

## Funcionalismo computacional

Bernardo Gonçalves Alonso  
Doutorando em Filosofia do PPGF-UFRJ

Hilary Putnam é o pai do Funcionalismo Computacional, teoria desenvolvida nos artigos “Minds and machines” (1960), “Robots: Machines or Artificially Created Life?” (1964) e “The nature of mental states” (1967c). Após enorme influência nos anos 60 e 70, e de importância significativa nos trabalhos de Fodor, Chalmers, Searle e Block, entre muitos outros, o Funcionalismo Computacional sofreu tantos ataques que mesmo seus arautos, Putnam e Fodor, não têm mais paciência para sua agenda filosófica. O objetivo deste artigo é apresentar os argumentos de Putnam a favor do Funcionalismo Computacional e algumas objeções. A relevância desta apresentação reside na síntese de idéias que se encontram dispersas em vários escritos ao longo de mais de 40 anos, fato que costuma confundir o leitor interessado no assunto.

Como uma aproximação preliminar, podemos caracterizar o funcionalismo como a visão de que a mente é a “organização funcional” do cérebro (Putnam 1960, 149; 1967a, 200). Outra formulação do funcionalismo diz que estados mentais são “estados funcionais” (Putnam 1967c, 30). Versões fortes ou brandas do funcionalismo podem ser formuladas dependendo do quanto da mente é considerada como sendo funcional. Todos os estados mentais são funcionais ou apenas alguns? Todos os aspectos dos estados mentais são funcionais, ou apenas os aspectos não-fenomenais?

Computacionalismo, para o presente propósito, é a visão de que a organização funcional do cérebro é computacional ou que estados neurais são estados computacionais. Para o computacionalismo vale o mesmo: versões fracas ou fortes podem ser formuladas dependendo do quanto da organização funcional do cérebro é considerada computacional. Funcionalismo + Computacionalismo = Funcionalismo Computacional. No jargão filosófico contemporâneo, a mente é o software do cérebro.

Putnam desenvolve o Funcionalismo Computacional em duas fases. Nos primeiros artigos Putnam (1960, 1964) não dá prosseguimento a uma teoria sobre a natureza dos estados mentais. Ao invés disso, utiliza a analogia entre mentes e máquinas para mostrar que:

(...)the various issues and puzzles that make up the traditional mind-body problem are wholly linguistic and logical in character. (PUTNAM, *Minds and Machines*, 1960: 362).

E somente em 1967 Putnam realiza a manobra adicional de identificar estados mentais com estados funcionais, sugerindo que:

(...)to know for certain that a human being has a particular belief, or preference, or whatever, involves knowing something about the functional organization of the human being. (PUTNAM, *The Mental Life of Some Machines*, 1967a: 424).

No artigo “The nature of mental states” (1967c), Putnam explicitamente propõe uma versão forte do funcionalismo, ao sugerir a hipótese de que a dor, ou o estado de estar com dor, é um estado funcional do organismo.

Ao propor uma analogia entre mentes e máquinas Putnam não sugere que questões associadas aos problemas mente-corpo também surgem em máquinas, alega, no entanto, que é possível clarificar questões pertinentes à mente por analogia à máquina. O tipo de máquina usada para a analogia é a Máquina de Turing, ainda o exemplo paradigmático de uma máquina de computar, um computador. Uma Máquina de Turing é um dispositivo abstrato, uma espécie de *teleprinter*, que consiste em um programa finito, um cabeçote *read-write* e uma fita de memória. A fita de memória é finita, embora indefinidamente prolongável, e é dividida em células, cada uma das quais contém um símbolo (*token*) de um alfabeto finito. A configuração inicial da fita é descrita como o “*input*”; a configuração final “*output*”. O cabeçote está sempre localizado acima de alguma das células. Ele pode escanear o símbolo impresso na célula, apagá-lo ou substituí-lo por outro símbolo. O programa consiste em um número finito de estados, por ex. A, B, C, D. A computação, que media um *input* e um *output*, progride gradualmente. A cada passo, o mecanismo *read-write* escaneia o símbolo de uma célula e a máquina, em seguida, realiza alguma das seguintes operações: (1) apaga o símbolo escaneado, substituindo-o por outro símbolo ou move o cabeçote para a célula imediatamente à direita ou à esquerda da célula escaneada; (2) muda o estado do programa da máquina; (3) *halting*. As operações que a máquina produz a cada passo são unicamente determinadas pelos símbolos escaneados e pelas instruções do programa.

A Máquina de Turing chama a atenção pela excessiva simplicidade e traz à tona argumentos em prol da superioridade das mentes sobre as máquinas. Porém, a Máquina de Turing é também teórica, na medida em que não precisa efetivamente ser construída, embora seja

essencial que seus componentes atômicos (a fita, o cabeçote, mover-se para esquerda ou para a direita, etc.) sejam tais que possam efetivamente ser implementados. Não pra menos, o objetivo de Turing era dar corpo ao processo mecânico geral de um ser humano. Sua análise não começou com máquinas de computar existentes na época, mas com livros de exercícios para crianças, aqueles de marcar quadrados em figuras. O conceito de máquina serve para capturar o que uma mente humana pode fazer quando está envolvida em um procedimento, que é a idéia do análogo. Mas é preciso ficar claro que há tantas Máquinas de Turing quanto diferentes métodos ou procedimentos. Hoje em dia é impossível evitar a imagem que não existia em 1936, a do computador. A “tabela de instruções” de uma Máquina de Turing é o equivalente ao programa de computador contemporâneo.

Como se dá a analogia com o computador? Turing escreveu o artigo “On computable numbers” (1936), onde fala da máquina universal. Infinitas Máquinas de Turing universais, formando subconjuntos de outras Máquinas de Turing. Máquinas com tabelas de instruções complexas o suficiente para lerem tabelas de outras Máquinas de Turing e aí poderem fazer o que as outras máquinas fizeram. O que poderia soar estranho tem paralelo inegável com um software que pode rodar em vários computadores. A maneira como tabelas / programas podem ler outras tabelas é crucial para a teoria de Turing, da mesma maneira que é essencial para programas de computadores modernos formarem dados que possam ser manipulados por outros programas. A Máquina de Turing permitiu uma formulação do que é computável, a saber, o que pode ser feito por uma Máquina de Turing.

Voltando a Putnam, em “Minds and Machines” (1960), o autor afirma que mentes e Máquinas de Turing não são análogas apenas no comportamento que são capazes de gerar, mas também na sua composição interna. Uma caracterização da Máquina de Turing é dada pelos programas que ela roda, que determina a ordem de sucessão de estados e os símbolos que serão impressos. Putnam se refere a esses estados como “estados lógicos” da máquina, estados que são descritos em termos formais ou lógicos, não físicos. No entanto:

(...)assim que a Máquina de Turing é fisicamente realizada, a máquina, como objeto físico, pode também ser caracterizada em termos físicos (p.ex. componentes eletrônicos). Os estados lógicos são o software e o objeto físico, hardware. (PUTNAM, *The Mental Life of Some Machines*. 1967a: 371). [Trad. do A.]

Assim como há duas descrições possíveis da Máquina de Turing,

há duas descrições possíveis do ser humano. Há uma descrição que se refere a estrutura física e química, que corresponde à descrição que se refere ao hardware do computador. Mas também outra:

(...)it would also be possible to seek a more abstract description of human mental processes in terms of 'mental states'... a description which would specify the laws controlling the order in which the states succeeded one another (PUTNAM, *Minds and Machines*, 1960: 373).

Tal descrição é análoga ao software: programas especificam leis que governam a sucessão de estados lógicos da máquina. Putnam novamente:

O mental e as descrições lógicas não são similares apenas por ambos diferirem da descrição física. São também similares por serem ambos, o pensamento e o 'programa', abertos à crítica racional (PUTNAM, *Minds and Machines*, 1960:373). [Trad. do A.]

Putnam afirma que poderíamos construir uma máquina que se comportasse de acordo com funções de preferências racionais (p.ex. regras de lógica indutiva ou até mesmo teorias econômicas) e mesmo as leis que governam a psicologia (1967b). Tal máquina de Turing pode ser vista como uma agente racional. Essa é uma analogia arrematadora entre homens e máquinas. A composição interna e o comportamento de ambos podem ser descritos, por um lado, em termos de estados físicos governados pelas leis da física, e por outro lado, de modo mais abstrato, em termos de estados lógicos e estados mentais, governados pelas leis da razão.

Putnam avança ainda mais na analogia entre máquinas e mentes, argumentando que a dor, ou qualquer outro estado mental, não é nem um estado cerebral ou uma disposição comportamental, mas um estado funcional. No artigo "The Nature of Mental States" (1967c) Putnam concorda que a dor está associada a uma certa experiência subjetiva, "quale" ou "qualia", na *parlance* filosófica, mas indaga:

(...) é necessário que um organismo tenha um cérebro para sentir dor? (PUTNAM, *The Nature of Mental States*, 1967c: 439)

Mas por que Putnam questiona a dor e qual seria a motivação de supor que a dor possa ser algo além de um estado cerebral? Buscar uma

definição para a dor implica em buscar o que é comum a todas as dores, ou equivalente que possa dar conta da dor fenomenal.

De uma maneira geral, Putnam busca a marca última do mental, aquilo que distingue um fenômeno mental do não-mental e o Funcionalismo, grosso modo, é a visão de que a marca do mental está associada com o papel que a mente exerce na vida de um organismo. Uma maneira de entendermos a abordagem funcional do mental é considerarmos as definições funcionalistas de outras entidades. Um coração é definido pelo papel que ele desempenha no corpo humano, bombear sangue, um carburador é definido pelo papel que tem no motor, misturar combustível e ar. O papel que um objeto desempenha é compreendido no contexto de um órgão maior e pode ser explicado pelas relações que ele mantém com outras partes do órgão maior. O material de que é feito não tem importância, desde que habilite o objeto a funcionar de maneira apropriada. De modo similar, o funcionalista argumenta que estados mentais são definidos por suas relações causais com outros estados mentais, *inputs* sensoriais e *outputs* motores.

Podemos considerar o Computacionalismo como a teoria de que organismos com mentes apresentam uma organização funcional, p.ex. uma descrição das instruções do organismo em termos de estados e suas relações com outros *outputs* e *inputs*. O Funcionalismo como a teoria de que possuir uma mente é ter uma certa organização funcional e que, qualquer propriedade mental é um certo tipo de organização funcional. Isso significa que estar com dor é possuir alguma propriedade que é característica dessa organização funcional. De modo geral, para qualquer propriedade mental M há uma propriedade funcional F tal que se mantenham as duas condições:

**UR F:** todo evento M é também um evento F – realização única

**SUP F:** todo evento F é também um evento M – superveniência.

E aí começam todos os problemas.

Dado o que vimos até agora, podemos tirar duas importantes conseqüências: Uma é que a dor, como um estado da organização funcional, é definida por suas relações causais com outros estados (p.ex. a crença de que tenho dor), *inputs* (levar um soco) e *outputs* (gritar 'ai'). A outra conseqüência é que a especificação da dor é redutiva no sentido em que é formulada em termos não-mentais. Ou seja, a especificação de um estado mental em termos de outros estados mentais é eliminada em favor de uma fórmula que contém termos lógicos (p.ex. existe um, e), variáveis ( $x$ ,  $S_1, \dots, S_n$ ) e termos biológicos e físicos (*inputs*, *outputs*), mas sem termos mentais.

Putnam demorou oito anos para modificar a afirmação de que estados mentais são estados de Máquinas de Turing complexas. A primeira razão para modificação é que o modelo da Máquina de Turing não pode representar perspicuamente o aprendizado nem a memória. Outro motivo foi o reconhecimento de que quando alguém está num estado de dor, está também em muitos outros estados mentais simultaneamente e a Máquina de Turing não dá conta de possíveis infinitos estados (teoria da computação, autômato probabilístico, Block e Fodor, 1972). Estas modificações não derrubaram o Funcionalismo. Este estava comprometido com o Computacionalismo, que estados mentais são estados computacionais, não o modelo da máquina de Turing. Pode ser que a organização funcional dos organismos cognoscentes seja melhor representada por redes neurais, máquinas conexionistas (Paul Churchland, 2005).

Muitos filósofos argumentaram contra o Funcionalismo. Para um exame detalhado dos argumentos contra o Funcionalismo Computacional e para compreender outras formas de funcionalismo recomendo o artigo de 1996 de Ned Block chamado “What is Functionalism?”. Vou apresentar algumas linhas do próprio Putnam contra a sua teoria. No livro “Representation and Reality” (1988), Putnam afirma que há um *Gap*, uma lacuna entre propriedades funcionais-computacionais de estados mentais, de um lado, e seus aspectos intencionais, do outro, significando que ou a SUP ou a UR é falsa. Por exemplo, contra a Realização Única Putnam aponta que o mesmo pensamento pode ser realizado em diferentes estruturas computacionais. O argumento é simplíssimo: o Funcionalismo é a teoria na qual um estado mental é definido por suas relações causais com outros estados mentais. Mas é possível que dois indivíduos, João e Maria, embora diferentes em suas organizações funcionais (por exemplo, Maria acredita numa proposição que João não acredita) acreditem que a água seja molhada. Ou não concebemos que João e Maria, ao acreditarem que a água está molhada, compartilham da mesma crença, o que parece um disparate, ou então devemos admitir que a mesma crença pode ser realizada em diferentes estados funcionais. Mas se esta última opção for o caso, então a cláusula da Realização Única deve ser abandonada. A mesma propriedade mental pode ser realizada em diferentes organizações funcionais.

Para evitar o embaraço de um argumento um tanto óbvio contra a realização única, despercebido durante anos, o funcionalista ainda poderia realizar dois movimentos: o primeiro seria fixar algumas crenças como “a água é molhada”, como analíticas. Isso tornaria a individuação de tais crenças invulnerável a realizações em diferentes organizações funcionais, movimento este que não ajuda muito (Fodor, 1992). O outro movimento seria apelar para fatos físicos, o que acabaria com as diferenças entre o

funcionalismo e o materialismo. Não foi à toa que muitos funcionalistas tiveram suas quedas fisicalistas.

Putnam diz que há uma lacuna entre o funcional e o intensional – o primeiro não determina o segundo. Parece óbvio que um programa pode ser interpretado de diferentes maneiras. Qual é a razão para pensarmos que os programas que os nossos cérebros rodam determinam o conteúdo de nossos pensamentos, crenças e estados intensionais? Vamos deixar claro o problema da múltipla interpretação. Um programa de computador é uma descrição matemática-formal que contém apenas operações lógicas e matemáticas que são definidas por um conjunto de símbolos (p.ex. 0 e 1). Mas os símbolos também têm uma dimensão semântica. Nós usamos o exemplo anterior da Máquina de Turing para somar 1 porque nós interpretamos o numeral '1' como representando o número UM, e '0' representando ZERO. Ao falarmos de semântica, dois elementos devem ser levados em conta. Um é a extensão do símbolo, que é o objeto ou o conjunto de objetos aos quais o símbolo se refere. Portanto a extensão do numeral '1' é o número UM e a extensão da palavra 'água' é o conjunto das coisas que consistem nas moléculas de H<sub>2</sub>O. A outra dimensão é o conteúdo do símbolo, que é o que faz o símbolo representar o que ele é. É o conteúdo do símbolo '1' que faz ele representar UM e não ZERO. Conteúdo é frequentemente associado a significado, sentido e intensão e sua natureza é altamente disputada no debate filosófico. No entanto, é bem menos disputado que (a) dois símbolos, p.ex. “água” e “H<sub>2</sub>O”, podem diferir em seus conteúdos, mas compartilham da mesma extensão e que (b) se dois símbolos têm o mesmo conteúdo, eles também têm a mesma extensão. Resumidamente, conteúdos determinam extensão, mas não vice-versa.

O problema fica evidente: há uma tensão entre a afirmação que o conteúdo mental é funcional-computacional e a afirmação de que o conteúdo mental determina a extensão. Se o funcionalismo estiver correto, então o conteúdo de pensamentos supervém em propriedades funcionais. Se dois organismos têm exatamente a mesma organização funcional, então o conteúdo de seus pensamentos, crenças, desejos, etc. devem ser os mesmos. *A fortiori*, seus pensamentos devem ser sobre as mesmas coisas. Se os dois organismos dizem que a água é molhada, os dois estão pensando em água e não em gatos. Por outro lado, sabemos que no caso de máquinas o programa – entidade sintático-formal – não determina a extensão dos símbolos sobre os quais as operações são definidas. Podemos inventar um dispositivo que interprete um como sendo zero e zero como sendo um. Organizações funcionais restringem os conjuntos de possíveis interpretações, mas não determinam uma interpretação única. É sempre possível que duas máquinas tenham a mesma organização funcional e que seus usuários interpretem os

símbolos pelos quais as operações são definidas de maneira diferente. De maneira similar, parece que dois organismos pensantes possam ter a mesma organização funcional embora as extensões do que dizem e pensam sejam diferentes. Nesse caso o Funcionalismo Computacional é falso. Os conteúdos dos nossos pensamentos não supervêm em propriedades funcionais.

Putnam ainda tenta fazer a manobra de adicionar a cláusula segundo a qual os *inputs* e *outputs* devem ser especificados biologicamente / fisicamente e não apenas formal-sintaticamente. Pensamos em água e não em gatos pois inputs físicos da percepção associados com este pensamento fixa o conteúdo do hidrogênio e oxigênio e descarta o conteúdo dos gatos ou qualquer maluquice. A organização funcional consistiria, portanto, em um programa implementado (autômato abstrato) mais (*plus*) I/O especificados fisicamente. Porém, no célebre artigo “The meaning of meaning” (1975), Putnam diz que apelar para I/O físicos não ajudaria tampouco. Dois indivíduos, Oscar e Toscar, podem ter exatamente a mesma organização funcional, incluindo os mesmo I/O motores e sensoriais fisicamente especificados e mesmo assim, seus conceitos, pensamentos, crenças, etc. podem ter conteúdos diferentes. Imaginem que Toscar viva em uma terra gêmea, que é exatamente como a Terra, exceto pelo fato de que o termo 'água' se refere a um líquido com a estrutura química XYZ. Oscar e Toscar não podem distinguir XYZ de H<sub>2</sub>O, pois os dois líquidos são idênticos na aparência, sabor, cheiro, textura, etc. Na Terra Gêmea há XYZ exatamente onde há H<sub>2</sub>O na Terra: rios, nuvens, torneiras, oceanos. É, pois, possível que Oscar e Toscar tenham exatamente a mesma organização funcional embora seus pensamentos sejam diferentes em conteúdo. Quando Oscar diz “a água é molhada”, ele está se referindo ao líquido que é H<sub>2</sub>O e Toscar, quando diz que “a água é molhada”, se refere ao líquido que é XYZ. Vamos supor que Oscar e Toscar vivem em 1600 e não sabem nada sobre a estrutura química da água em seus respectivos mundos. O que Oscar e Toscar sabem sobre água faz referência ao líquido que é familiar aos respectivos ambientes. No entanto os líquidos aos quais Oscar e Toscar se referem são bem diferentes. Porém, dado que o conteúdo determina a extensão e que as extensões são diferentes, os pensamentos de Oscar e Toscar devem diferir em conteúdo também. Portanto, conteúdos mentais não supervêm propriedades funcionais.

## Bibliografia

- Block, N. & Fodor, J. (1972). "What Psychological States Are Not," *Philosophical Review*, 83, 159-181
- Block, N. (1996). "What is Functionalism?" *The Encyclopaedia of Philosophy Supplement*. New York: MacMillan Reference Books.
- Churchland, P. (2005). "Functionalism at Forty: A Critical Retrospective," *Journal of Philosophy, American Philosophical Association*, 33–50.
- Fodor, J. A. (1992). "The big idea: Can there be a science of mind?" *Times Literary Supplement*. pp. 5-7.
- Putnam, H. (1960). "Minds and Machines". *Dimensions of Mind: A Symposium*. S. Hook. Collier, 138-164.
- Putnam, H. (1964). "Robots: Machines or Artificially Created Life". *The Journal of Philosophy, American Philosophical Association*, 668-691
- Putnam, H. (1967a). "The Mental Life of Some Machines". *Intentionality, Minds, and Perception*. H. Castañeda Detroit, Wayne State University Press; 177-200.
- Putnam, H. (1967b). "Psychological Predicates". *Art, Philosophy, and Religion*. Pittsburgh, PA, University of Pittsburgh Press.
- Putnam, H. (1967c). "The Nature of Mental States". *Philosophy of Mind, Contemporary Readings*. New York, Routledge.
- Putnam, H. (1975). "The Meaning of 'Meaning'". *Language, Mind and Knowledge , Minnesota Studies in the Philosophy of Science*, VII, Minneapolis: University of Minnesota Press.
- Putnam, H. (1988). *Representation and Reality*. Mass, MIT Press.
- Turing, A.M. (1936). "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society* (2), 42: 230-265.