

CDD: 501

PRAGMÁTICA DA INVESTIGAÇÃO CIENTÍFICA: UMA ANÁLISE COMPORTAMENTAL UTILIZANDO CADEIAS DE MARKOV DE PRIMEIRA ORDEM

LUIZ HENRIQUE DE A. DUTRA¹

*Departamento de Filosofia
Universidade Federal de Santa Catarina, UFSC
Florianópolis, SC, Brasil
lhdutra@hotmail.com*

CEZAR A. MORTARI

*Departamento de Filosofia
Universidade Federal de Santa Catarina, UFSC
Florianópolis, SC, Brasil
c.mortari@ufsc.br*

IVAN FERREIRA DA CUNHA

*Departamento de Filosofia
Universidade Federal de Santa Catarina, UFSC
Florianópolis, SC, Brasil
clockwork.ivan@gmail.com*

JERZY A. BRZOZOWSKI

*Centro de Filosofia e Ciências Humanas
Universidade Federal da Fronteira Sul (UFFS)
Erechim, RS, Brasil
jerzyab@yahoo.com*

¹ Este artigo foi produzido a partir de uma pesquisa realizada pelo Núcleo de Epistemologia e Lógica (NEL) da Universidade Federal de Santa Catarina (UFSC) em 2008 e 2009, com o apoio do CNPq.

Resumo: Este artigo apresenta alguns resultados preliminares da análise do comportamento científico utilizando cadeias de Markov de primeira ordem. Derivamos uma matriz de transição a partir das frequências apresentadas por um cientista ao publicar um artigo de um dado tipo seguido por um artigo de outro tipo. De modo a conduzir este estudo, primeiro determinamos tipos de artigos e então classificamos sequências (trajetórias) de artigos de acordo com tais tipos. Em seguida, utilizamos um programa de computador para derivar as matrizes de transição. Apesar da não-ergodicidade das cadeias de Markov resultantes, descobrimos que é altamente provável que artigos descrevendo resultados experimentais apareçam ao final da trajetória.

Palavras-chave: pragmática da investigação científica, naturalismo, análise comportamental, comportamento científico, cadeias de Markov.

PRAGMATICS OF SCIENTIFIC INVESTIGATION: A BEHAVIORAL ANALYSIS USING FIRST-ORDER MARKOV CHAINS

Abstract: This paper presents some preliminary results of the analysis of scientific behavior using first-order Markov chains. We derive a transition matrix from the frequencies with which a scientist publishes an article of a given type followed by an article of another type. In order to conduct this study we first determined article types and then classified sequences (trajectories) of articles according to such types. Afterwards we ran the trajectories through a software for deriving the transition matrices. Despite the non-ergodicity of the resulting Markov chains, we found that it is highly probable that articles describing experimental results are at the end of the trajectory.

Keywords: pragmatics of scientific investigation, naturalismo, behavior analysis, scientific behavior, Markov chains.

1. Introdução

Em um artigo publicado recentemente (DUTRA et al. 2011), propusemos que a atividade científica fosse estudada de um ponto de vista pragmático, em parte empírico, envolvendo uma análise do comportamento (verbal) manifesto dos cientistas em seus artigos científicos publicados. Propusemos então que tal análise fosse pautada pela possibilidade de recursos matemáticos, o que permitiria que regularidades do comportamento científico fossem observadas, mapeadas e tratadas rigorosamente, ainda que em termos probabilísticos. Trata-se de uma

abordagem nomológica a um estudo pragmático da ciência, abordagem que procura encontrar correlações nomológicas que possam descrever alguns aspectos da atividade científica. O mencionado artigo era de caráter programático. O presente artigo, por sua vez, apresenta um primeiro resultado de tal programa de pesquisa (cf. também DUTRA, 2008, caps. 3 e 8, e DUTRA, 2006).

O instrumental matemático aqui adotado é o da análise por cadeias de Markov – uma ferramenta frequentemente utilizada para encontrar padrões no comportamento de escolha de consumidores (SUPPES, 2003). Apesar de haver diferenças importantes entre os padrões comportamentais de consumidores e cientistas, a atividade científica pode ser caracterizada como uma atividade que envolve escolha. SHIMP (2007) expressou recentemente a necessidade de uma análise quantitativa do comportamento dos próprios analistas quantitativos de comportamento. Enquanto tal análise reflexiva de comportamento não estiver disponível, propomos um modelo simples de cadeias de Markov que poderia ser um passo em direção à consolidação da análise quantitativa do comportamento científico em geral.

O objetivo do presente estudo é o de aplicar uma análise em termos de cadeias de Markov a uma sequência de tipos de artigos publicados por determinado cientista. Presumimos que publicar um artigo de determinado tipo, e não de outro, é um comportamento de escolha. Como tal, publicar artigos é uma atividade que pode estar sujeita a formas parecidas de determinação que afetam outros tipos de comportamento de escolha. Acreditamos que é razoável assumir que algumas formas de comportamento científico são mais reforçadoras que outras. Por exemplo, em muitas circunstâncias, é mais fácil realizar experimentos com um modelo existente do que propor um novo modelo para dar conta de certo fenômeno. Estudar uma sequência de artigos pode então trazer alguma luz a respeito de quais atividades científicas são as mais reforçadoras.

Nossa abordagem é uma forma de behaviorismo molar², apesar de diferente do behaviorismo teleológico de RACHLIN (1994). Acreditamos que comportamentos molares podem ser levados em consideração de maneira puramente matemática, como no caso considerado por HERRNSTEIN (1997),

² O termo “molar” refere-se a um nível de descrição mais macroscópico, em oposição a um nível mais “molecular” (em analogia com a terminologia da química).

com sua lei de igualação. De acordo com tal lei, havendo comportamentos alternativos (e complementares) em determinado contexto, a proporção (temporal) de resposta relativa para cada comportamento será proporcional à taxa de reforço relativo de cada um desses comportamentos. Dessa forma, seguindo essa proposta, a escolha é apenas um comportamento apresentado no contexto de outro comportamento (mais amplo; cf. HERRNSTEIN, 1997, p. 22). Ou, mais exatamente, é a tendência de igualar o tempo gasto em uma atividade à sua taxa relativa de reforço, dado certo contexto de atividades alternativas e complementares, presumindo que o comportamento total do contexto é invariante.

HEYMAN (1979) sugeriu que um modelo em termos de cadeias de Markov poderia explicar a base de nível molecular do padrão de igualação correspondente em nível molar. Em um esquema com comportamentos de intervalos variáveis concorrentes ou alternativos, de forma semelhante à análise de Herrnstein, acima mencionada, o modelo de cadeias de Markov de Heyman prevê que a probabilidade de mudar de uma alternativa de reforço para outra (a probabilidade de transição) é proporcional às frequências gerais de reforço relativo. Esse resultado sugere que um modelo de cadeias de Markov pode ser usado para inferir taxas de reforço a partir de probabilidades de transição. Pressupondo então a perspectiva molar de Herrnstein, o modelo de Heyman se desenvolve na direção da análise que pretendemos fazer, utilizando uma ferramenta matemática mais poderosa.

No presente artigo, descrevemos probabilidades de transição para uma trajetória de investigação, isto é, uma série de artigos a respeito de certo modelo publicado por um cientista. As probabilidades de transição representam a frequência com a qual o cientista mudou de um tipo de artigo para outro. Como ilustração, podemos pensar em uma trajetória de investigação composta apenas por dois artigos de divulgação. Nesse caso, a frequência de passar de um artigo de divulgação para outro artigo de divulgação seria 1,0. Procuramos retratar o todo das probabilidades de transição para oito tipos de artigos (descritos abaixo), apresentados na forma de matrizes de transição de cadeias de Markov.

Já que a pesquisa que este artigo relata é a fase empírica de uma investigação pragmática, adotamos aqui a forma de apresentação mais direta dos artigos científicos, de modo a informar as discussões posteriores, de caráter analítico, e torná-las mais compreensíveis. Assim, as seções a seguir foram

nomeadas “Metodologia”, “Resultados” e “Discussão”. Voltaremos aos aspectos filosóficos relativos a essa abordagem nas considerações finais.

2. Metodologia

2.1. Tipos de artigo

De modo a realizar o programa de pesquisa sugerido acima, o primeiro passo foi determinar os tipos de artigos. Depois de alguns estudos preliminares, chegamos aos seguintes tipos:

(Tipo A) Artigo que propõe um novo modelo para lidar com determinado fenômeno ou coleção de fenômenos.

(Tipo B) Artigo em que são apresentadas generalizações e/ou pequenas correções ao modelo ou a uma aplicação do modelo. Tais generalizações e correções podem se referir a um modelo proposto em um artigo do tipo A ou a uma aplicação do modelo, apresentada em um artigo do tipo C ou do tipo E.

(Tipo C) Artigo que propõe uma aplicação do modelo. Este tipo de artigo mostra que o modelo apresentado em um artigo do tipo A pode ser útil para explicar algum outro grupo de fenômenos dentro da mesma área de conhecimento em que o modelo foi originalmente proposto.

(Tipo D) Artigo que traz explicações sobre a aplicação do modelo; essas explicações podem se referir a uma aplicação do modelo apresentada originalmente em artigos do tipo C ou do tipo E.

(Tipo E) Artigo que propõe uma aplicação do modelo em uma área do conhecimento diferente daquela para a qual o modelo foi originalmente proposto. Este tipo de artigo é parecido como o de tipo C, mas trata de um grupo diferente de fenômenos.

(Tipo F) Artigo que apresenta resultados experimentais da aplicação do modelo; pode se referir a um artigo do tipo C ou do tipo E.

(Tipo G) Artigo que revisa ou divulga o modelo e/ou suas aplicações. Seu propósito é tornar pública a pesquisa em desenvolvimento, seja entre o público em geral, seja na própria comunidade científica.

(Tipo H) Artigo apresentando falhas no modelo, ou em um modelo concorrente.

No início, estávamos trabalhando com a possibilidade de classificar um artigo como híbrido, isto é, um artigo apresentando características de dois ou mais tipos. No entanto, descobrimos alguns problemas metodológicos com essa técnica de classificação; devido ao fato de que a maioria dos artigos seria híbrida, nossos dados apresentariam poucas ocorrências de cada tipo híbrido. Em particular, os artigos do tipo H nunca seriam encontrados em forma pura, por assim dizer, simplesmente porque sempre que uma falha fosse encontrada, seja no próprio modelo, seja em outros modelos, tal falha precisaria ser corrigida ou explicada. Isso é óbvio mesmo do ponto de vista mais ingênuo da ciência: normalmente, não se critica algo sem oferecer uma alternativa. Assim, os artigos do tipo H seriam sempre encontrados em formas híbridas – junto com uma reexplicação do fenômeno ou uma correção do modelo.

Diante de tais considerações, optamos por trabalhar apenas com os tipos predominantes de cada artigo, e, dessa forma, mudamos a definição do artigo de tipo H da seguinte maneira:

(Tipo H) Artigo no qual uma falha em um modelo é detectada e corrigida. Tal correção pode ser por meio de uma nova aplicação do modelo, ou por meio de uma correção do modelo, ou ainda por meio de uma nova explicação do fenômeno.

2.2. A ferramenta de análise Markov

Do ponto de vista comportamental, nosso estudo permite a inferência de leis estocásticas que dirigem o comportamento científico. De modo a computar tais medidas de probabilidade, alimentamos com nossos dados o software Remarkov, um script Ruby para a análise Markov elaborado especialmente para nosso projeto. Dada uma trajetória (t) ou uma série (s), isto é, a sequência de publicação de artigos como input em texto composto pelas letras correspondentes aos tipos de artigos, o Remarkov gera matrizes de transição, diagramas de estados, calcula taxas de entropia, e realiza análises estatísticas. Nesse aspecto, nosso software opera de maneira similar ao *Discussion Analysis Tool* de JEONG (2005), com a vantagem de ser independente de plataforma e ser capaz de operar em ordens maiores que 1. Os outputs são apresentados na forma de tabelas HTML e grafos Graphviz.

3. Resultados

3.1. Classificação

De posse das categorias de artigos, pudemos descrever uma série de artigos científicos como uma série de ocorrências dos tipos de artigos. Dessa forma, podemos nos referir a tal série como uma trajetória de investigação, ou simplesmente uma trajetória. Nosso objeto de análise foi a trajetória de um físico, o sujeito da pesquisa. Tal escolha foi devida à grande quantidade de artigos disponíveis no website desse autor. Estudamos 158 artigos publicados por ele entre 1990 e o início de 2009, e classificamos tais artigos de acordo com nossos tipos de artigos, acima apresentados. O resultado foi a série de ocorrências nomeada como ‘trajetória 1’ (t1), mostrada na Fig. 1.

1	AACCBFFFCF	CCDCFGECCD	FDCBGDGEDC	BDFHFGCDFF	FCFFFFFFHF
51	DGBFFGCCDE	CFCGEDDEFF	FFFEDEBGBHF	CECFEEGFDG	BEEEGFGDBG
101	GFEFFFDGFG	FBFGDFGDFD	GHDDGGFFHE	GBFFGGHFEF	FGEEFCHGHF
151	FDDECFEE				

Figura 1. Trajetória 1 (t1), compreendendo 158 artigos publicados entre 1990 e 2009. Os números à esquerda (1, 51, 101 e 151), assim como a organização dos signos em grupos de dez, têm como objetivo facilitar a contagem e a localização de pontos específicos na trajetória

1	AAAABBCCCF	CCDCFGECCD	FDCBGDGEDC	BDFHFGDFF	FFFFFFDHF
51	DGBFFDEFFF	FFCFGGDFD			

Figura 2. Trajetória 1, série 1 (t1.s1), compreendendo 70 artigos publicados entre 1990 e 2005.

1	AAAACCGED	DEFFEDEBGBH	CEEGDBEEE	GFGBBGFEFF	FDGFBGFFG
51	FFGHDDGGFF	HEGBFFGGHF	EFFGEEFGHG	HFGFDECFEE	

Figura 3. Trajetória 1, série 2 (t1.s2), compreendendo 90 artigos publicados entre 1999 e 2009, exceto os dois primeiros artigos (veja a Tabela 1 para mais detalhes).

Os dois artigos do tipo A no início da trajetória 1, t1.a1-2, descrevem um modelo matemático amplo, relacionado com a geometria fractal, modelo esse que poderia ser aplicado a diferentes áreas. Tendo proposto o modelo naqueles dois artigos, o sujeito continuou a investigar suas consequências em dois domínios de pesquisa, que podem ser tratados como duas subtrajetórias distintas, ou séries, t1.s1 (Fig. 2) e t1.s2 (Fig. 3). Tais séries, entretanto, foram entendidas como sendo a respeito do mesmo modelo matemático apresentado inicialmente. Para refletir tal dependência, os dois artigos tipo A, t1.a1-2, foram considerados parte tanto de t1.s1 quanto de t1.s2. Assim, t1.a1-2 são os mesmos artigos que t1.s1.a1-2 e que t1.s2.a1-2, como mostrado na Tabela 1.

Considerando, porém, que nosso objetivo é explicar um padrão de comportamento como uma função do contexto em que tal comportamento ocorre, alguns artigos tiveram que ser reclassificados depois dessa divisão de t1 em t1.s1 e t1.s2. Esse é o caso de t1.s1.a3-4, classificados como tipo A na primeira série, e tipo C quando consideramos a trajetória toda. Tais artigos foram considerados como do tipo A por causa do papel que desempenham na série.

Isso quer dizer que, se olharmos para a trajetória toda, veremos propostas de aplicação empírica do modelo mais amplo sobre matemática pura. No entanto, se dirigirmos nossas atenções para a primeira série, veremos que tais aplicações adquirem o caráter de submodelos durante os anos seguintes. Da mesma forma, os artigos do tipo C t1.s1.a7-8 são propostas de aplicação empírica do modelo tal como ele é apresentado nos artigos t1.s1.a3-4. Há referência explícita a esses artigos e não aos artigos t1.s1.a1-2 que originaram a série. Quando consideramos a trajetória geral, notamos que t1.a7-8 (que são os mesmos que t1.s1.a7-8) são predominantemente relatórios de experimentos. Isso acontece porque, ao levar em consideração os artigos t1.a1-2, eles estão se referindo aos experimentos propostos nos artigos do tipo C t1.a3-4.

Uma situação similar acontece na segunda série. Os artigos do tipo A t1.s2.a3-5 correspondem à sequência 'G C C' sublinhada em t1.a56-58. Tais artigos são considerados como sendo do tipo A por causa do papel que desempenham na série. A segunda série, de fato, começa com o artigo do tipo G t1.a56, publicado em uma revista de divulgação popular, que tinha como objetivo chamar a atenção da comunidade científica para a possibilidade de explicar alguns fenômenos com o modelo de geometria fractal que tinha tido bastante sucesso na primeira série. Depois disso, o sujeito publicou dois artigos técnicos mostrando como o modelo poderia ser aplicado àqueles fenômenos. É

importante ressaltar que, no momento em que aqueles artigos foram publicados, não era possível classificá-los como sendo do tipo A: eles seriam certamente classificados como artigos do tipo C. A característica de ser do tipo A só emerge depois que outros artigos aparecem, trazendo aplicações, experimentos, explicações etc., que fazem referência àquele artigo e às possibilidades que ele oferece.

Outra consideração a ser feita é que não é o caso que, quando a segunda série começa, a primeira série termine. As duas séries se sobrepõem por algum tempo, mas, à medida que a primeira série começa a dar menos resultados, isto é, à medida que o número de artigos publicados em relação à primeira série começa a escassear, o número de artigos relativos à segunda série se torna maior. A Tabela 1 mostra as duas séries agrupadas por ano de publicação de seus artigos constituintes.

Ano	1ª série	2ª série
1990	A A (comuns às duas séries)	
1991	A A B B	
1992	C C C F C	
1993	C D C F	
1994	G E C C D	
1995	F D C B	
1996	G D G E D C B D	
1997	F H F G F D F	
1998	F F F F F F F F	
1999	D H F D G B F F D E F	A A A C
2000	F F	C G E D D E F
2001	F F C F	F E D B G B H C E
2002	F G D	E E G D B E E E G F G
2003		B B G F E F F F
2004	F F	D F G B F G F F G
2005	D	F F G H* D D G* G F
2006		F H* E* G B* F
2007		F G G* H* F* E** F F* G** E** E* F G* H** G
2008		H* F G** F* D* E** C** F
2009		E** E*

Tabela 1. Série 1 e série 2 agrupadas agrupadas por ano de publicação de seus artigos constituintes.

A Tabela 1 mostra que em 1998, o ano anterior à mudança de foco da pesquisa do sujeito, houve apenas a publicação de artigos do tipo F. Notemos também que em 1997, os artigos de tipo F representam mais do que 50% dos artigos publicados. Isso quer dizer que, depois de quase oito anos, a pesquisa tinha se tornado uma atividade de fazer experimentos. Podemos dizer isso a partir do comportamento manifesto do sujeito, que consistia na publicação majoritária de artigos do tipo F, isto é, relatórios de experimentos. Depois dessa sequência de artigos do tipo F, o sujeito mudou a área de pesquisa, resultando na diminuição de publicações da primeira série que mencionamos acima. Até o ano de 2009, quatro anos haviam se passado sem publicações relativas à primeira série.

É possível notar que, na segunda série, há uma ocorrência crescente de duas novas aplicações do modelo original – uma relacionada às Ciências Sociais e outra à Bioquímica. Os artigos relativos a essas áreas estão marcados na tabela acima com * e com ** respectivamente. Se retirarmos tais artigos da série, já que é possível que eles se tornem artigos do tipo A em novas subtrajetórias, notamos um padrão interessante (Fig. 4). A predominância de artigos do tipo F, ou ainda, a recorrência do padrão ‘G F F’, pode ser notada ao final da série. Isso se torna ainda mais evidente se considerarmos a Tabela 2, em que os artigos não marcados da segunda série estão agrupados por ano.

1	AAAAACCGED	DEFFEDBGBH	CEEEGBESEE	GFGBBGFEFF	FDGFBFGFFG
51	FFGDDGFFGF	FGFFGFF			

Figura 4. Trajetória 1, série 2, com os artigos marcados removidos (t1.s2u).

É possível notar na Tabela 1 que os artigos marcados começaram a aparecer em 2005 e 2006, enquanto que em 2003 e 2004, de acordo com a Tabela 2, há predominância de artigos dos tipos F e G. A partir de 2005, notamos uma diminuição de artigos não marcados na segunda série. O padrão persiste até 2008, ano no qual apenas dois artigos, ambos do tipo F, foram publicados em tal trajetória. Baseado no padrão de, por assim dizer, degeneração, observado no final da primeira série, é possível inferir que a segunda série também está degenerando. Se for assim, podemos prever a emergência de, pelo menos, uma

nova série na trajetória de pesquisa do nosso sujeito. Os artigos que estabelecem as pontes entre o modelo e as ciências sociais e a bioquímica são candidatos a aparecerem como sendo do tipo A numa nova série.

Ano	Tipo (t1s2 não marcados)
1990	A A
1991-1998	
1999	A A A C
2000	C G E D D E F
2001	F E D B G B H C E
2002	E E G D B E E E G F G
2003	B B G F E F F F
2004	D F G B F G F F G
2005	F F G D D G F
2006	F G F
2007	F G F F G
2008	F F
2009	

Tabela 2. Artigos não marcados da segunda série, agrupados por ano.

3.2. Análise Markov

Utilizamos o software Remarkov para analisar a trajetória t1 e as séries t1.s1, t1.s2, e t1s.2u, o que resultou em quatro matrizes de transição e quatro digramas de estado. Tais resultados são apresentados nas Tabelas 3-6 e nas Figs. 5-8. Em cada matriz de transição, a primeira coluna, à esquerda, mostra os estados presentes e a primeira linha, no alto, os estados futuros. Como é usual em matrizes estocásticas, cada coluna representa um vetor probabilidade, isto é, a soma das probabilidades de transição é igual a 1. Assim, o valor em qualquer célula representa a probabilidade de um artigo do tipo (linha) ser seguido por um artigo do tipo (coluna). Por exemplo, vemos na Tabela 3 que um artigo do tipo C sucede um artigo do tipo A com probabilidade de 0,5 (segunda linha, terceira coluna).

	A	B	C	D	E	F	G	H	Entropia
A	0.50	0	0.50	0	0	0	0	0	1.00
B	0	0.09	0	0.09	0.09	0.36	0.27	0.09	2.29
C	0	0.15	0.20	0.20	0.05	0.30	0.05	0.05	2.51
D	0	0.10	0.14	0.14	0.14	0.29	0.19	0	2.50
E	0	0	0.21	0.16	0.26	0.21	0.16	0	2.30
F	0	0.02	0.12	0.10	0.10	0.41	0.20	0.06	2.38
G	0	0.16	0.08	0.16	0.16	0.20	0.12	0.12	2.76
H	0	0	0	0.13	0.13	0.63	0.13	0	1.57

Tabela 3. Matriz de transição de primeira ordem para t1. Números em negrito indicam um valor-z (desvio padrão) fora do intervalo [-2, 2], veja abaixo para mais detalhes.

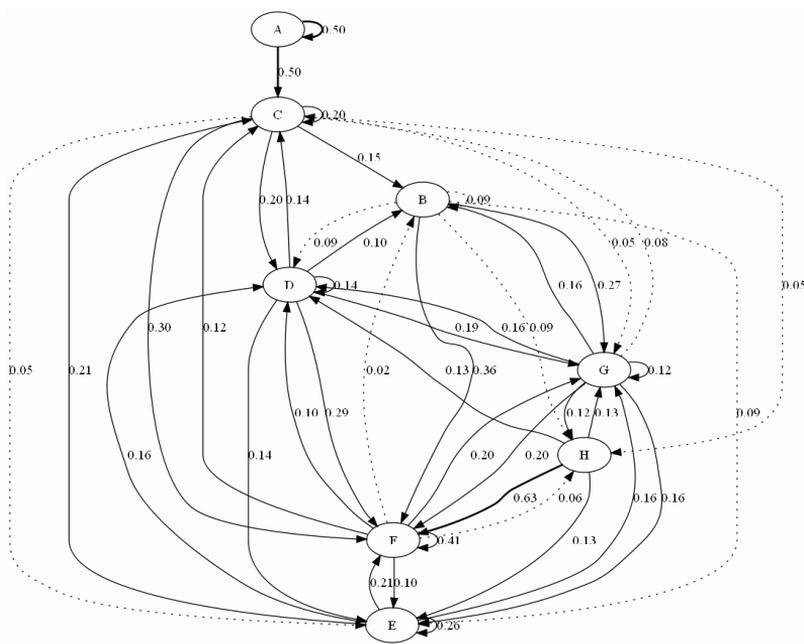


Figura 5. Diagrama de estados para t1. Setas em negrito indicam transições com $p \geq 0,50$; já setas pontilhadas indicam transições com $p < 0,10$.

	A	B	C	D	E	F	G	H	Entropia
A	0.75	0.25	0	0	0	0	0	0	0.81
B	0	0.20	0.20	0.20	0	0.20	0.20	0	2.32
C	0	0.18	0.36	0.18	0	0.27	0	0	1.93
D	0	0	0.27	0	0.09	0.36	0.18	0.09	2.11
E	0	0	0.33	0.33	0	0.33	0	0	1.58
F	0	0	0.07	0.22	0	0.56	0.11	0.04	1.75
G	0	0.17	0	0.33	0.33	0.17	0	0	1.92
H	0	0	0	0	0	1.00	0	0	0.00

Tabela 4. Matriz de transição de primeira ordem para t1.s1.

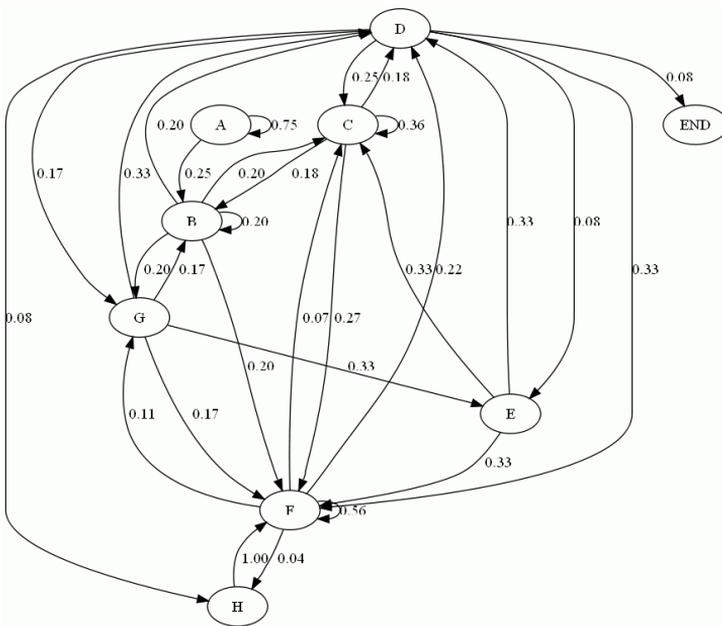


Figura 6 Diagrama de estados para t1.s1.

	A	B	C	D	E	F	G	H	Entropia
A	0.80	0	0.20	0	0	0	0	0	0.72
B	0	0.14	0	0	0.14	0.29	0.29	0.14	2.23
C	0	0	0.25	0	0.25	0.25	0.25	0	2.00
D	0	0.25	0	0.25	0.25	0.13	0.13	0	2.27
E	0	0	0.06	0.13	0.38	0.25	0.19	0	2.11
F	0	0	0	0.08	0.17	0.33	0.38	0.04	1.97
G	0	0.21	0	0.05	0.11	0.32	0.11	0.21	2.39
H	0	0	0.17	0.17	0.17	0.33	0.17	0	2.27

Tabela 5. Matriz de transição de primeira ordem para t1.s2.

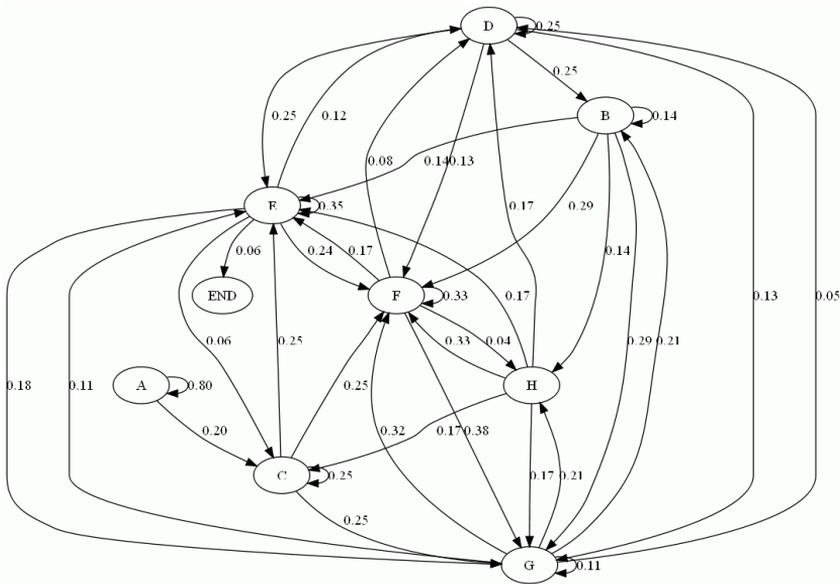


Figura 7 Diagrama de estados para t1.s2.

	A	B	C	D	E	F	G	H	Entropia
A	0.80	0	0.20	0	0	0	0	0	0.72
B	0	0.17	0	0	0.17	0.17	0.33	0.17	2.27
C	0	0	0.33	0	0.33	0	0.33	0	1.58
D	0	0.29	0	0.29	0.14	0.14	0.14	0	2.23
E	0	0	0	0.20	0.40	0.20	0.20	0	1.92
F	0	0	0	0.05	0.10	0.57	0.40	0	1.60
G	0	0.21	0	0.14	0.07	0.57	0	0	1.60
H	0	0	1.00	0	0	0	0	0	0.00

Tabela 6. Matriz de transição de primeira ordem para t1.s2u.

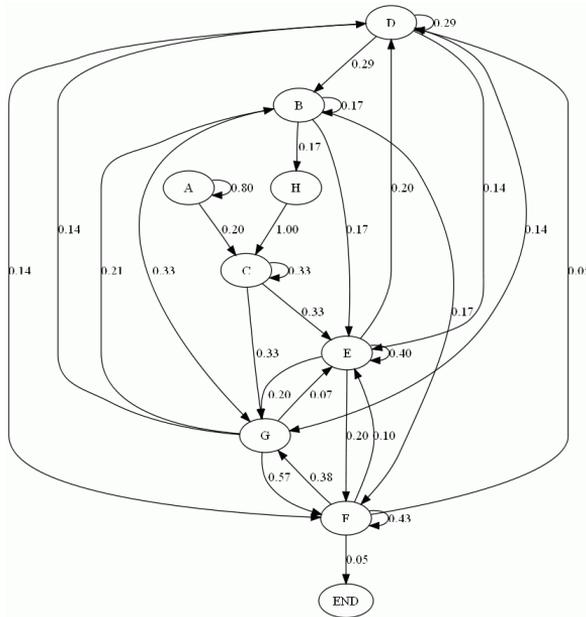


Figura 8 Diagrama de estados para t1.s2u.

4. Discussão

Os altos valores de probabilidade para que um artigo do tipo A tenha como sucessor outro artigo do tipo A são devidos a erros de amostragem. Há apenas uma transição de A para A em t1, três em t1.s1 e t1.s2, e quatro em t1s2u. Em todas as trajetórias, tais transições representam menos do que 1% de todas as transições.

Considerando a matriz de transição para t1 (Tabela 3), podemos notar que o estado G é aquele com a maior entropia. De acordo com o critério de SUPPES (2003), isso representaria o estado com a maior liberdade de escolha, isto é, um cientista poderia escolher publicar um artigo de qualquer tipo depois de um artigo do tipo G. Esse resultado, entretanto, não se aplica a cada uma das séries em particular. Quando as séries são tomadas individualmente, B é geralmente o estado mais livre. Isso pode ser devido ao fato de que os modelos científicos exigem manutenção periódica e regular.

O padrão F recorrente, no caso de t1.s1, é evidente na matriz de transição correspondente (Tabela 4), na qual a probabilidade de um artigo do tipo F ocorrer depois de outro artigo do tipo F é 0,56, com um valor-z de 2,26. O padrão 'G F F' em t1.s2u também é saliente na matriz de transição, apesar de os valores-z não serem tão significativos.

Outro padrão relevante é a recorrência de artigos do tipo E ao longo de t1.s2.a21-30. Tal padrão também acontece na série não marcada, onde isso é ainda mais proeminente. A recorrência de artigos do tipo E pode estar restrita ao modelo particular considerado em t1.s2, e pode ser explicada pelo fato de que o modelo é atraente para diversos domínios de explicação fora daquele em que foi originalmente proposto.

Devido ao fato de que o tamanho da amostra é relativamente pequeno, a cadeia de Markov derivada é não ergódica e, assim, dependente de um estado inicial. Dessa forma, as distribuições de probabilidade estacionária não podem ser calculadas com precisão, e não conseguimos comparar nossos valores de entropia com aqueles de SUPPES (2003). A não ergodicidade pode também ser devida ao desenho experimental: a escolha de incluir artigos do tipo A pode ter viciado a análise. Eliminar os artigos dos tipos A e H da análise resultaria em uma cadeia quase ergódica, cujas lacunas (isto é, transições com probabilidade zero) podem ser preenchidas ao estudar outras trajetórias de investigação.

Uma imagem mais completa da pragmática da investigação científica só pode ser esboçada ao realizar mais análises como esta desenvolvida aqui. É possível que cada área de pesquisa científica tenha seu próprio perfil de matrizes de transição. Na área estudada aqui (Física), um modelo matemático às vezes precede sua aplicação empírica. Em outras áreas – por exemplo, a Psicologia e a Biologia –, isso raramente é o caso, e pode fazer com que nossa taxonomia de artigos mude radicalmente. Entretanto, pode haver alguns elementos em comum, tais como uma grande taxa de entropia para artigos de divulgação e revisão (artigos do tipo G). Se toda pesquisa científica é igualmente restrita pelos mesmos tipos de reforço, as taxas de entropia também deveriam se mostrar como algo que não varia muito entre diferentes áreas científicas.

5. Considerações finais

Em DUTRA et al. (2011), o artigo que deu origem à pesquisa relatada aqui, discutiu-se o valor dos métodos matemáticos. A questão levantada naquele texto é a de como uma metodologia quantitativa pode ser utilizada sem desfigurar o fenômeno a ser analisado. No presente artigo, temos uma amostra do que é possível fazer com observações empíricas e a utilização de ferramentas matemáticas para encontrar regularidades em tais observações. Esses procedimentos não são comuns na filosofia, apesar de diversos autores já terem levantado a bandeira do naturalismo, isto é, a ideia de que o conhecimento e a ciência devem ser tomados como fenômenos naturais e, como tal, devem ser estudados utilizando a metodologia das ciências naturais. Nossa pesquisa, assim como outras pesquisas de inclinação naturalista na filosofia da ciência, está ainda dando seus primeiros passos, de modo que ainda não é possível dar um veredicto a respeito do naturalismo. Uma primeira impressão é a de que outros tipos de estudo são necessários para uma visão mais completa e abrangente da ciência. Deve ficar claro, no entanto, que a filosofia perde muito ao negligenciar os métodos empíricos e matemáticos em sua missão de compreender a ciência.

Contudo, a questão filosófica crucial é o que ela ganharia com esse tipo de abordagem. Um estudo empírico do comportamento cognitivo, mesmo que generalizemos seus resultados, está longe de responder às inquietações epistemológicas tradicionais. Mesmo que os resultados empíricos sejam relevantes e que possamos argumentar que eles representam com considerável

fidelidade determinados fatos cognitivos (no caso da pesquisa relativa a este artigo, o comportamento verbal técnico de um cientista quando publica seus artigos), do ponto de vista filosófico mais tradicional, sempre se poderia argumentar que não podemos tomar tais fatos como sugestivos de normas epistemológicas. A epistemologia tradicional, de caráter normativo, se interessa, antes, por formas racionais (ou lógicas) por meio das quais possamos validar os resultados da pesquisa científica. A discussão sobre a natureza da epistemologia, se ela deve ser normativa ou apenas descritiva, é hoje uma antiga polêmica desde o célebre artigo de QUINE (1969), “Epistemology naturalized”, e não pretendemos entrar aqui de novo nessa discussão. Ela pode sempre continuar, apelando para diferentes orientações filosóficas e acrescentando novos argumentos em favor de uma ou de outra posição. Independentemente disso, estudos de caráter mais empírico, entre os quais o mais célebre é a *Estrutura* de KUHN (1970 [1962]), têm contribuído para o entendimento de aspectos da atividade científica negligenciados pela perspectiva tradicional e normativa, que visa sempre ao produto mais abstrato da atividade científica (a teoria, por exemplo), e desconsidera o processo que conduz a tais produtos abstratos. O que tornou tão célebre a obra de Kuhn foi sua discussão detalhada de certos aspectos desse processo.

Se um estudo pragmático da atividade científica, como o de Kuhn, pode contribuir para nosso entendimento da ciência real, acreditamos que o uso de métodos alternativos, como aquele que empregamos na pesquisa relatada neste artigo, também o podem, obviamente. A multiplicação de descrições do comportamento científico, é claro, não pode atrapalhar nossa compreensão da ciência, mesmo que não nos permita avançar – se o quisermos – nas discussões normativas tradicionais da filosofia do conhecimento. Apenas ela conflita com a concepção tradicional, intelectualista, do conhecimento humano e da ciência, segundo a qual temos controle consciente e voluntário de nosso comportamento cognitivo. Para tal concepção, falarmos de padrões do comportamento verbal do cientista sugere que estamos querendo suprimir a razão e a liberdade, noções tão caras à filosofia tradicional. Mas, como mostra o texto de SUPPES (2003) aqui citado, não se trata de argumentar que não há liberdade e racionalidade, mas de compreender essas noções empiricamente.

No presente artigo, assim como naquele publicado anteriormente, DUTRA et al. (2011), procuramos ir na mesma direção, como já mencionamos, mas acrescentamos à discussão duas noções que são comuns na análise empírica

do comportamento e que, à primeira vista, parecem conflitar com os conceitos tradicionais. Trata-se das noções de molaridade e de controle. Na literatura da tradição behaviorista esses são temas importantes. Mas o filósofo está pouco habituado a essas noções, e o pouco que pode entender delas lhe parece conflitar com sua imagem intelectualista do conhecimento humano. De fato, conflita, mas por razões um tanto diferentes daquelas que o filósofo tradicional possa imaginar.

A molaridade e o controle entram em cena nos eventos cognitivos que discutimos neste artigo da seguinte maneira, que resume e generaliza os resultados empíricos acima apresentados: o comportamento verbal do cientista pode ser mais bem compreendido quando tomamos porções maiores do que ele comunica a seus pares (molaridade), como em uma longa série de artigos científicos que ele publica. Além disso, o tratamento matemático que demos a tais fatos cognitivos mostra que seu comportamento verbal está sob o controle de variáveis ambientais, nesse caso, sua própria história de publicações, a trajetória que seu comportamento verbal técnico seguiu, trajetória essa que exhibe um padrão. O mais interessante na interpretação dos dados empíricos e desse padrão é que não temos aí razões para argumentar em favor da supressão da liberdade, mas, ao contrário, justamente, em favor do comportamento verbal menos determinado, que se mostra como uma possibilidade de fato. Como vimos, resgatando a ideia de Suppes de que isso está relacionado com a entropia, trata-se apenas de ver a liberdade e a racionalidade de outra forma.

Por outro lado, se isso ainda estiver longe das aspirações do filósofo tradicional, não há razão para desmotivar as pesquisas empíricas sobre o conhecimento humano e a ciência, mas, ao contrário, para multiplicá-las. Se uma análise pragmática da investigação científica for ter relevância mesmo para a filosofia intelectualista tradicional, só a proliferação dessas pesquisas poderá mostrar.

Submetido: 14.11.2014; Aceito: 07.04.2015

Referências bibliográficas

- DUTRA, L.H.A. (2006). Comportamento intencional e contextos sociais: uma abordagem nomológica. *Abstracta* 2(2): 102-28.
- (2008). *Pragmática da investigação científica*. São Paulo: Loyola.
- DUTRA, L.H.A.; MORTARI, C.; BRZOZOWSKI, J. & BATISTA, T. (2011). Pragmática da investigação científica: uma abordagem nomológica. *Scientiae Studia* 9(1): 167-87.
- HERRNSTEIN, R.J. (1997). *The matching law: papers in psychology and economics*. Cambridge (MA): Harvard University Press.
- HEYMAN, G.M. (1979). A Markov model description of changeover probabilities on concurrent variable-interval schedules. *Journal of the Experimental Analysis of Behavior* 31: 41-51.
- JEONG, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education* 26(3): 367-83.
- KUHN, T. S. (1970) [1962]. *The structure of scientific revolutions*. Chicago: The University of Chicago Press. Em português: (1978). *A estrutura das revoluções científicas*. Trad. B.V. Boeira & N. Boeira. São Paulo: Perspectiva.
- QUINE, W.V.O. (1969). Epistemology naturalized. In: ———. *Ontological relativity and other essays*. New York: Columbia University Press, p. 69-90. Em português: (1975). Epistemologia naturalizada. Trad. Andréa Loparić. Coleção Os Pensadores 52: Ryle/Austin/Quine/Strawson. São Paulo: Abril Cultural, p. 163-75.
- RACHLIN, H. (1994). *Behavior and mind: the roots of modern psychology*. New York: Oxford University Press.
- SHIMP, C. P. (2007). Quantitative behavior analysis and human values. *Behavioural Processes* 75(2): 146-55.
- SUPPES, P. (2003). Rationality, habits and freedom. In: DIMITRI, N.; BASILI, M. & GILBOA, I. (orgs.). *Cognitive processes and economic behavior*. New York: Routledge, p. 137-67.