na sua opinião, uma definição prematura seria inadequada (idem, pp.38 e 39). Ele diz que fornecerá "*apenas alguns comentários descritivos relativos ao uso do termo **consciência***" (idem), e que devemos confiar na nossa compreensão intuitiva do seu significado[2] (idem). Ele escreve muitas vezes a expressão "*conscious awareness*" ou simplesmente "*awareness*" no lugar de "*consciousness*". Para o dicionário *Oxford Advanced Learners* as definições de "*awareness*" e de "*consciousness*" são parecidas:

---

[2] I shall give just a few descriptive comments concerning my use of the term 'consciousness'. When all is said and done, we must ultimately rely upon our intuitive comprehension of its meaning.
PENROSE (1996b) p.39

- *awareness*: conhecendo algo; conhecendo que algo existe e é importante; estando interessado em algo;
- *consciousness*: 1- o estado de ser capaz de usar os seus sentidos e recursos mentais para entender o que está acontecendo; 2- o estado de estar atento (*aware*) sobre algo.

O que PENROSE designa com o termo *awareness* é o aspecto passivo do fenômeno da consciência (PENROSE,1996b p.39). Embora sem fornecer definições completas, PENROSE esboça algumas explicações sobre os termos-chave de sua argumentação. Ele coloca que o uso que faz do termo **entendimento** (*understanding*) implica que algum elemento de *awareness* esteja presente. Coloca também que só usa o termo **inteligência** quando supõe que esteja envolvido algum **entendimento**. Para ele inteligência sem entendimento é um termo sem sentido[3] (op. cit. p.38). Assim, para ele o uso do termo **inteligência** implica na presença de "*awareness*", pois:

a) inteligência requer entendimento; e
b) entendimento requer "*awareness*".

Ao delinear estas explicações, PENROSE lembra que o seu argumento tem como centro mostrar que um aspecto do que aqui ele chama de **entendimento** não pode ser simulado por nenhum processo computacional. Ele coloca também que um argumento que estabelece a natureza não-computacional do **entendimento** também estabelece a natureza não-computacional da **inteligência**. E mais ainda, como *awareness* é algo necessário para o entendimento, então uma explicação não-computacional para o fenômeno de *awareness* explica também uma natureza não-computacional para o **entendimento** (PENROSE, 1996b p.38). Para PENROSE esta explicação não-computacional é resultado de processos quânticos que ocorrem em estruturas internas dos neurônios.

Como PENROSE (1996b, pp. 37-39) defende a natureza não-computacional de *awareness* e como, segundo ele, **entendimento** humano implica em *awareness*, podemos ver que sua argumentação é circular. De fato, se ele propõe que **inteligência**

---
[3] 'intelligence' without understandinf is a misnomer
PENROSE (1996b) p.38

implica em **entendimento** e que **entendimento** implica em *awareness*, então, por transitividade, **inteligência** implica em *awareness*. Se, também, segundo PENROSE, somente humanos são dotados de *awareness*, então a máquina foi excluída já no início da argumentação, pelas próprias proposições iniciais. Cabe lembrar também que o segundo livro de PENROSE foi escrito em 1994 e que possivelmente PENROSE não avaliava que o campo da ciência que trabalha hoje com a computação quântica teria o desenvolvimento que vem apresentando.

Como já comentado, para PENROSE a consciência tem dois aspectos: um passivo e outro ativo. E não é sempre clara a distinção entre os dois. Ele dá como exemplos do aspecto passivo a percepção de uma cor; a sensação de dor; e a apreciação de uma música. Ele exemplifica o aspecto ativo com o ato de se levantar da cama. Já relembrar-se de um fato antigo envolveria os dois aspectos. A consciência também estaria envolvida na formulação de um plano e certamente em qualquer atividade que requeira **entendimento**. Ainda, ele explica que podemos estar em um grau de consciência passiva, e até mesmo da ativa quando sonhamos. O aspecto passivo tem relação com as sensações e o ativo com as questões de “livre arbítrio” (*free will*) (PENROSE, 1996b p.39-40). Por falta de um termo melhor para traduzir “*awareness*”, ele será traduzido neste trabalho pelo termo consciência.

PENROSE (1997) menciona pela primeira vez o termo “IA forte” (op. cit. p.16), que denota um ponto de vista segundo o qual um computador digital poderia não apenas simular a inteligência humana, mas também estar plenamente consciente – no sentido comentado por PENROSE – de suas ações. A “IA forte” postula que a atividade mental é simplesmente a realização de uma seqüência bem definida de operações lógicas, que PENROSE chama de algoritmo. Segundo ele, os defensores da “IA forte” argumentam que este feito ainda não ocorreu devido ao fato de que o software para obtê-lo é de um nível de complexidade que ainda não foi alcançado pela tecnologia atual.

REDHEAD (2000, p. 915) vê tanta identificação entre as idéias de LUCAS e de PENROSE que chama a argumentação do último de **argumento LUCAS-PENROSE**. Ele resume dizendo que o argumento se apóia essencialmente na capacidade de matemáticos humanos em saberem que a sentença de GÖDEL é verdadeira, ao passo que um computador produzindo seqüências lógicas de axiomas nunca chegariam à esta sentença. Assim, humanos saberiam de algo que um computador não poderia saber.

Em "*Shadows of The Mind*" PENROSE resume em quatro os pontos de vista, alternativos e excludentes, que resumem o debate sobre a questão da I.A.( PENROSE, 1996b p.12):

**A.** Todo pensamento é computação; em particular, a sensação de estar consciente é criada meramente através da execução de computações apropriadas[4];

**B.** A consciência é uma característica de processos quânticos que ocorrem no cérebro. Qualquer ação física (incluindo, portanto, os processos quânticos) pode ser simulada computacionalmente. Mas a simulação computacional não pode, por ela mesma, produzir a consciência[5];

**C.** A consciência emerge de uma determinada ação física produzida no cérebro, mas esta ação física não pode sequer ser simulada completamente de modo computacional[6];

**D.** A consciência não pode ser explicada em termos físicos, computacionais, ou quaisquer outros termos científicos[7].vem apresentando

Ele também sugere que esta listagem em quatro pontos de vista tem o objetivo de sintetizar vários outros, ou seja, pode haver pontos de vista intermediários entre os listados acima. O ponto de vista **A** é relacionado ao termo "IA forte". PENROSE apóia o ponto de vista **C**, apesar de respeitar de certa forma os defensores do ponto de vista **A**. No entanto, em **A** o *hardware* deixa de ter importância, pois tudo com que temos contato seria produzido por um gigantesco sistema formal (o *software*).

[4] All thinking is computation; in particular, feelings of conscious awareness are evoked merely by the carrying out of appropriate computations
[5] Awareness is a feature of the brain's physical action; and whereas any physical action can be simulated computationally, computational simulation cannot by itself evoke awareness
[6] Appropriate physical action of the brain evokes awareness, but this physical action cannot even be properly simulated computationally
[7] Awareness cannot be explained by physical, computational, or any other scientific terms

A questão da interdependência entre *hardware* e *sofware* tem dois aspectos. o primeiro aspecto relaciona-se com os computadores digitais atuais. Fundamentando o *software* destes computadores está a álgebra booleana, que necessita um *hardware* capaz de manipular as seqüências de símbolos binários desta álgebra. Caso não fosse desenvolvido um *hardware* capaz de executar tal manipulação, a álgebra booleana não teria tido esta importância e, portanto, o *software* teria características diferentes. O segundo aspecto considera que, dado que existe *hardware* capaz de fazer operações com os símbolos binários, o *software* fica independente de um *hardware* particular, já que o programa não depende de nenhum computador específico para funcionar ZYTKOW (1998, p.86). Toda vez que o programa for posto a funcionar, ele realizará o mesmo modelo, independente de diferenças entre processadores e circuitos de memória.

Voltando ao ponto de vista **A**, que representa a crença dos que defendem a "IA forte", percebemos que este ponto de vista implica também que a parte física do mundo não é importante e, portanto, tanto cérebros quanto computadores passam a ser irrelevantes, tornando-se assim meros processadores das regras do sistema formal. Isto porque, para este ponto de vista, a sensação de estar consciente é criada meramente através da execução de computações apropriadas.

ROSA (2002), considerando isto, convida-nos a questionar se o ponto de vista **A** não é uma versão moderna do idealismo de BERKELEY (1685-1753). Segundo a *Encyclopedia of Philosophy* (1972), a palavra **idealismo** vem da palavra grega **ιδεα**, que significa a aparência de alguma coisa vista, e só começou a ser utilizada como uma categoria filosófica no século XVIII, embora Platão já a usasse como um termo técnico dentro da sua filosofia para representar um *universal* - por exemplo *doçura* - em contraste com um *particular* - por exemplo *algo doce*, ou para representar um padrão ideal - como por exemplo *beleza absoluta*. Em *Principles of Human Knowledge* BERKELEY (1978) postula o princípio fundamental da sua filosofia. Este diz que existir é ser percebido, "**Esse est percipi**". Uma árvore existe porque estou - ou alguém está - a percebendo, e esta árvore **existe na idéia** desta mente que a percebe. Ou seja, para BERKELEY a única existência é a da idéia, sendo a matéria apenas a extensão desta idéia. No aforismo XVIII do *Principles of Human Knowledge* (op. cit.) BERKELEY pergunta: "*Mas, se fosse possível que substâncias sólidas, com forma e* *movimento pudessem existir sem a mente, correspondendo às idéias que temos dos corpos, como seria possível sabermos?*"[8] (BERKELEY, 1978 § XVIII). Ele nos propõe, no aforismo XX (op. cit.), a fazer uma comparação entre nossas impressões e as de uma inteligência isolada de corpos externos que seja afetada com a mesma seqüência de sensações ou idéias que nós. E nos pergunta: "esse ser inteligente não teria todas as razões para acreditar na existência de substâncias corpóreas, representadas por suas idéias, e trabalhando com elas na sua mente, que nós temos para acreditar na mesma coisa?

Os quatro pontos de vista listados acima envolvem o termo **consciência**, para o qual não temos uma definição satisfatória. As dificuldades em obter definições para temos como **mente** e **consciência** devem-se, segundo PENROSE, ao estado atual da nossa ciência e, para sanar este problema, ele prescreve uma reconsideração sobre o que consideramos "ciência". Ele pondera que, para acomodar a questão da **mente**, precisaremos de uma ciência mais abrangente que a atual. Para ele, os resultados de GÖDEL estão nos forçando a aceitar algo parecido com o ponto de vista **C**, o que nos obriga a nos ajustarmos às suas implicações (PENROSE, 1996b p.50).[9]

ROSA (2002) concorda com PENROSE. Para ele, a questão da inteligência nos conduz inevitavelmente ao problema da mente humana, para a qual falta uma explicação científica. Longe de ser trivial, esta explicação não é consensual nem sequer no âmbito da neurociência, envolvendo aspectos filosóficos polêmicos (ROSA, 2002 cap. VI).

Há também quem defenda que a consciência simplesmente não é explicável cientificamente, pois não pode ser capturada nos modelos científicos e nem reduzida a um sistema computacional. Entre estes, estão ZYTKOW (1998) e HORST (1998). ZYTKOW menciona que os modelos computacionais expandiram o reducionismo ao

---

[8] But, though it were possible that solid, figured, movable substances may exist without the mind, corresponding to the ideas we have of bodies, yet how is it possible for us to know this?

[9] It may well be that in order to accommodadte the mistery of mind, we shall need a broadening of what we presently mean by 'science', but I see no reason to make any clean break with those methods that have served us so extraordinarily well. If, as I believe, the GÖDEL argument is consequently forcing us into acceptance of some form of viewpoint C, then we shall also have to come to terms with some of its other implications.

permitir modelar entidades pertencentes à mente, para as quais não há correspondentes físicos. Por exemplo, um modelo cognitivo pode incluir metas, crenças, planos, esquemas de inferência, etc (ZYTKOW,1998 p.86).

ZYTKOW lembra que modelos científicos exercem influência no que pensamos sobre os aspectos da natureza. Como exemplo, podemos citar o conceito de **força** na física clássica. A maioria das pessoas aceita que "força" é algo que existe na natureza, provavelmente devido à popularização da mecânica newtoniana, na qual este conceito é fundamental. No entanto, por mais útil que o conceito seja dentro do modelo, não há prova de que este nome designe alguma coisa real (ROSA, 2002 cap. III). D'ALEMBERT, por exemplo, considerava este conceito supérfluo. Para ele, tudo o que pode ser diretamente observado é a matéria e seu movimento através do espaço (idem).

Partindo da constatação de que modelos científicos nos levam a acreditar que os correspondentes às partes do modelo existem realmente no mundo físico, ZYTKOW aponta que algo similar acontece em relação aos modelos computacionais de aspectos mentais. Ele nos alerta que podemos ser levados a acreditar que um modelo de consciência funcionalmente adequado é realmente consciente (idem, p.87).

ZYTKOW divide tudo que ocorre em nossa consciência em dois componentes: conteúdos de informação (*information contents*) e conteúdos intrínsecos (*intrinsic contents*).

Os conteúdos de informação podem ser expressos em símbolos. Símbolos primitivos podem ser combinados para formar expressões complexas. Estes conteúdos podem ser compartilhados por humanos e sistemas de computação e podem ser medidos em números de símbolos primitivos. Eles são usados quando nos comunicamos através de uma linguagem para descrever nossas percepções, sentimentos ou crenças. Mas, para ZYTKOW, há um outro componente, que ocorre em todos os estados mentais, que vai além dos conteúdos de informação (ZYTKOW,1998 p.88). Estes últimos são o que ZYTKOW chama de conteúdos intrínsecos.

Os conteúdos intrínsecos não são passíveis de descrição, nem por um modelo científico e nem por um modelo computacional (ZYTKOW, op. cit pp.83,87-92). Para explicar este conceito ZYTKOW usa o termo **qualia**. Este termo também é usado por PENROSE (1996b, p.40 e 52) e por SKOKOWSKI (2002) que o identificam com a expressão "experiência subjetiva", uma qualidade fenomênica, e não física. SKOKOWSKI (op. cit.) escreve todo um artigo, intitulado "I, zombie" sobre a questão do aspecto **qualia**.

ZYTKOW defende que os conteúdos intrínsecos não podem ser capturados por modelos científicos e também não podem ser reduzido a um sistema computacional, pois não podem ser definidos e nem medidos de uma maneira cientificamente aceitável (op. cit. p. 90). Ele também realiza um tipo de demonstração para mostrar que, se fosse possível inserir conteúdos intrínsecos num sistema de computação, estes conteúdos não teriam razão de ser. Isto porque o sistema agiria e reagiria do mesmo modo, ou seja, nós, exteriores ao sistema, não teríamos como saber se ele os possui.

HORST esclarece alguns termos e suposições de ZYTKOW, nos comentários que faz ao artigo deste. Lembra que as idéias expostas por ZYTKOW são reminiscentes de idéias já discutidas por ele mesmo em "*Symbols, Computation and Intentionality: a critique of the Computational Theory of Mind*". E, apesar de algumas discordâncias menores, endossa a opinião de ZYTKOW.

Mesmo não havendo uma explicação científica satisfatória para a mente, parece que há concordância em diferenciar a concepção de **mente** da concepção de **cérebro**. DAMASIO (2002), expõe esta diferença, ao explicar que o corpo e o cérebro de qualquer pessoa são observáveis por terceiros; a mente, porém, é observável somente por seu possuidor. Sendo assim, vários indivíduos confrontados com o mesmo corpo ou cérebro podem fazer as mesmas observações sobre aquele corpo ou cérebro, mas nenhuma observação comparável de terceiros é possível para a mente de alguém. Conforme DAMASIO, o corpo e o cérebro são públicos, expostos, externos, ou seja, entidades objetivas. A mente é uma entidade inequivocamente subjetiva[10] (DAMASIO, 2002 p.4).

Mais um tema polêmico é colocado por PENROSE quando, com base no teorema de GÖDEL, ele coloca uma concepção platônica em contraposição à concepção formalista de Hilbert. Ele chega a dizer que um ponto de vista platônico foi uma das motivações iniciais de GÖDEL. De fato, Martin Gardner, que prefacia PENROSE (1997) deixa claro o platonismo de PENROSE: para este, não só o universo "existe", como a verdade matemática também tem misteriosa independência e intemporalidade próprias. Além disso, a mente, em alguns momentos estabelece contato com esta verdade objetiva (op. cit., prefácio).

Gardner enquadra PENROSE como platônico porque a concepção platônica crê que os objetos matemáticos - números, séries, conjuntos, classes, teoremas, funções, etc. - existem e sempre existiram em um mundo abstrato, ideal e atemporal, no qual todo o conhecimento matemático já está formado, com toda a integridade e coerência entre as suas partes garantida. Nesta concepção, o desenvolvimento da matemática se dá através de **descobertas**, ao contrário de um desenvolvimento feito por **invenção**, que é o ponto de vista de quem tem outras concepções sobre a natureza da matemática, como por exemplo os intuicionistas e os construtivistas. Os opositores da concepção platônica questionam que não há explicação para o modo como se dá o contato entre nossas mentes e este mundo abstrato e imutável, no qual residem os objetos matemáticos.

Uma justificativa para o ponto de vista platônico é dada por CONNES, um matemático platônico que desenvolve um debate com o neurobiólogo CHANGEUX em "*Conversations on mind, matter and mathematics*" (CHANGEUX e CONNES, 1999). CONNES (op. cit. p. 17-18) evoca a coerência que as estruturas matemáticas exigem e que, para ele, independem do nosso sistema de percepção. Se alguém "trabalhar corretamente", não há como não chegar ao resultado correto. Ele usa também a

[10] Anyone's body and brain are observable to third parties; the mind, though, is observable only to its owner. Multiple individuals confronted with the same body or brain can make the same observations of that body or brain, but no comparable direct third-person observation is possible for anyone's mind. The body and its brain are public, exposed, external and unequivocally objective entities. The mind is a private, hidden, internal, unequivocally subjective entity.

sensação de frustração que sentimos quando fazemos um cálculo de duas maneiras diferentes e não chegamos ao um resultado único[11]. No entanto, CONNES não explora ou define o que chama de "trabalhar corretamente". Mas, se acompanharmos o seu diálogo com CHANGEUX, veremos que CONNES relaciona "trabalhar corretamente" a trabalhar segundo o modelo da lógica aristotélica tradicional. Então, percebemos que sua argumentação enfraquece se admitirmos a possibilidade de outro modelo de lógica. Por exemplo, uma lógica que admita um terceiro valor-verdade, além dos valores verdadeiro e falso, contrariando, assim, o princípio do terceiro excluído da lógica aristotélica. Além disso, como explicar as diferenças a que se chega trabalhando com paradigmas científicos diversos, ou, o que é equivalente, com sistemas formais que trabalham com conjuntos de axiomas diferentes? Um bom exemplo para este questionamento pode ser dado se compararmos resultados da geometria euclideana com resultados da geometria não-euclideana.

Dentro da filosofia matemática, o platonismo confunde-se com o realismo, que é a linha de pensamento que defende que os números e os objetos da matemática existem em um plano abstrato independente de nosso mundo, sendo que o trabalho dos matemáticos é de **descobrir** - em contraposição a inventar - estes objetos.

Alguns trechos de PENROSE (1997) deixam clara a sua concepção platônica. Em suas próprias palavras:

- como tantas outras idéias matemáticas, em especial as mais belas e fundamentais, a idéia da computabilidade parece ter uma realidade platônica própria (PENROSE, 1997 p.77);
- Ao mesmo tempo, freqüentemente, parece haver alguma realidade profunda nesses conceitos matemáticos que vão além das deliberações mentais de qualquer matemático particular. É como se o pensamento humano fosse, em vez disso, guiado para uma verdade exterior e eterna – uma verdade que tem realidade e só se revela parcialmente a qualquer de nós (op. cit. p.104);

[11] (...) if you perform a calculation in two different ways, without arriving at the same result, you experience a very real sense of frustration. For me, this is what mathematical reality is all about : there exists, quite inexplicably, a coherence independent of our system of sensory perception, which guarantees that if one works correctly, one will always detect mistakes – a coherence that entirely surpasses the coherence yielded by sensible intuition, the direct intuition of phenomena.

- Parece que esta estrutura simplesmente não pertence à nossa mente, e, sim, tem uma realidade própria. O conjunto de Mandelbrot não é uma invenção da mente humana: foi uma descoberta. Como o monte Everest, o conjunto de Mandelbrot simplesmente existe! (idem pp.104-105);
- Da mesma forma, o sistema de números complexos tem uma realidade profunda e intemporal, que vai além das construções mentais de um matemático em particular (idem, p.105);
- Há alguma coisa de absoluto e "divino" na verdade matemática (idem, p.124);
- É a "independência matemática" do conjunto de Mandelbrot que lhe dá existência platônica (ibid).

Ele mesmo coloca que, para que o leitor possa acompanhá-lo, é necessário o entendimento desta postura (op. cit. p.128).

No entanto, o ponto de vista platônico sempre foi polêmico. Uma das principais críticas à filosofia platônica é a questão da captação das verdades: Como os humanos, vivendo neste mundo imperfeito, com seus corpos e mentes materiais e mortais, podem interagir com o mundo ideal, não material e atemporal dos objetos matemáticos ? Qual é o ponto em comum ? A resposta platônica para esta questão é a intuição. Mas esta intuição não é explicada, nem o modo como adquiri-la ou refiná-la (HERSH, 1997 p.63).

PENROSE (1996b, pp.418-419) responde a esta questão usando o teorema de GÖDEL. Para ele, os resultados obtidos por GÖDEL serviriam para ilustrar a natureza misteriosa das nossas percepções matemáticas. Aqui ele volta a usar a expressão "*conscious awareness*", que podemos interpretar como "um estado de estar alerta", para responder ao problema da interação entre nossas mentes e o mundo ideal dos objetos matemáticos.[12]

[12] the arguments from GÖDEL's theorem serve to illustrate the deeply mysterious nature of our mathematical perceptions . We do not just 'calculate', in order to form these perceptions, but something else is profoundly involved - something that would be impossible without the very conscious awareness that is, alfter all, what the world of perceptions is all about.

## 4.1 – Criticas e contra-argumentações

O segundo livro de PENROSE (1996b), que foi escrito para responder a criticas feitas aos argumentos expostos no primeiro (PENROSE 1997) gera, por sua vez, sua própria série de criticas e comentários. Alguns destes são apresentados pela revista eletrônica PSYCHE que os reúne no "*Symposium on Roger Penrose's* Shadows of the Mind" (http://psyche.cs.monash.edu.au/symposia/penrose/). Entre os comentários, podemos destacar os de FEFERMAN (1995), CHALMERS (1995) e McCULLOUGH (1995). Este simpósio termina com a resposta de PENROSE aos vários participantes (PENROSE, 1996).

FEFERMAN considera irretocável a apresentação do teorema de GÖDEL feita por PENROSE, mas aponta alguns erros pontuais na discussão posterior que este faz. Entre estes o fato que PENROSE menciona que a sentença que diz que o sistema é $\omega$-consistente não é derivável no sistema. Aqui FEFERMAN nos lembra o que vimos no cap. II deste trabalho. Na demonstração da proposição XI, GÖDEL não precisa da restrição mais rigorosa da $\omega$-consistência. Ele usa apenas o requisito – mais fraco – da consistência simples.

Apresentados os erros, FEFERMAN resume dizendo que se PENROSE usa como base os resultados de GÖDEL, deveria tê-los explorado com mais correção. Mas, os erros em si não tornam a argumentação inválida (FEFERMAN, 1995). FEFERMAN também critica a forte posição platônica assumida por PENROSE e as alegações que este faz sobre a posição idêntica que GÖDEL também teria. Para FEFERMAN, o platonismo de GÖDEL é mais relativo do que PENROSE apregoa. Ao responder a FEFERMAN, PENROSE nota que para seguir sua argumentação não é necessário ter um ponto de vista platônico "forte", relativizando, de certa forma, a posição mais radical assumida em PENROSE(1996b), (1997) e (1998). No entanto, em (1996a) diz que "*é difícil pensar em discutir conceitos abstratos de outro modo*" pois as provas matemáticas lidam com abstrações – idéias que podem ser transmitidas e que não são específicas de nenhum indivíduo.

Para mostrar que as limitações não são inerentes ao computador mas acontecem também com a mente humana, McCULLOUGH (1995) propõe um experimento de pensamento. Neste experimento supomos que uma pessoa tem à sua frente dois botões, um impresso com a palavra NÃO e outro com a palavra SIM. Pede-se a esta pessoa que pressione um dos botões em resposta a uma série de perguntas. O que acontece se perguntarmos a esta pessoa "Você vai pressionar o botão 'NÃO' ?" Vamos agrupar as possibilidades em dois grupos:

a) ela tem a intenção de pressioná-lo.
   Se pressionar 'SIM' a resposta estará errada, pois pressionou o botão 'SIM'. Se pressionar 'NÃO', a resposta também está errada, pois ela tem a intenção de pressioná-lo;
b) ela **não** tem a intenção de pressioná-lo.
   Se pressionar 'NÃO' a resposta estará errada, pois acabou pressionando-o. Se pressionar 'SIM' a resposta também está errada, pois ela não tem a intenção de pressioná-lo.

McCULLOUGH conclui que há um aspecto de incompleteza nas respostas da pessoa porque há perguntas que não podem ser respondidas corretamente. A pessoa sabe o que quer, qual seria a resposta, mas não tem como expressar esta resposta corretamente. Assim, existiria uma lacuna entre o conhecimento interno da pessoa e o que ela pode comunicar externamente. E esta característica não é uma deficiência de inteligência, pois de nada adianta "raciocinar melhor" para conseguir uma resposta correta.

Analisando o argumento de McCULLOUGH vemos que só há incompleteza nas respostas da pessoa porque esta foi colocada no esquema rígido das regras do jogo proposto. Ou seja, a incompleteza está no esquema e não na pessoa, pois esta poderia responder de forma distinta, por exemplo, apertando os dois botões simultaneamente, fazendo sinais com as mãos, ou qualquer outra resposta alternativa.

CHALMERS (1995) resume o argumento no cap. 3 de PENROSE (1996b) usando uma demonstração por contradição. Esta demonstração usa o conceito de sistema *sound*, que traduziremos aqui por "confiável". Um sistema "confiável" é aquele no qual todas as fórmulas demonstráveis são verdadeiras. Ou, como PENROSE (1996b, p. 73) coloca, o sistema é "confiável" se ele não gera respostas erradas.

**1)** Supomos que o pensamento de CHALMERS é reproduzido por um sistema formal F. Ou, mais resumidamente, CHALMERS **é** F. Entre as proposições que ele sabe serem verdadeiras está a premissa de que ele é F;

**2)** Dado que CHALMERS sabe que é F, ele sabe que F é "confiável" (pois ele sabe que ele mesmo é "confiável"). Além disso, se F' é o sistema composto por F mais a proposição "eu sou F", ele sabe que o sistema ampliado F' é "confiável". CHALMERS justifica este passo ao lembrar que acrescentando uma proposição verdadeira a um sistema "confiável", chegamos a outro sistema "confiável";

**3)** Assim, ele sabe que G(F'), isto é, a sentença de GÖDEL do sistema F', é verdadeira. Cabe relembrar que G(F') diz que não é demonstrável em F', o que é verdade apesar de não poder ser reconhecida por F'.

**4)** Mas o sistema F' não pode perceber que G(F') é verdadeira, pelo teorema de GÖDEL. Lembramos que a sentença de GÖDEL, isto é G(F') é a expressão que GÖDEL usou para demonstrar seu teorema. Ou seja, GÖDEL provou que, dentro de F', G(F') é indecidível.

**5)** pelo que supusemos, todavia, CHALMERS é efetivamente equivalente a F', pois ele é F mais o conhecimento que ele é F;

**6)** chegando a esta contradição, mostra-se que a suposição inicial é falsa e, portanto, F não reproduz o pensamento de CHALMERS;

**7)** e, generalizando: seu pensamento não pode ser reproduzido por nenhum sistema formal.

Para PENROSE (1996a), o que CHALMERS expõe acima é o argumento mais importante de PENROSE (1996b) e, por isto mesmo, também o item mais importante do simpósio promovido na revista eletrônica PSYCHE sobre PENROSE (1996b). PENROSE surpreende-se com o fato de que, entre todos os debatedores participantes do simpósio, somente CHALMERS tenha chamado atenção a este argumento contra o modelo computacional de "entendimento".

## 4.2 – A argumentação de REDHEAD

REDHEAD (2000) sintetiza toda a argumentação pró e contra a possibilidade de uma IA “forte” ao apresentar um argumento que usa um **raciocínio indutivo** para considerar que a aritmética é consistente. Mais especificamente, REDHEAD refere-se ao sistema formal que tem como fórmula iniciais os axiomas de PEANO (1858-1932) para a aritmética. Os princípios propostos por PEANO foram originalmente escritos em latim em 1889 (PEANO, 1977 p. 83) e incluem:

- quatro noções aritméticas iniciais: “número”, “um”, “sucessor” e “é igual a”; e
- nove axiomas relativos a estas noções.

No cap. I desta tese vimos que a indução, representada na parte ascendente do arco do saber Aristotélico, não é totalmente fundamentada pela lógica. No entanto, REDHEAD (2000, p.916) usa o fato de que temos bastante evidência indutiva sobre a consistência da aritmética para inferir **indutivamente**, que ela é consistente. Esta evidência está baseada no fato de que já se passaram mais de cem anos desde que PEANO propôs seus axiomas e até hoje não surgiu nenhuma contradição. REDHEAD propõe usar este fato como uma validação indutiva da consistência da aritmética. DELONG (1971) concorda com esta avaliação e acrescenta mais razões, que ele reconhece serem empíricas, para esta validação (DELONG,1971 p.216):

- apesar de muitas tentativas em se achar inconsistências na aritmética durante um grande período de tempo, nenhuma até agora foi encontrada;
- a aritmética tem sido usada e investigada numa ampla gama de problemas e nenhuma inconsistência surgiu até agora;
- mesmo que se venha a encontrar inconsistência, provavelmente a aritmética poderá ser especificada de forma a manter-se útil como é mas sem o que causou a inconsistência. Neste caso, DELONG faz uma comparação com o que aconteceu com o cálculo diferencial e integral, cuja (falta de) fundamentação foi criticada por cerca de 3 séculos. Esta crítica teve razão de ser pois, enquanto os resultados resultavam corretos, havia incoerências nas demonstrações. No entanto, o cálculo agora está livre de falhas e com a

mesma funcionalidade, graças à nova fundamentação concebida por Bolzano (1781-1848), Cauchy (1789-1857) e Weierstrass (1815-1897).

Naturalmente, estas considerações não tornam a inferência indutiva inatacável. A questão, mencionada por REDHEAD, sobre o fato de terem se passado mais de cem anos desde que PEANO propôs os axiomas que levam seu nome, também merece comentários. Isso porque o período de um século é relativamente curto para validar uma teoria. Por quantos anos permaneceu válida a teoria que dizia que a terra era plana? Outro exemplo pode ser dado se considerarmos a mecânica newtoniana. NEWTON (1642-1727) publicou seu "*Principia*" – obra que fundou a mecânica newtoniana – em 1686 (NEWTON, 1974 p. xi). Portanto, mais de dois séculos foram necessários para o aparecimento das teorias quântica e da relatividade (ROSA, 2002). Estas teorias tornaram claras as limitações da mecânica newtoniana. A teoria da relatividade de Einstein quebrou a validade da mecânica newtoniana para velocidades muito altas, próximas à da luz (ROSA, 2002), e a mecânica quântica mostra que o determinismo inerente à mecânica newtoniana não pode descrever o que ocorre com o comportamento de partículas atômicas (EINSTEIN, 1996 p.85-86). De qualquer forma, o que REDHEAD pretende não é fundamentar indutivamente a aritmética. O que ele pretende é mostrar que chegaríamos a uma inconsistência caso tivéssemos um programa de inteligência artificial capaz de aceitar verdades via indução. É o que será mostrado a seguir.

REDHEAD (op. cit.) também comenta a argumentação de LUCAS discutida no capítulo anterior deste trabalho, no sentido de que pode-se adicionar a sentença de GÖDEL (**G**) ao sistema ***P***. Neste caso, teríamos um novo sistema estendido, ***P'***, no qual **G** seria decidível, sendo um dos axiomas de ***P'***. Todavia, o teorema de GÖDEL poderia ser aplicado à ***P'*** e este teria sua própria sentença indecidível **G'**. Esta, claro, poderia ser parte de outro sistema, mais estendido que ***P'***, digamos, ***P''***, mas ***P''*** também teria sua **G''** e assim *ad infinitum*. Como ele resume: em qualquer estágio uma sentença de GÖDEL poderia sempre ser incorporada ao sistema dedutivo, e não existiria sistema dedutivo imune. Assim, ele coloca duas proposições (REDHEAD, op. cit. p.915):

- uma "fraca" (*weaker claim*): para todas as sentenças de GÖDEL existe um computador capaz de produzi-la;

- e outra “forte” (*stronger claim*): existe um computador que pode produzir todas as sentenças de GÖDEL.

Ele coloca que a proposição “fraca” é usada por alguns para justificar que o pensamento aritmético pode ser totalmente programado. No entanto, o que ele chama de argumento “LUCAS-PENROSE” não trata da proposição “fraca”, mas sim da proposição “forte”. O centro do argumento “LUCAS-PENROSE” é a negação da proposição “forte”. Mesmo porque o “computador que pode produzir todas as sentenças de GÖDEL” não se enquadra na premissa de “finitude” do programa formalista.

Para sua argumentação, REDHEAD (op. cit. p. 915) relembra que sabemos que a sentença de GÖDEL (**G**) é verdadeira porque estamos no nível da metalinguagem e que, neste nível, usamos considerações semânticas para reconhecer esta verdade. Parte também do princípio de que um computador não pode reconhecer o significado de **G**, por ser um mecanismo apenas sintático. Neste ponto ele propõe uma tentativa para libertar o computador da limitação revelada pelo teorema de GÖDEL usando uma maneira apenas sintática de se reconhecer a veracidade de **G**, ou seja de o computador provar **G**. Assim o computador poderia ultrapassar a limitação ditada pelo resultado de GÖDEL. Esta maneira sintática lançará mão do recurso da inferência indutiva. REDHEAD contra-argumentará, mais adiante, para mostrar que mesmo esta abordagem está destinada a falhar.

Para tanto, REDHEAD usa um recurso que o próprio GÖDEL demonstra como sua proposição XI: o fato de que se o sistema for consistente, então a fórmula que diz que ele é consistente não pode ser demonstrada nele mesmo (GÖDEL, 1992 p.70). O capítulo II deste trabalho apresenta um esboço da proposição XI de GÖDEL.

REDHEAD escreve a sentença na forma $\mathbf{CON} \Rightarrow \mathbf{G}$ , onde:

**CON** é a sentença que assegura que ***P*** é consistente e **G** é a sentença de GÖDEL.

A sentença $\mathbf{CON} \Rightarrow \mathbf{G}$, como visto, é demonstrável no sistema ***P*** (GÖDEL, idem).

REDHEAD prossegue dizendo que, se o computador pudesse provar **CON** através do processo indutivo, ele poderia usar o condicional **CON ⇒ G** para obter uma prova de **G** por *modus ponens*. O *modus ponens* é uma regra de inferência que permite inferir **B** a partir de **A** e de **A ⇒ B** e é uma das duas únicas regras utilizadas por GÖDEL na sua demonstração (ver cap. II deste trabalho).

REDHEAD (op. cit. p. 916) propõe, portanto, usar o recurso da inferência indutiva que temos de que a aritmética é consistente como recurso para tornar o computador capaz de validar **CON**. Sendo assim, tendo **CON** verdadeiro e também o condicional **CON ⇒ G** verdadeiro, por *modus ponens*, o computador teria uma maneira, a principio somente sintática, de verificar **G**. Quer dizer, um computador programado com algumas regras de lógica indutiva poderia provar **G**. Está fora do escopo deste trabalho discutir sobre lógica indutiva, mas podemos ver algumas reflexões o tema em ALISEDA (2004) e NORTON (2003). Uma discussão mais filosófica sobre o problema da indução é apresentado por CARGILE (1998).

No entanto, o próprio REDHEAD contesta o caráter absolutamente sintático deste raciocínio para contra-argumentar. Ele começa acrescentando que poderíamos adicionar a negação de **G**, ou seja, **~G** aos axiomas de PEANO. A este sistema corresponde uma interpretação não-padrão de aritmética, ainda conforme REDHEAD. Neste sistema, é claro que **G** é falsa – pois **~G** foi colocada como um axioma. No entanto, **CON ⇒ G** ainda é derivável neste sistema. E, desde que se aplicássemos, como antes, o argumento indutivo, chegaríamos a **G** por *modus ponens*. Assim, teríamos que o sistema é inconsistente, pois deriva **~G** , por ser um axioma e, ao mesmo tempo, **G**.

Por que isto acontece? REDHEAD responde que isto ocorre porque a proposição **CON**, de fato, não pode ser apenas sintática. De fato, ela só expressa a consistência da interpretação padrão (*standard*) da aritmética de PEANO, ou seja, da interpretação que tínhamos em mente quando criamos o sistema formal (*intended interpretation*).

CARVALHO e OLIVEIRA (1998) ressaltam a intencionalidade na criação de um sistema formal quando relatam que tais sistemas não ocorrem de modo arbitrário. Estes sistemas são criados com o objetivo de retratar ou reproduzir mecanicamente algum processo ou alguma estrutura que queremos estudar. Para tanto, os sistemas formais devem capturar nosso entendimento do aspecto da realidade que queremos analisar (CARVALHO e OLIVEIRA, 1998 pp. 215-216).

Assim, REDHEAD coloca que, para provar a veracidade de **G** usando **CON** $\Rightarrow$ **G** o argumento original não foi puramente sintático, e nem poderia ser, pois tinha subjacente nossa intenção de interpretá-lo como a aritmética de PEANO. E, como computadores não têm intenções, REDHEAD conclui dizendo que, afinal, verificar que **G** é verdadeira é uma tarefa que transcende o computador.

# CONCLUSÃO

Este trabalho faz reflexões sobre a relação entre as limitações dos sistemas formais e a possibilidade de uma inteligência artificial "plena". Ou, para usar o termo cunhado por LUCAS e PENROSE, "forte". A questão de se os computadores digitais serão ou não capazes de reproduzir uma inteligência similar à inteligência humana tornou-se um ponto polêmico. Para ilustrar tal polêmica e refletir sobre as colocações dos principais autores na área, foram escolhidas algumas argumentações sobre as quais tecemos considerações. Este trabalho também acrescenta mais um argumento sobre as dificuldades de se reproduzir em um sistema formal a inteligência e a criatividade humana. Exemplifica-se as idéias usando o desenvolvimento das geometrias não-euclideanas e as correspondentes alterações nos axiomas de sistemas formais correspondentes. O processo criativo que culminou no surgimento das geometrias não-euclideanas exemplifica as questões relativas à substituição de axiomas num sistema dedutivo.

Segundo DORIA (1987, p.75-76) e PENROSE (1996, p.20-21), as limitações seriam as mesmas em se tratando de computadores digitais, pois a tese de CHURCH nos garante que os sistemas formais do tipo utilizado por GÖDEL são equivalentes a máquinas de TURING.

As questões da incompleteza e da prova de consistência, que permeiam os argumentos contra a possibilidade de uma "IA forte", relacionam-se às metas do programa formalista. Este foi criado com a finalidade de resolver problemas relativos à fundamentação da matemática e de eliminar os paradoxos relativos à teoria dos conjuntos e às questões relacionadas ao infinito. Para este fim, o programa formalista cria a metamatemática. Esta é uma nova área de estudo no qual os sistemas formais passam a ser as "máquinas" cujas regras de produção geram novas fórmulas a partir de fórmulas iniciais, e onde a semântica é eliminada, restando apenas a sintaxe. Outra meta do programa formalista era a de mostrar a consistência da matemática. Como vimos no capitulo II, as provas de GÖDEL quanto à incompleteza e quanto à impossibilidade de se provar a consistência de um sistema formal acabaram por frustrar as metas principais deste programa.

Os resultados de GÖDEL acabaram servindo como argumento nas discussões sobre a possibilidade de uma inteligência artificial similar à inteligência humana. No capítulo III mostrou-se como um sistema formal, por mais enriquecido que possa ser, continua sujeito à limitação da incompleteza e à necessidade de um avaliador externo para julgar sua consistência.

O capítulo IV mostrou as propostas de PENROSE e de contra-argumentações que suas idéias geraram. Vimos neste capítulo que PENROSE trabalha com duas linhas de argumentação: a primeira, que tem como base o teorema de GÖDEL para justificar que a inteligência humana tem um aspecto inalcançável por qualquer computador digital; e a segunda, que tenta explicar que a parte não-algoritmica de nossos processos mentais dá-se graças a fenômenos quânticos que ocorrem em estruturas celulares internas aos neurônios. Já que PENROSE (1996a) coloca que a segunda linha de argumentação só faz sentido dada a veracidade da primeira, pode-se pensar numa pesquisa que verifique esta hipótese, ou seja, os processos quânticos realmente realizam ações não-algoritmicas? caso a resposta seja “sim”, seriam estes processos o único modo de se realizar ações não-algoritmicas?

Vimos que PENROSE patina quando argumenta sobre termos como inteligência, entendimento (*understanding*) e consciência (*consciousness* e *awareness*), pois ele mesmo não fornece definições claras para estes termos. Sabendo que há várias interpretações para estes termos, como podemos avançar numa proposta de aprofundamento científico nestas questões, sem definições para os seus termos-chave? Como a discussão poderá prosseguir com cada participante tendo uma idéia do que representa cada termo? FRANCIS BACON certamente ficaria surpreso em saber que quase quatro séculos após a publicação do seu “*Novum Organum*”[1] (BACON, 1999) ainda erramos ao não dar a devida atenção à eliminação das ambivalências que as palavras carregam. BACON postulou que, para o desenvolvimento do conhecimento seria necessário livrar nossas mentes de quatro mitos básicos, entre eles o que BACON

[1] BACON publicou o “*Novum Organun*”, em latim, em 1620

chamou de “ídolos do mercado”, que consistem nos erros provenientes da ambigüidade das palavras. Uma mesma palavra tem sentidos diferentes para os interlocutores. Para BACON, a interdependência entre palavras e razão pode resultar numa concordância aparente, numa falsa controvérsia ou em ficções. Os homens acreditam que dominam as palavras, mas estas adquirem força e esta força acaba por influenciar o intelecto. Este processo tornaria as ciências sofísticas e inativas (BACON, 1999 p. 46). Ainda para BACON (op. cit. pp. 46-47) com freqüência acontece que grandes disputas entre intelectuais acabam em controvérsias em torno de palavras e nomes, “*caso em que melhor seria restaurar a ordem, começando pelas definições*” (idem).

A falta de definição dos termos de certa maneira lembra a situação que ocorreu no século XIX com os paradoxos da lógica., quando FREGE, RUSSELL, HILBERT e outros foram forçados a reconhecer as limitações da linguagem natural, devido à sua ambivalência. Talvez a dificuldade em explicitar sobre o que estamos realmente discutindo quando falamos em “IA forte” mostre que PENROSE tem razão em pelo menos um aspecto: quanto este aponta que é necessário algum novo ingrediente na ciência atual para que ela possa explicar o que são e como se dão coisas como “pensamento” e “consciência”.

Ao considerar todas estas dificuldades constata-se que uma pergunta feita na introdução deste trabalho não pode ser respondida. Tendo como ponto de partida uma visão analítica e racionalista, para copiar algo seria preciso primeiramente conhecer este algo. Como portanto copiar completamente a mente humana se, pelo menos no momento, não conhecemos completamente sua abrangência e potencialidades?

Considerando o problema exposto nos parágrafos acima, pode-se fazer outra sugestão de pesquisa, que seria o projeto – ambicioso – de definir propriamente os termos **mente**, **inteligência** e **consciência**.

As argumentações de LUCAS, REDHEAD e a que CHALMERS sintetiza de PENROSE (1996b) são bastante convincentes de que humanos conseguem executar algo “além do algoritmo”, embora esta expressão continue necessitando definição precisa. Para ilustrar isto, o experimento proposto por McCULLOUGH (cap. IV deste trabalho) é ilustrativo. Como ele corretamente expõe, qualquer dos dois botões que a pessoa venha a pressionar provoca uma resposta errada. No entanto, mesmo fora das regras impostas, nada nos impediria de pressionar os dois botões ao mesmo tempo, ou de verbalizar a resposta. Ou fazer qualquer outra coisa que transmitisse a mensagem de que estamos numa situação "de não poder responder". E é difícil imaginar que uma máquina, na mesma situação, não teria a iniciativa, ou a intenção para fazer o mesmo.

Talvez o próprio termo **metamatemática**, criado por HILBERT para denominar o estudo das proposições de um sistema dedutivo seja esclarecedor o bastante para percebermos a necessidade de "avaliadores externos" ao sistema, que são os matemáticos, ou seja, humanos. Ou seja, talvez o "além de" designado pelo prefixo *meta* denunciasse, mesmo antes da prova de GÖDEL, as limitações dos sistemas formais.

Considerando as argumentações dos vários autores apresentados, vemos que todas são suscetíveis de contra-argumentações. Colocando-nos como espectadores destes debates podemos ter uma idéia do desenvolvimento da ciência atualmente. Este desenvolvimento se dá nas mais variadas direções e chega-se a conclusões às vezes contraditórias, gerando debates intermináveis. Estes fatos nos levam a questionar se não estaria a ciência, e não apenas os sistemas formais, sujeita a limitações. Embora não tenhamos uma prova para este fato, somos levados a aceitar que, pelo menos em algum grau, teremos que conviver com a incerteza. O texto de Clarice Lispector sintetiza:

> *Não entendo.*
> *Isso é tão vasto que ultrapassa qualquer entender. Entender é sempre limitado. Mas não entender pode não ter fronteiras. Sinto que sou muito mais completa quando não entendo. Não entender, do modo como falo, é um dom. Não entender, mas não como um simples de espírito. O bom é ser inteligente e não entender. É uma benção estranha, como ter loucura sem ser doida. É um desinteresse manso, é uma doçura de burrice. Só que de vez em quando vem a inquietação: quero entender um pouco.*
> *Não demais: mas pelo menos entender que não entendo.*

# BIBLIOGRAFIA


ALISEDA, Atocha, 2004, **Logics in scientific discovery**. Foundations of Science **9** pp. 339-363.

ARISTÓTELES (384-322 A.C.), 2001 **Da geração e da corrupção**. Tradução de Renata M. P. Cordeiro. Apresentação de Ana M. A. Goldfarb. São Paulo: Landy.

BACON, Francis (1561-1626), 1999, **Novum organum,** Coleção Os Pensadores. São Paulo: Editora Nova Cultural Ltda.

BERKELEY, George (1685-1753), 1978, **The Principles of Human Knowledge**. publicado por Hutchins, R. M.; Great Books of the Western World n. 35, William Benton Publisher: London.

BOOLE, Geoge (1815-1864), 1958, **An investigation of the laws of thought**. New York: Dover.

BOOLOS, George e JEFFREY, Richard, 1974, **Computability and logic**. New York: Cambridge University Press.

BOYER, Carl B., 1974, **História da matemática**. Tradução: Elza F. Gomide. São Paulo: Edgard Blücher.

CALUDE, Cristian S., 2002, **Chaitin Ω numbers, Solovay machines, and Gödel incompleteness**. Theoretical Computer Science 284 pp. 269-277.

CARGILE, James., 1998, **The problem of induction**. Philosophy **73** pp. 247-275.

CARVALHO, Roberto L. de e OLIVEIRA, Claudia M. G. M., 1998, **Modelos de computação e sistemas formais**. Rio de Janeiro: DCC/IM, COPPE/Sistemas, NCE/UFRJ.

CHAITIN, Gregory J., 1990, **A Random Walk in Arithmetic**. In: New Scientist 125, No. 1709 (24 March), pp. 44-46.

____________, 1993, **Randomness in Arithmetic and the Decline & Fall of Reductionism in Pure Mathematics**. EATCS Bulletin, No. 50 (June 1993), pp. 314-328.

____________, 2002. **Computers, paradoxes and the foundations of mathematics**. American Scientist, vol. 90 pp. 164-171.

____________, 2004, **Leibniz, randomness & the halting probability**. Mathematics Today, V. 40 Nº 4 (August), pp. 138-139.

CHALMERS, David J., 1995, **Minds, machines and mathematics**. PSYCHE – An interdisciplinary journal of research and consciousness 2(9), June 1995. http://psyche.cs.monash.edu.au/v2/psyche-2-09-chalmers.html.

CLAPHAM, Christopher, 1996, **Concise dictionary of mathematics**. Oxford: Oxford University Press.

CLELAND, Carol E., 2004, **The concept of computability**. Theoretical Computer Science **317** pp. 209-225.

DAGHLIAN, Jacob, 1995, **Lógica e álgebra de Boole**. 4. ed. São Paulo: Atlas.

DAMASIO, Antonio R., 2002, **How the brain creates the mind**. Scientific American Special. Volume 12, number 1 pp. 4-9.

DELONG, Howard, 1971, **A profile of mathematical logic**. London: Addison-Wesley Publibishing Company.

DORIA, Francisco A. et alli., 1987, **A máquina e seu avesso**. Rio de Janeiro: Francisco Alves.

EINSTEIN, Albert (1879-1955), 1996, **Out of my later years**. New York: Wings Books.

EISENSTADT, Stuart A. e SIMON, Herbert A., 1997, **Logic and thought**. Minds and Machines **7**, pp. 365-385.

ERNEST, Paul, 1998, **Social constructivism as a philosophy of mathematics**. New York: State University of New York Press.

EUCLID (c. 300 A.C.), 1958, **The Thirteen books of the Elements**, 3v. 2 ed. New York: Dover.

FEFERMAN, Solomon, 1995, **Penrose's gödelian argument**. PSYCHE – An interdisciplinary journal of research and consciousness 2(7), May 1995. http://psyche.cs.monash.edu.au/v2/psyche-2-07-feferman.html.

FERREIRA, Aurélio Buarque de H., 1997, **Novo dicionário da língua portuguesa**. 2.ed. Rio de Janeiro: Nova Fronteira.

FREGE, Gottlob (1848-1925), 1970, "Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought". In: **Frege and Gödel. – Two fundamental texts in mathematical logic**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 1-82,

____________, 1977, "Letter to Russell". **In: From Frege to Gödel - A source book in mathematical logic, 1879-1931**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 126-128.

GÖDEL, Kurt (1906-1978), 1970, "On formally undecidable propositions of Principia Mathematica and related systems 1". In: **Frege and Gödel. – Two fundamental texts in mathematical logic**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 83-108.

____________, 1979, "Acerca de proposições indecidíveis de sistemas matemáticos formais". In: **O teorema de Gödel e a hipótese do contínuo**. Ed. e tradução de Manuel Lourenço. Lisboa: Fundação Calouste Gulbenkian, pp. 291-358.

____________, 1992, **On formally undecidable propositions of Principia Mathematica and related systems 1**. Introduction by R. B. BRAITHWAITE. New York: Dover.

____________, 2001a, "On formally undecidable propositions of Principia Mathematica and related systems 1". In: **Kurt Gödel collected works**. Vol 1. Edited by Solomon Feferman et al. New York, Oxford University Press, pp. 145-195.

____________, 2001b, "On completeness and consistency". In: **Kurt Gödel collected works**. Vol 1. Edited by Solomon Feferman et al. New York, Oxford University Press, pp. 235-237.

____________, 2001c, “The formalist’s way of founding mathematics - review”. In: **Kurt Gödel collected works**. Vol 1. Edited by Solomon Feferman et al. New York, Oxford University Press, pp. 248-249.

GRATTAN-GUINESS, I., 1979, **In memoriam Kurt Gödel: his 1931 correspondence with Zermelo on his incompletability theorem**. Historia Mathematica **6**, pp. 294-304.

GREENBERG, Marvin J., 1994, **Euclidean and non-Euclidean geometries: development and history**. 3rd ed. New York: H. Freeman and Company.

GREY, Jeremy, 1979, **Non-euclidean geometry – a re-interpretation**. Historia Mathematica 6, pp. 236-258.

GUILLEN, Michael, 1987, **Pontes para o infinito – o lado humano das matemáticas**. Lisboa: Gradiva.

HATCHER, William S., 1968, **Foundations of mathematics**. London: W. B. Saunders Co.

HEIJENOORT, J. V., 1977, **From Frege to Gödel - A source book in mathematical logic, 1879-1931**. Cambridge: Harvard University Press.

HERSH, Reuben, 1997, **What is mathematics, really** . London: Random House.

HEYLIGHEN, Francis, 1999, **Advantages and limitations of formal expression**. Foundations of Science **4** pp. 25–56.

HILBERT, D. e ACKERMANN, W., 1950. **Principles of mathematical logic**. New York: Chelsea.

HILBERT, David (1862-1943), 1977a, “On the infinite”. **In: From Frege to Gödel - A source book in mathematical logic, 1879-1931**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 367-392.

____________, 1977b, “The foundations of mathematics”. In: **From Frege to Gödel - A source book in mathematical logic, 1879-1931**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 464-489.

HOFSTADTER, D. R., 1980, **Gödel, Escher, Bach: an eternal golden braid**. London: Penguin Books.

HORST, Steve, 1998, **Comments on Zytkow's article**. Foundations of Science **1** pp. 103-109.

HUME, David (1711-1776), 1978, **On human nature and the understanding**. New York: Macmillan Publishing.

KANT, Immanuel (1724-1804), 1997, **Crítica da razão pura**. Lisboa: Fundação Calouste Gulbenkian.

KNEALE, William e KNEALE, Martha, 1971, **The development of logic**. Oxford: Oxford University Press.

KNEEBONE, G. T., 2001, **Mathematical logic and the foundations of mathematics**. New York: Dover Publications Inc.

KRAJEWSKI, Stanislaw, 2004, **Gödel on Tarski**. Annals of Pure and Applied Logic 127, pp. 303– 323.

LEIBNIZ, Gottfried W (1646-1716), 1951, "Towards a universal characteristic". In: **G. B. Leibniz, Selections**. Edited by Philip P. Wiener. New York: Scribner's Sons, pp. 17-25.

LEWIS, Harry R. e PAPADIMITRIOU, Christos H., 2000, **Elementos de teoria da computação**. Tradução de E. Furmankiewicz. Porto Alegre: Bookman.

LISPECTOR, Clarice (1920-1977), 2004, **Aprendendo a Viver**. Rio de Janeiro: Rocco.

LOECKX, Jacques, 1972, **Computability and decidability**. Berlin: Springer-Verlag.

LOSEE, John, 1977, **A historical introduction to the philosophy of science**. Oxford: Oxford University Press.

LUCAS, J. R., 1964, "Minds, machines and Gödel". In: **Minds and machines**, edited by Alan Ross Anderson. New Jersey: Prentice-Hall, pp. 43-59.

____________, 2005, **A paper read to the Turing Conference at Brighton on April 6th, 1990**. http://users.ox.ac.uk/~jrlucas/Godel/brighton.html - acesso em 10/02/2005.

MACHADO, N. J., 1991, **Matemática e Língua Materna**. São Paulo: Cortez.

MARTINS, Roberto de A., 1995, **A influência das geometrias não-euclidianas no pensamento físico so século XIX**. Revista da Sociedade Brasileira de História da Ciência, n. 13, pp. 67-80. Rio de Janeiro.

McCULLOUGH, Daryl, 1995, **Can humans escape Gödel?** PSYCHE – An interdisciplinary journal of research and consciousness 2(4), April 1995. http://psyche.cs.monash.edu.au/v2/psyche-2-04-mccullough.html.

NAGEL, Ernest and NEWMAN, James R., 1958, **Gödel's Proof**. New York: New York University Press.

NEWTON, Isaac (1642-1727), 1974, **Mathematical principles of natural philosophy and his system of the world**. 8. ed. Berkeley: University of California Press.

NORTON, John, 2003, **A material theory of induction**. Philosophy of Science, 70, pp. 647-670.

Oxford Advanced Learner's Dictionary, 2003, 6.ed. Ed. Sally Wehmeier. New York: Oxford University Press.

PEANO, Giuseppe (1858 - 1932), 1977, "The Principles of arithmetic, presented by a new method". In: **From Frege to Gödel - A source book in mathematical logic, 1879-1931**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 83-97.

PECKAUS, Volker, 2002, **Hilbert's Paradox**. Historia Mathematica 29, pp. 157–175.

PENROSE, Roger, 1996a, **Beyond the doubting of a shadow**. PSYCHE – An interdisciplinary journal of research and consciousness 2(23), January 1996. http://psyche.cs.monash.edu.au/v2/psyche-2-23-penrose.html.

____________, 1996b, **Shadows of the mind**. **A search for the missing science of consciousness**. Oxford: Oxford University Press.

____________, 1997, **A mente nova do rei: computadores, mentes e as leis da física**. Tradução de Waltensir Dutra. Rio de Janeiro: Campus.

PENROSE, Roger et. al, 1998, **O grande, o pequeno e a mente humana**. Org. de Malcolm Longair; tradução Roberto Leal Ferreira. São Paulo: Fundação Editora UNESP.

PUTNAM, Hilary, 1964, “Minds and machines”. In: **Minds and machines**, edited by Alan Ross Anderson. New Jersey: Prentice-Hall, pp. 43-59.

RADU, Mircea, 2003, **A debate about the axiomatization of arithmetic: Otto Hölder against Robert Graßmann**. Historia Mathematica 30, pp. 341–377.

REDHEAD, Michael, 2000, **Roger Penrose – The large, the Small and the Human Mind – Review**. *Brit. J. Phil. Sci*. **51** pp. 913-917.

____________, 2004, **Mathematics ans the Mind**. *Brit. J. Phil. Sci*. **55** pp. 731-737.

REICHENBACH, H, 1958, **The philosophy of space & time**. New York: Dover.

ROSA, Luiz P., 2002, **A ruptura de paradigmas no fim do milênio - da incerteza ao caos e à complexidade**. Rio de Janeiro: COPPE-UFRJ.

ROSSER, J. B., 1936, **Extensions of some theorems of Gödel and Church**. The Journal of Symbolic Logic Vol. 1 Number 3, pp. 87-91.

____________, 1979, “Uma exposição informal das demonstrações dos teoremas de Gödel e Church”. In: **O teorema de Gödel e a hipótese do contínuo**. Ed. e tradução de Manuel Lourenço. Lisboa: Fundação Calouste Gulbenkian, pp. 385-403.

RUSSELL, Betrand (1872-1970), 1977, “Letter to Frege”. In: **From Frege to Gödel - A source book in mathematical logic, 1879-1931**. Edited by HEIJENOORT, J. V. Cambridge: Harvard University Press, pp. 124-125.

RUSSELL, Stuart J. e NORVIG, Peter, 1995, **Artificial intelligence – A modern approach**. New Jersey: Prentice Hall.

SEARLE, John R., 1980, **Minds, brains and programs**. Behavioral and Brain Sciences 3, pp. 417-424.

SINGH, Simon, 1998, **Fermat's enigma**. New York: Anchor Books.

SKOKOWSKI, Paul, 2002, **I, Zombie**. Consciousness and Cognition **11** pp. 1-9.

SOARE, Robert I., 1996, **Computability and recursion**. The Bulletin of Symbolic Logic. Vol. 2, Number 3, pp. 284-321.

STRATHERN, Paul, 1997, **Wittgenstein em 90 minutos**. Tradução de Maria Helena Geordane. Rio de Janeiro: Jorge Zahar Ed.

SUPPES, Patrick, 1966, **Introduction to logic**. 9. ed. New York: D. Van Nostrand Co.

TARSKI, Alfred (1902-1983), 1969, **Truth and proof**. Scientific American. 220, n. 6, pp. 63-77.

____________, 1995, **Introduction to logic and to the methodology of deductive sciences**. New York: Dover.

**The Encyclopedia of Philosophy**, 1972, Paul Edwards (editor in chief). New York, London: Collier MacMillan Publishers.

THIELE, Ruediger e WOS, Larry, 2002, **Hilbert's Twenty-Fourth Problem**. Journal of Automated Reasoning **29**, pp. 67–89.

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ENGENHARIA, 1997, **Norma para elaboração gráfica de teses**. Rio de Janeiro: UFRJ/COPPE.

VILKKO, Risto, 1998, **The reception of Frege's *Begriffsschrift***. HISTORIA MATHEMATICA **25**, pp. 412–422.

WANG, Hao, 1954, **The formalization of mathematics**. The Journal of Symbolic Logic, vol. 19, n. 4, pp. 241-266.

____________, 1981, **Some facts about Kurt Gödel**. The Journal of Symbolic Logic, vol. 46, n. 3, pp. 653-659.

____________, 1991, **Reflexiones sobre Kurt Gödel**. Tradução para o espanhol de Pilar Castillo Criado. Madrid: Alianza Universidad.

WITTGENSTEIN, Ludwig (1889-1951), 1994, **Tractatus Logico-Philosophicus**. Tradução, apresentação e estudo introdutório de Luiz Henrique Lopes dos Santos; [Introdução de Bertrand Russell]. 2. ed. São Paulo: Editora da Universidade de São Paulo.

ZYTKOW, Jan M., 1998, **Consciousness defies scientific and computational reduction**. *Foundations of Science* **1**, pp. 83-101.