



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília



**CULTURA
ACADÊMICA**
Editora

Big Data e Autonomia:

Continuidade ou Revolução?

Edna Alves de Souza

Maria Eunice Quilici Gonzalez

Como citar: SOUZA, E. A.; GONZALEZ, M. E. Q. Big Data e Autonomia: Continuidade ou Revolução? *In:* MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (org.).

Informação, Conhecimento, Ação Autônoma e Big Data: Continuidade ou Revolução?

Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2019. p. 25-46.

DOI: <https://doi.org/10.36311/2019.978-85-7249-055-9.p25-46>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

***BIG DATA* E AUTONOMIA: CONTINUIDADE OU REVOLUÇÃO?**

*Edna Alves de Souza
Maria Eunice Quilici Gonzalez*

O cenário inusitado da contemporaneidade, caracterizado pelo crescimento exponencial do volume de dados coletados, armazenados e organizados em uma variedade de formatos digitais e a velocidade de coleta, nunca antes vistos, ilustra aspectos do que atualmente se compreende, vagamente, por *Big Data*. A utilização de dados massivos, por vezes sem a permissão dos usuários, apresenta desafios à ciência e à vida cotidiana. Um desses desafios, de especial interesse para a presente investigação, consiste na capacidade de agentes racionais se adequarem a esse cenário, mantendo-se em sintonia com ele sem prejuízo de sua autonomia.

Neste contexto, dois problemas centrais direcionam a presente investigação:

<https://doi.org/10.36311/2019.978-85-7249-055-9.p25-46>

P_1 : Estamos vivenciando uma revolução ou uma mera reforma social com o emprego de *Big Data* na ciência e na ação cotidiana?

Discutimos também possíveis implicações dos impactos dos *Big Data*, já sentidos, na ação autônoma. Interessa-nos, em particular, elaborar considerações éticas da forma com que instituições públicas e privadas, fazendo uso de técnicas de análise de *Big Data*, atingem pessoas, das mais diversas formas. Nesse contexto, o segundo problema pode ser assim formulado:

P_2 : Quais as possíveis implicações éticas da manipulação de *Big Data* na ação autônoma?

Não temos respostas definitivas, nem mesmo consensuais, para os problemas P_1 e P_2 , mas julgamos relevante a reflexão filosófica sobre eles, pois as implicações para a ciência e a vida cotidiana do que podemos chamar de Era de *Big Data* são amplas e, aparentemente, graves.

Consideramos a postura falibilista, caracterizada pela abertura à possibilidade de reformulação, ou abandono de opiniões e crenças previamente estabelecidas, diante do reconhecimento de erro ou de novidade relevante, um desiderato científico-epistemológico de grande valor prático. Nesse sentido, sugerimos que posicionamentos extremos, sejam eles pessimistas ou otimistas, em relação às implicações éticas e epistemológicas da Era de *Big Data*, sejam colocados em análise da forma mais crítica possível.

O capítulo está estruturado em duas partes. Na primeira, apresentamos uma contextualização dos *Big Data*, em termos de seu desenvolvimento histórico. Essa contextualização conduz a uma reflexão sobre a dificuldade de uma carac-

terização exaustiva sobre o tema de *Big Data*. Tal dificuldade pode ser compreendida ao considerarmos *Big Data* enquanto elemento resultante de uma imbricada relação cultural, acadêmica e científico-tecnológica. Em suas várias dimensões política, econômica, acadêmica, social e cultural, estudos sobre *Big Data* estão frequentemente vinculados às posições extremistas de otimismo exagerado ou pessimismo exacerbado. Relacionada a essa dualidade polarizada na consideração dos *Big Data* (de um lado as apostas miraculosas dos entusiastas e de outro os alertas para o extremo perigo dos críticos radicais) estão as perspectivas de sua inovação ou continuidade. Na segunda parte, refletimos sobre possíveis implicações do emprego de recursos de *Big Data* na ação autônoma. Por fim, argumentamos que o apelo aos recursos de análise de *Big Data*, mesmo que somatório e não revolucionário, no paradigma metodológico vigente, nos remetem para além de considerações epistemológicas, àquelas também de cunho ético.

1 DOS DADOS AOS *BIG DATA*

A criação de “bancos de dados” é uma prática antiga e bancos de dados digitais já existem há um bom tempo. Como Mayer-Schonberger & Cukier (2013) ressaltam, o registro de dados mais antigo parece remontar a pelo menos 8.000 a.C., quando os Sumérios usaram continhas de barro para controle de bens comercializados. A contagem em larga escala pode ser exemplificada com o censo praticado pelos antigos governos egípcios e chineses, a fim de manterem o controle governamental por meio da coleta de dados sobre a população. O surgimento do que hoje denominamos ciência ocidental (a antiga Filosofia Natural), segundo Zingano (2005), também é marcado pela coleta, organização, armazenamento e análise

de dados. A exploração, o registro e a classificação, de forma metódica, de dados sobre o mundo natural, empreendidos por Aristóteles e a fundação da Biblioteca de Alexandria, são exemplos emblemáticos da relação existente entre bancos de dados e a prática científica. A era digital, por sua vez, veio facilitar e acelerar o registro e o processamento de dados.

De acordo com Mayer-Schonberger & Cukier (2013), por volta de 1880 deflagrou-se nos Estados Unidos a necessidade de novas técnicas de processamento de dados, uma vez que a Constituição Americana prescrevia a realização do censo a cada década. Porém, dada a aceleração do crescimento populacional e a demora em se completar o censo, a informação obtida ficava obsoleta antes mesmo de se tornar disponível. Assim, eles relatam que foi preciso oito anos para que o censo de 1880 fosse concluído e a expectativa era a de que o término do próximo censo ultrapassaria dez anos, o que contrariaria, inclusive, a própria exigência legislativa. Eles ressaltam ainda que, dada essa situação problemática, o inventor americano Herman Hollerith foi contratado para usar sua ideia de cartões perfurados e máquinas de tabulação para a elaboração do censo de 1890. Ele, em um feito memorável, conseguiu diminuir o tempo de tabulação de oito anos para menos de um ano, marcando, assim, o início do processamento automatizado de dados, que serviu de base para o que mais tarde se tornou a IBM (*International Business Machines*). Mas, tal feito, como método de coleta e análise de grande quantidade de dados, ainda era lento, complexo e dispendioso.

Na chamada Era de *Big Data*, a novidade que merece maior reflexão e discussão diz respeito à crescente escala, sofisticação e onipresença da coleta e análise de dados, bem como a utilização por pesquisadores de técnicas avançadas de análise

de correlações para identificação de padrões de informação, potencialmente úteis, para subsequente inferência ou tomada de decisão. Tudo isso realizado em tempo real!

Tecnologias informacionais de comunicação (TICs), como as redes móveis e a computação em nuvem ou névoa, deram origem a mundos de informação incompreensivelmente grandes, tornando mais complexos os métodos de coleta e análise de *Big Data*. Os analistas de dados (humanos e máquinas), a partir do uso de técnicas avançadas de correlação, podem escrutinar enormes quantidades de dados, em pouco tempo, para prever comportamentos, situações e eventos de maneiras inimagináveis há bem pouco tempo. Estamos falando de dados armazenados em *terabytes*, *petabytes*, *exabytes* ou *yottabytes* (um trilhão de *terabytes*!).

Nesse cenário, definir ‘*Big Data*’ não é uma tarefa simples e parece exigir considerações interdisciplinares. São levantadas questões que oscilam sobre a origem do termo à imprecisão de seu significado, passando por aplicações distintas em áreas diversas. O rastreamento das primeiras ocorrências do termo diz pouco a seu respeito além de seu forte apelo comercial. Ibekwe-SanJuan & Bowker (2017, p. 192) observam que um analista de dados, Douglas Laney, no entanto, contribuiu decisivamente para a atual caracterização dos *Big Data* “[...] ao cunhar os ‘três Vs’ populares e cativantes dos *Big Data* (volume, variedade e velocidade)”.

O exponencial aumento de volume, velocidade e variedade de dados processados, exige um planejamento, em especial, por parte dos usuários neste novo cenário que está se instalando no século XXI. Nas palavras de Laney (2001, p.1): “O e-comércio, em particular, explodiu os desafios do geren-

ciamento de dados em três dimensões: volumes, velocidade e variedade”.

A digitabilidade, o volume, a velocidade e a variedade, na coleta e análise de dados, parecem ser os aspectos reconhecidos cientificamente como característicos dos *Big Data*, apesar da persistente imprecisão quanto à natureza dos dados. Aos três Vs iniciais, têm sido acrescentados muitos outros Vs, que não são tão distintivos como os originais, mas que os orbitam, como o V de valor e o de veracidade. Em um espírito entusiasta e de divulgação, há quem diga que “estamos agora a caminho dos 100 Vs dos *Big Data* e da Ciência dos Dados!” (SHAFER, 2017, p. 3).

Do ponto de vista físico (quantitativo), esclarecem Ibekwe-SanJuan & Bowker (2017, p. 192):

[...] os *Big Data* representam volumes de dados que os algoritmos tradicionais de bancos de dados são incapazes de lidar e que requerem infraestruturas e algoritmos de computação mais robustos e distribuídos, como *hadoop clusters*, *grid infrastructure* e *cloud clusters*.

O crescimento constante de bancos de dados, de computação ubíqua e de desenvolvimento de *softwares* capazes de, dentre outras coisas, interconectar sistemas tecnológico-informacionais e proceder à análise de dados, com recursos algorítmicos variados, parece indicar que uma “revolução silenciosa” está a caminho. A confluência de diversos fatores está tornando os *Big Data* uma força poderosa por si só, geradora de um tipo radicalmente novo de “infraestrutura de conhecimento” comum e científico. Daí o questionamento, expresso em P₁, se a Ciência que hoje faz uso de *Big Data* em suas investigações representa uma continuidade, uma convergência, de práticas

científicas tradicionais ou se está nos conduzindo a uma revolução científico-tecnológica, no sentido kuhniano.

Não é nosso objetivo discutir aqui as várias interpretações da concepção kuhniana de revolução científica, mas apenas ressaltar um de seus sentidos, qual seja, que após uma revolução, um novo paradigma surge sob o qual se assentará a pesquisa científica futura. Nesse sentido, cabe questionar se a sociedade contemporânea está passando por uma revolução, que reconfigura hábitos e relações, desde as mais básicas, existentes entre seus elementos, dando lugar a um novo modelo de organização (que se poderia chamar de universo da cultura digital), ou se está passando apenas por uma reforma, com a adição de um novo espaço público, da realidade virtual, que, nem sempre (ou quase nunca) está em sintonia com a realidade em seu sentido tradicional, ou seja, o mundo natural.

A literatura sobre os *Big Data* é permeada de exemplos, como os apresentados por Mayer-Schonberger & Cukier (2013), que pretendem, dentre outras coisas, ilustrar o aspecto revolucionário ocasionado por essa nova forma de lidar com informação disponível através de computação ubíqua, que atua ininterruptamente na coleta de dados sobre o cotidiano das pessoas, através de celulares, cartões de crédito, câmeras, entre outros. Vejamos alguns deles: (a) Google, por meio dos seus mecanismos de pesquisa, considerando o momento e a localização das buscas de usuários, consegue prever fenômenos, como surtos de violência e tendências de desemprego, antes mesmo que os resultados estatísticos de pesquisas oficiais sejam oficialmente divulgados pelo Estado. (b) Técnicos de empresas de cartões de crédito registram e analisam vastas quantidades de dados com informação sobre hábitos e ações financeiras pessoais para tentar detectar fraudes e identificar

tendências de compra do consumidor. (c) Pesquisadores da área médica analisam os registros de saúde de milhares de pessoas para tentar identificar correlações úteis entre tratamentos médicos e resultados de saúde. (d) Empresas realizam estudos de “mineração de dados” sobre informações pessoais em sites de redes sociais, na tentativa de identificar preferências sutis do consumidor e criar estratégias eficientes de *marketing*. Os dados de “geolocalização” permitem que instituições públicas e privadas, como grandes empresas, analisem dados de dispositivos móveis para fazer inferências intrigantes sobre a vida das pessoas e a economia: dados coletados do rastreamento da localização de celulares – por exemplo, a periodicidade com que as pessoas vão aos *shoppings* – pode ser um indicador econômico para medir a demanda dos consumidores.

As situações indicadas acima são amplamente comentadas por Mayer-Schonberger & Cukier que, não por acaso, intitulam a sua obra *Big Data: A revolution that will transform how we live, work, and think*. Elas ilustram aspectos impactantes dos *Big Data* para questões da vida em sociedade. O alegado caráter revolucionário acarretado pelos *Big Data* na vida cotidiana, e também na ciência, estaria na subjacente metodologia adotada e desenvolvida pelos estudiosos de *Big Data*, ou seja, o foco implícito nos exemplos acima: o trato, já em curso, dos dados a partir da identificação de *correlações* consideradas relevantes para a análise e previsão do fluxo de eventos no mundo.

Independente de os *Big Data* representarem uma revolução ou uma continuidade (com algumas novidades impactantes) para atividades científicas e cotidianas, concordamos com Bollier (2010, p. 1) que suas “[...] implicações para os negócios, o governo, a democracia e a cultura são enormes”. Daí

a dificuldade de separar caracterizações de *Big Data* de seus elementos valorativos mais amplos que, comumente, conduzem às posturas dicotômicas, conforme veremos a seguir.

Para além da perspectiva meramente quantitativa, as técnicas de análise de *Big Data* podem ser vistas, em uma perspectiva metodológica, como um conjunto de métodos de análise que incorpora uma diversidade de conhecimentos, técnicas de programação e tecnologias. Desse modo, Boyd & Crawford (2012, p. 663) caracterizam as suas diferentes dimensões cultural, tecnológica e acadêmica que, em suas palavras, “[...] repousam na interação de tecnologia, análise e mitologia”. Elas sugerem que a tecnologia desenvolvida, visando a aplicação na análise de *Big Data*, permite maximizar o poder de computação e de precisão algorítmica para coletar, analisar, vincular e comparar grandes conjuntos de dados. A análise, no contexto dos *Big Data*, possibilita a extração de padrões em grandes conjuntos de dados, a fim de se fazer intervenções econômicas, sociais, técnicas e legais. A mitologia, por sua vez, consistiria na crença generalizada de que os *Big Data* oferecem uma forma superior de inteligência e conhecimento, com a aura de verdade, objetividade e precisão, que pode gerar percepções antes impossíveis.

Em síntese, indicamos até aqui concepções de *Big Data* que enfatizam o seu papel impactante na contemporaneidade. Nossa tendência, entretanto, é admitir que sem uma clara definição atual de seu escopo, os recursos de *Big Data* promovem apenas uma reforma, antes que uma revolução, nos modos de organização de hábitos individuais e coletivos existentes a partir de uma longa história evolutiva humana. Sem dúvida, técnicas de análise e de organização de dados podem auxiliar a pesquisa literária, bibliográfica, política, acadêmica, ao otimizar a reu-

nião e o cruzamento de dados massivos, em diferentes épocas e contextos; indivíduos e grupos podem aumentar suas tendências consumistas e exacerbar suas tendências agressivas com os recursos das TICs. Mas é principalmente na ciência que os *Big Data* estão trazendo novidades consideradas revolucionárias.

Anderson (2008), por exemplo, decreta o fim da teoria científica tradicional, ao afirmar, no subtítulo de seu artigo, que “o dilúvio de dados torna o método científico obsoleto”. Para ele, em uma era de computação em nuvem e conjuntos de dados massivos, o verdadeiro desafio não é criar novas taxonomias ou modelos científicos, mas peneirar os dados de novas maneiras para encontrar correlações significativas.

Em contraste com a visão otimista *do* uso de *Big Data* na ciência, críticos dos *Big Data* insistem que esse instrumento de manipulação de dados massivos poderia ser visto, no máximo, como um complemento possível e não uma alternativa ou um substituto da investigação científica tradicional. Embora técnicas de *Big Data* sejam boas para detectar correlações, não são elas, por si só, que especificam quais correlações são significativas e o porquê de o serem. Ilustrativamente, o site *Spurious Correlations*, de Tyler Vigen (2015), ficou conhecido por trazer uma série de correlações espúrias engraçadas, mas com o propósito de denunciar o uso irresponsável de correlações estatísticas. Um programa computacional que registra correlações em grandes bases de dados foi desenvolvido, com resultados, no mínimo provocantes, que relacionam, por exemplo, a redução do consumo per capita de margarina à diminuição de divórcios.

Radicalismos expressos por Anderson (2008) e outros entusiastas e promotores dos *Big Data*, bem como, em outro extremo, por Vigen (2015) e demais críticos aguerridos dos *Big Data*, indicam uma tendência na literatura corrente sobre

o tema: perspectivas de inovação ou continuidade na consideração de *Big Data* parecem estar relacionadas à dualidade polarizada – de um lado o otimismo exagerado de seus entusiastas e de outro o pessimismo exacerbado de seus críticos radicais.

Consideramos que o problema da produção e da organização acelerada de dados massivos e do uso de técnicas de análise de *Big Data* na investigação científica, enquanto envolvimento em questões e análises do tipo “sociotecnológicas”, pode conduzir a retóricas utópicas ou distópicas (de acordo com as propensões dos envolvidos), que se confundem com a própria caracterização do que é entendido por *Big Data*.

Além disso, talvez uma das chaves para o entendimento de possíveis efeitos dos *Big Data* na Ciência, bem como na ação autônoma, esteja na análise do conceito de dado. Essa análise da dinâmica do conceito de dado parece ter o potencial de revelar mudanças sutis na ação coletiva que estão ocorrendo no, ou com o, uso de *Big Data*, não sendo obscurecido de antemão pelos vieses utópicos ou distópicos de questões sociotecnológicas.¹

Na próxima seção refletimos a respeito de algumas possíveis implicações do emprego de recursos de *Big Data* na ação autônoma.

2 INFORMAÇÃO, *BIG DATA* E GRAUS DE AUTONOMIA DA AÇÃO

Em seus instigantes estudos sobre a natureza da informação, Gregory Bateson (1972, p. 460) a caracteriza como: “[...] *uma diferença que faz uma diferença*”. Essa aparente enig-

¹ Essa hipótese é apresentada e discutida em nosso artigo *Big Data: Philosophy of Science Contemporary Issues* (no prelo).

mática caracterização expressa um dos pilares da chamada Escola de Palo Alto, que inaugurou a *Sistêmica Qualitativa*. Em contraste com os pressupostos da teoria matemática da comunicação (SHANNON & WEAVER, 1949) – que propiciou, entre outros, subsídios formais para o desenvolvimento de métodos analíticos de *Big Data* –, ênfase é dada à abordagem sistêmica ao estudo do *significado* presente na comunicação.

A *Sistêmica Qualitativa* começou a se estruturar na década de 1950, quando Bateson desenvolve um estudo interdisciplinar sobre a natureza da informação significativa, com a ajuda de filósofos, psicólogos, antropólogos, biólogos, neurocientistas, entre outros (destacando-se Birdwhistell, Goffman, Hall, Jackson, Scheffen, Sigman, Watzlawick). O que interessa à escola de Palo Alto é o entendimento da dinâmica constitutiva de *padrões informacionais* prenhes de significado (que não se reduzem a enormes conjuntos de dados) e de seus possíveis efeitos na ação de organismos contextualmente situados.

Inspirado em hipóteses da Cibernética, em especial na noção de *feedback circular* (WIENER, 1950), e na Teoria de Tipos Lógicos (WHITEHEAD & RUSSELL, 1910), Bateson (1972) sugere uma ponte explicativa entre o que ele denomina *fatos da vida e do comportamento* e a *dinâmica estrutural* constitutiva de padrões de ordem diversas que “ligam” seres vivos, em várias escalas, possibilitando a comunicação de *informação significativa*.

Em sua formulação mais geral sobre a natureza da informação, Bateson argumenta que *informação significativa*, presente em inúmeras formas de comunicação entre seres vivos, emerge de relações *qualitativas* (e não apenas numéricas) estabelecidas entre seres vivos e meio ambiente. A complexa dinâmica de relações contextualmente situada engendraria ti-

pos de *padrões informacionais*, estruturados em várias escalas. Tais padrões, por sua vez, emergem de ações habituais comuns aos organismos que compartilham, não apenas vínculos informacionais e emocionais, mas também aspectos de estruturas biológicas. Bateson (1979) menciona, por exemplo, o padrão que liga funcionalmente as pinças dos caranguejos e os dedos polegar e indicador nas mãos de humanos; o padrão constituído de dois olhos e nariz no centro do rosto, que liga seres humanos a outros animais, como elefantes, macacos e felinos em geral, permitindo um tipo de vínculo fundamental entre diferentes espécies. Uma vez que a comunicação significativa extrapola o domínio das estruturas linguísticas, estas serão analisadas como um caso particular de tipos de padrões informacionais no domínio da pragmática humana.

Com o advento da linguagem proposicional humana, estruturas dinâmicas, ordenadas em redes de padrões informacionais, possibilitaram a emergência de inúmeros planos de comunicação, gerando *disposições* para a ação de agentes morais, com diferentes graus de autonomia. Nesse contexto informacional, diferenças (entre elementos constitutivos de padrões informacionais) podem fazer diferença (não apenas para a comunicação verbal, mas também no plano da ação individual e coletiva).

Um pressuposto basilar da *Sistêmica Qualitativa*, de especial interesse para nós, é a conhecida hipótese batesoniana do *duplo-vínculo* (*double-bind*), segundo a qual relações de dependência entre seres vivos se expressam através de padrões informacionais, muitas vezes contraditórios e conflitantes, que condicionam, por meio da repetição, suas ações. Na Psicologia, o *duplo-vínculo* é caracterizado como: “[...] uma situação na qual, não importa o que uma pessoa faça, ela ‘não

pode vencer'. [...] uma pessoa aprisionada no duplo vínculo pode desenvolver sintomas esquizofrênicos” BATESON et al., 1956, p. 251). É a partir desse pressuposto que investigamos o problema P_2 , sobre possíveis implicações da manipulação de *Big Data* na ação autônoma, focalizando os duplos vínculos que parecem estar sendo estabelecidos na contemporaneidade entre usuários das TICs.

De acordo com Anderson (2008, p. 1), a “[...] era do *Petabyte* é diferente porque *mais* é diferente”. Concordamos que “mais” pode fazer diferença quando se trata de acúmulos de padrões de dados correlacionados. Contudo, resta explicitar o papel da diferença que envolve graus de autonomia.

Entendemos que a ação de uma pessoa (P), ou grupo de pessoas (GP), expressa certo *grau* de autonomia se pelo menos duas condições forem satisfeitas: (1) P (ou GP) possui informação antecipadora de oportunidades de escolha, e (2) P (ou GP) possui capacidade de escolha, independente de pressão externa. Satisfeitas (1) e (2), consideramos que a ação de P (ou de GP) será tanto mais autônoma quanto mais independente de coação, ou imposição externa, for a sua escolha, e quanto mais informação, relacionada a cada escolha, estiver ao seu alcance.²

No que concerne aos graus de autonomia individual e coletiva, nossa hipótese é que vivenciamos atualmente uma relação de “duplo vínculo”. Por um lado, as TICs propiciam, no plano individual, inúmeras facilidades; o acesso atraente e eficiente à informação enciclopédica disponível na *Internet*, por exemplo, libera os indivíduos da penosa necessidade de memorizar dados necessários à resolução de problemas. Elas

² Uma análise mais detalhada do conceito de ação autônoma pode ser encontrada em Gonzalez (2017).

também possibilitam a desejada comunicação, quase imediata, entre pessoas de várias culturas, idades e gêneros. Por outro lado, vivenciamos a invasão de privacidade, de *bullying*, e uma uniformidade cansativa, pouco criativa, de comportamentos no ambiente digital, promovida pela ubiquidade das TICs que tendem a acelerar e padronizar as atividades cotidianas, sobrecarregando os indivíduos com massiva quantidade de informação alienante. É nesse contexto que recursos de análise de *Big Data* permitem que empresas como a Amazon e a Netflix, por exemplo, “criem necessidades” e antecipem os desejos de seus clientes, que nem sempre possuem condições econômicas de compra, comprometendo os graus de autonomia de suas ações.

Agências governamentais e empresariais, como a Google, Facebook, Instagram, estão descobrindo rapidamente usos estratégicos de algoritmos, estruturadores de grandes bancos de dados. Longe de serem neutros, esses algoritmos são empregados também para detectar padrões de ação, laços familiares, sonhos e medos, preferências de compra, entre outros, registrados em traços deixados por usuários das TICs, direcionando, muitas vezes de forma invisível, a dinâmica de formação de opinião dos indivíduos. Aprisionados em um duplo vínculo de “amor e ódio”, muitos indivíduos sentem-se obrigados a aceitar as condições impostas por essas agências para que possam usufruir de seus benefícios.

No plano coletivo, algumas das consequências práticas, positivas, da aplicação de resultados de análises de *Big Data*, dizem respeito, de um lado, à possibilidade de uso, desejável, de correlações de dados para prever e prevenir, idealmente, desastres ecológicos, bem como ações sociais indesejáveis relacionadas à segurança e aos direitos humanos. Por outro lado,

já vivenciamos, em várias sociedades, a crescente e conflituosa polaridade na dinâmica de opiniões em questões de gênero, de política, dentre outras. Estimulados pela influência das mídias digitais, grupos de amigos e familiares estão se desintegrando no confronto de opiniões muitas vezes apoiadas em notícias falsas. Novas formas de vigilância também já estão sendo aplicadas envolvendo algoritmos de rastreamento para extrair informações a partir de traços deixados por indivíduos e grupos sociais. Vivenciamos, também, aplicações de algoritmos manipuladores de grande quantidade de dados na robótica, em especial nas áreas bélicas, que vêm colocando em risco a autonomia da ação humana coletiva. O desconforto resultante desse “duplo vínculo” no emprego de técnicas de análise de *Big Data* – que por um lado otimista trazem benefícios atraentes aos usuários das TICs, mas por outro também os asfixiam – parece estar gerando uma situação cotidiana de “esquizofrenia” individual e coletiva.

No contexto científico, o emprego de algoritmos manipuladores de *Big Data* facilita a coleta de dados, que atualmente seriam impossíveis de serem coletados sem o subsídio desse tipo de recurso. Contudo, o emprego por vezes pouco crítico de técnicas de análise mecânica e de algoritmos que privilegiam a ocorrência de correlações, com prejuízo para a analiticidade da metodologia científica tradicional, indica uma possível revolução silenciosa nos modos de se fazer ciência. De forma ilustrativa, como explicam Santana e Fernandes (2016), na astronomia, o pesquisador recentemente tem passado de recursos do telescópio para os de *Big Data*; ou seja, mais do que perscrutar os céus, de forma mais ou menos direta e objetiva, com equipamentos como o telescópio de mão, em busca de algo novo que explique algum fenômeno ou ofereça resposta a algum problema, hoje, o astrônomo, com conhecimentos em estatística e programação,

emprega recursos de programas computacionais para minerar dados coletados por equipamentos de alta tecnologia, como as sondas espaciais. Os telescópios agora têm suas funções estendidas, pois são equipados com sensores que coletam, constantemente, enorme quantidade de dados o que torna praticamente impossível exaurir a capacidade humana de extração de conhecimento a partir dos dados coletados. Os dados permanecem valiosos não só para a Astronomia como para a Ciência em geral, devido a seu potencial não explorado, ensejando sua reutilização para outras finalidades. Atualmente, a teorização, por exemplo, sobre bilhões de galáxias, a partir da análise de *petabytes* de dados, exige infraestruturas e algoritmos de computação sofisticados. Contudo, é preciso saber que critérios de relevância são empregados para procurar o que se quer descobrir. Uma possibilidade seria observar tudo (o máximo possível), levantando uma quantidade massiva de dados, e tornar explícitos os critérios de relevância empregados para minerar esses dados em busca de correlações significativas para um determinado propósito. Ao se perscrutar “tudo”, corre-se o risco de se perder, em um primeiro momento, a delimitação do problema, do objeto de pesquisa, no sentido tradicional dos termos, e; em um segundo momento, ao ter no banco de dados o contexto da investigação, de se distanciar do contexto original da realidade.

O cenário acima esboçado requer reflexão epistemológica, metodológica e ética, e em certos casos intervenção, sobre possíveis perigos do emprego de recursos analíticos de *Big Data* no trato não só de imensa quantidade de dados, mas também de pessoas. Acreditamos que tal reflexão, se realizada a partir de uma perspectiva da sistêmica qualitativa, pode abrir horizontes para o desenvolvimento de uma ética que respeite a ação autônoma no exercício de maximização de interesses comuns.

CONCLUSÕES PROVISÓRIAS

Iniciamos este capítulo com a formulação do problema P_1 : Estamos vivenciando uma revolução ou uma mera reforma com o emprego de *Big Data* na ciência e na ação cotidiana? Ressaltamos que a Era de *Big Data* não consiste, simplesmente, no abandono da analiticidade, da metodologia científica tradicional, mas na articulação dessas, indicando um novo modo de se fazer ciência e de se agir no cotidiano, a partir de padrões que podem ser identificados por meio de algoritmos rastreadores de grandes volumes de dados.

O problema P_1 nos remete à consideração de que, uma vez que agora dispomos de conjuntos massivos de dados e técnicas avançadas de manipulação dos mesmos, restaria ainda lugar para a metodologia científica tradicional, que visa a busca de explicações para a ocorrência dos fatos. Tradicionalmente, as explicações científicas estão fortemente apoiadas na noção de causalidade e não apenas em massivas correlações de dados detectadas por algoritmos, cuja estrutura pode depender de interesses, em geral comerciais. Embora Pietsch (2013; 2016) tenha sugerido, entusiasticamente, que a consideração de quantidade massiva de dados aumenta a chance de se identificar por meio de algoritmos de *Big Data* relações causais legítimas (e não meramente correlações), não estamos seguras do valor epistemológico dessa hipótese. Dentre outros problemas, o autor parece incorrer em uma falácia: a da petição de princípio, ou seja, assume de antemão que, no conjunto de dados massivos coletados com recursos de *Big Data*, há relações causais que os algoritmos de análise de *Big Data* seriam capazes de explorar e identificar (que é justamente o que se desejaria descobrir ou demonstrar).³

³ Contribuímos para com o debate sobre correlação *versus* relação causal no já mencionado artigo *Big Data: Philosophy of Science Contemporary Issues* (no prelo).

Nas ciências exatas e biológicas, as contribuições de recursos de análise de *Big Data* parecem mais se somar à prática tradicional do que a revolucionar, em um sentido kuhniano de uma alteração radical e profunda em sua estrutura metodológica. Nas ciências humanas, o emprego otimista de técnicas de *Big Data* ainda encontra resistências (justificadas) quanto ao valor de cálculos estatísticos e de correlações descontextualizadas na análise de problemas, cuja compreensão requer informação de subsídios socioculturais, contextualmente situados.

Em ambos os casos, o ideal de inter/multi e transdisciplinaridade na pesquisa enfrenta dificuldades no estudo de problemas cuja solução demanda um diálogo ético, francamente cooperativo, entre pesquisadores que valorizem diferenças ambientais, biológicas, contextuais e culturais. Nesse sentido, considerações sobre a possibilidade de uma revolução científica silenciosa, que poderia estar a caminho na era de *Big Data*, requerem a colaboração filosófico-interdisciplinar voltada à atenção de suas consequências epistemológicas e éticas para as futuras gerações. Metaforicamente, poderíamos dizer que se, atualmente, na ciência, o que é procurado consiste em uma agulha digital perdida em um palheiro de *Big Data*, então, persiste pelo menos a necessidade do cuidado para evitar acidentes e erros, nessa tarefa, que possam trazer consequências indesejáveis.

No que diz respeito à vida cotidiana, embora o apelo aos recursos de *Big Data* possa ter efeitos considerados impactantes, reconfigurando hábitos, comportamentos e entendimentos de nós mesmos e da realidade que nos cerca, acreditamos que uma revolução ainda não ocorreu, não obstante o ufanismo de seus promotores. A herança da história evolutiva humana, geradora de hábitos e disposições de longa duração,

parece sustentar alguns dos passos da reforma de padrões informacionais, aparentemente em curso, na era de *Big Data*.

Julgamos que a escolha de critérios de relevância aplicados na elaboração de algoritmos, não neutros, para a seleção de correlações a serem privilegiadas em situações e estudos específicos, quando desvinculada de considerações éticas pode acarretar resultados indesejáveis a curto e longo prazo. Nesse contexto, sugerimos uma reflexão a partir da *Sistêmica Qualitativa*, destacando, dentre os pressupostos dessa abordagem, a *hipótese Batesoniana do duplo-vínculo*, sobre a dinâmica de relações de dependência que se expressam por meio de padrões informacionais ambíguos, propiciadores de situações conflituosas.

Este é o caso do problema P_2 – sobre possíveis implicações éticas da manipulação de *Big Data* na ação autônoma –, para o qual sugerimos uma resposta não muito otimista, na perspectiva falibilista. Contudo, apenas o tempo histórico revelará, com mais propriedade, se de fato vivenciamos uma revolução ou uma mera reforma social com o emprego de *Big Data* na ciência e na ação cotidiana.

REFERÊNCIAS

ANDERSON, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*, June 23, 2008. Disponível em: <https://www.wired.com/2008/06/pb-theory/>. Acesso em: 30/09/2018.

BATESON, G. *Steps to an Ecology of Mind*. Collected essays in anthropology, psychiatry, evolution, and epistemology. London: Jason Aronson Inc., 1972.

BATESON, G. *Mind and nature*. A Necessary Unity. New York: E. P. Dutton, 1979.

BATESON, G.; JACKSON, D.; HALEY, J. & WEAKLAND, J. Toward a theory of schizophrenia. *Behavioral Science* [1956] 1(4): 251-264.

BOLLIER, D. *The Promise and Peril of Big Data*. Washington: The Aspen Institute, 2010. (Report)

BOYD, D. & CRAWFORD, K. Critical Questions For Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* Vol. 15, No. 5, June 2012, pp. 662–679.

GONZALEZ, M. E. Q. Autonomous action in complex mechanical systems: A real dilemma? In: Adams, F.; Pessoa Jr., O. & Kogler Jr., J.E. (eds.) *Cognitive science: recent advances and recurring problems*. Wilmington (DE): Vernon Press, v. 1, p. 17-30, 2017.

IBEKWE-SANJUAN, F. & BOWKER, G. C. Implications of Big Data for Knowledge Organization. *Knowl. Org.* 44 (2017).

LANEY, D. 3D Data Management Controlling Data Volume Velocity and Variety. 2001. Disponível em: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Acesso em: 05/09/2018.

MAYER-SCHONBERGER, V. & CUKIER, K. *Big Data. A Revolution That Will Transform How We Live, Work and Think*. Houghton Mifflin Harcourt, 2013.

PIETSCH, W. *Big Data – The New Science of Complexity*. 2013. Disponível em: <http://philsci-archive.pitt.edu/9944/>. Acesso em: 24/06/2019.

PIETSCH, W. The Causal Nature of Modeling with Big Data. *Philos. Technol.* (2016) 29: 137. pp 137–171. Disponível em: <https://doi.org/10.1007/s13347-015-0202-2>. Acesso em: 24/06/2019.

SANTANA, E. S.; FERNANDES, I. F. Dados Astronômicos: uma proposta de implementação para banco de dados. *Caderno de Física da UEFS*. 16 (02): 2402.1-20 2018. Disponível em: <http://dfisweb.uefs.br/caderno/vol16n2/s4-a2-Dados%20Astronomicos.pdf>. Acesso em: 24/06/2019.

SHAFER, T. The 42 V's of Big Data and Data Science. 2017. Disponível em: <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>. Acesso em: 05/09/2018.

SHANNON, C. E.; WEAVER, Warren. [1949] *A mathematical theory of communication*. Urbana: University of Illinois Press, 1998.

SOUZA, E. A.; GONZALES, M. E. Q.; VICENTINI, M. R. *Big Data*, Pós-verdade e Novo Construtivismo: ainda há um lugar para o realismo? *Semeiosis – Semiótica e Transdisciplinaridade em Revista*, São Paulo, vol. 10, 2019 (no prelo).

SOUZA, E. A.; GONZALES, M. E. Q.; VICENTINI, M. R. *Big Data*: Philosophy of Science Contemporary Issues. 2019 (no prelo).

VIGEN, Tyler. Spurious correlations. New York: Hachette Books, 2015. *Spurious Correlations*. Disponível em: <http://tylervigen.com/spurious-correlations>. Acesso em 03/09/2018.

WHITEHEAD, A. N. & RUSSELL, B. [1910]. *Principia Mathematica*. Cambridge: Cambridge University Press, 1963.

WIENER, N. *Cibernética e sociedade*. O uso humano de seres humanos. Trad. de José Paulo Paes. São Paulo: Cultrix, 1950.

ZINGANO, M. *Platão & Aristóteles*. O fascínio da Filosofia. 2. ed. São Paulo: Odysseus, 2005. (Imortais da Ciência)