Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Using big data to solve real problems through academic and industry partnerships

Stephen R Mitroff[1] and Benjamin Sharpe[2]

Big data has revolutionized a number of industries as it provides a powerful tool for asking and answering questions in novel ways. Academic researchers can join this trend and use immense and complex datasets to explore previously intractable questions. Yet, accessing and analyzing big data can be difficult. The goal of this chapter is to outline various benefits and challenges of using big data for academic purposes, and to provide thoughts on how to succeed. The primary suggestion is for academics to collaborate with appropriate industry partners to simultaneously achieve both theoretical and practical advances.

Addresses
[1] Department of Psychology, The George Washington University, United States
[2] Kedlin Company, United States

Corresponding author: Mitroff, Stephen R (mitroff@gwu.edu), Mitroff, Stephen R (mitroff@gwu.edu)

Current Opinion in Behavioral Sciences 2017, **18**:91–96

This review comes from a themed issue on **Big data in the behavioural sciences**

Edited by **Michal Kosinski** and **Tara Behrend**

http://dx.doi.org/10.1016/j.cobeha.2017.09.013

2352-1546/© 2017 Elsevier Ltd. All rights reserved.

To understand the future impact of big data projects that are driven by academic–industry partnerships, experts from a variety of communities were asked to answer the following question:

"*Looking forward to the next 5 to 10 years, how do you think big data might be able to help you do your job better*?"

Their answers are provided throughout this chapter and gathered together in Box 1.

Introduction

The term 'big data' has become ubiquitous, and it generally refers to very large and complex datasets that can be analyzed to reveal hidden patterns and insights. From an academic perspective, big data holds promise for completely revolutionizing entire fields by unlocking new avenues of research with more power than ever before. While it is clear that academics can use big data to solve real problems, there are many hurdles that can thwart well-intended plans. For example, one must capture the data in a useable format, have the necessary skill and resources to process the data, and have the appropriate industry and/or government connections to make the analyses practically useful. Such hurdles can be significant hindrances, yet overcoming them can have dramatic and meaningful results.

The current goal is to discuss how big data can be used to advance both science and applied practices (see Figure 1). We start with our own case study of ***Gamifying Airport Security*** to highlight the promises and struggles of our approach to leverage big data from a game to further science and improve real world airport security.

> "The future of scientific analysis will be dramatically changed by big data. The traditional approach is to ask a question and then collect data to answer it. Big data does the reverse. You gather data without a purpose, play with the data and get a sense of what it can tell you. The traditional bases of soccer competition are eroding fast; Organisations have more and more data on hand but they have far less room for errors in execution, so decision-making has to be sharper and better informed. All in all, these factors call for superior analytics and deeper insights into what makes an organisation work. Analytics will involve the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions."
> **Tony Strudwick**
> **Director of Performance, Manchester United**

Gamifying Airport Security: using mobile data to advance science and aviation security

The two authors of this chapter are a seemingly unlikely pair — a cognitive psychology professor (Mitroff) and a mobile app developer (Sharpe) — however it is exactly this unlikely combination of skillsets that has allowed the formation of a successful partnership. Dr. Mitroff studies aspects of visual search — how individuals find targets amongst distractors. Mr. Sharpe has developed a number of mobile apps, one of which is *Airport Scanner (Kedlin Company,* www.airportscannergame.com*)* — a game wherein the player serves as an airport security officer

Box 1 The future impact of big data across industries.

To understand the future impact of big data projects that are driven by academic–industry partnerships, experts from a variety of communities were asked to answer the following question:

“***Looking forward to the next 5 to 10 years, how do you think big data might be able to help you do your job better***?”

“The future of scientific analysis will be dramatically changed by big data. The traditional approach is to ask a question and then collect data to answer it. Big data does the reverse. You gather data without a purpose, play with the data and get a sense of what it can tell you. The traditional bases of soccer competition are eroding fast; Organisations have more and more data on hand but they have far less room for errors in execution, so decision-making has to be sharper and better informed. All in all, these factors call for superior analytics and deeper insights into what makes an organisation work. Analytics will involve the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.”

Tony Strudwick
Director of Performance, Manchester United

“Continuing on this path of collaboration with “unlikely” partners provides us a new way of solving problems — as we move forward and have enormous amounts of data available to us, we need people who understand how to use these data to solve emerging problems. Dr. Mitroff and Sharpe have the uncanny ability to insightfully explore vast amounts of data to offer astute, remarkable recommendations and solutions.”

Bonnie Kudrick
Program Manager, US Transportation Security Administration

“As performance specialists, therapists, applied sport scientists, nutritionists, etc., we must listen to the voice of the athletes we train (and those that we don’t, but that we can learn from listening to). Said best by Matt Nurse at Nike NSRL, one key voice to listen to is their biometric data. This is a voice that tells a story at a depth specificity that no verbalization or visualization can provide. By itself, data are useless. But added to the other voices of the athlete (subjective and objective) we have the best ability to provide impactful solutions and management guidance to reach full athletic potential. Sometimes these data speak instantly, and other times retrospectively. Either way, without it we are truly flying the high performance plane blind.”

Lance Walker, MS, PT
Director, Michael Johnson Performance

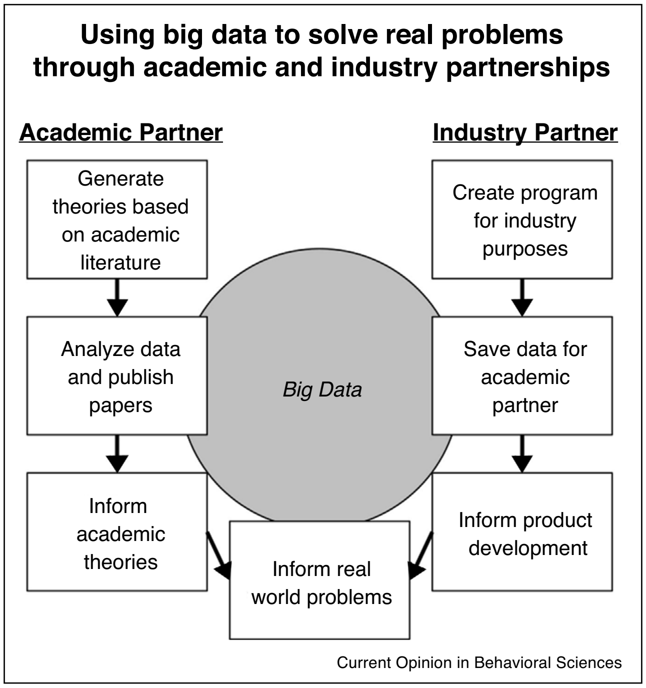
“Advanced Analytics and Big Data are the motors of digital innovation, disruption and modern business transformations. Successfully competing in business is quickly becoming an entirely different ballgame: Winning requires mastery of data, of creating original insights, and of navigating transformative change when applying such insights to large organizations. My clients are using Big Data to improve their decision making, their processes, products and services. They develop new offerings, and the best find ways to out-innovate their competition and disrupt their respective industries. As a Bain Partner and facilitator of business transformations these are truly exciting and energizing times. We are leaders in advanced analytics and are investing heavily into Bain’s talent, our tools and techniques as we harness the power of Big Data to help our clients create such high levels of economic value that together we set new standards of excellence in our respective industries.”

Rasmus Wegener, Ph.D.
Partner, Bain & Company

“Big data is already reshaping the field of cancer control. The great promise over the next 5–10 years is the ability to get intensive longitudinal data on individuals. Anyone with a smartphone will be able to easily volunteer to contribute data on their cognitive function, activity levels, sleep habits, and a wide range of other behavioral variables. This will allow us to break the logjam on many outstanding problems in cancer prevention and control.”

Todd Horowitz, Ph.D.
Program Director, National Cancer Institute, National Institutes of Health

Figure 1


Using big data to solve real problems through academic and industry partnerships
Academic Partner
Industry Partner
Generate theories based on academic literature
Analyze data and publish papers
Inform academic theories
Big Data
Inform real world problems
Create program for industry purposes
Save data for academic partner
Inform product development
Current Opinion in Behavioral Sciences

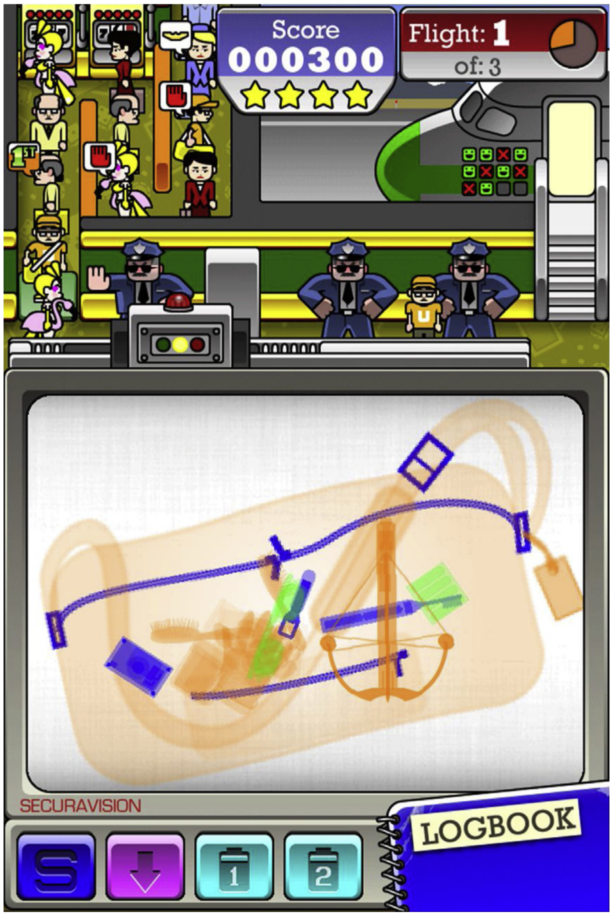
Academic–industry partnership. Centered on big data, academic and industry partners can pursue parallel path to advance academic theory, industry goals, and solve real-world problems.

looking for prohibited items in passenger bags (i.e. looking for targets amongst distractors).

In 2012, Dr. Mitroff had been studying the nature of visual search from a number of perspectives, and was actively working with the US Department of Homeland Security to explore how professional airport security officers search for threats. At this time, *Airport Scanner* became a highly success game, reaching #1 in the Apple App Store, resulting in millions of downloads worldwide. Dr. Mitroff came across the game and realized its potential for research purposes and emailed Mr. Sharpe hopeful for a collaboration. The two connected by video chat (they were on opposite sides of the US) and by the end of their first discussion, Mr. Sharpe saw the potential in forming a partnership and jumped at the opportunity. Specifically, Mr. Sharpe was enticed by three opportunities: firstly, that a game meant for entertainment could have an altruistic purpose to improve airport security, secondly, an opportunity to boost downloads from marketing and public relations from the app being used for research, and thirdly, the potential for future opportunities to expand the partnership and potentially further commercialize the product.

Several steps were taken to establish a formal partnership; a data sharing agreement was established between the university and company, the app was modified to collect and store the data, a web-based interface was created for

Figure 2



Current Opinion in Behavioral Sciences

Sample bag (trial) from Airport Scanner. Users tap with their finger on prohibited items (e.g. the crossbow in this example) they notice. All aspects of the trial are saved, including the location, identity, and rotation of each prohibited and allowed item in the bag.

the university team to efficiently access and analyze the data, and approval was obtained to use the data for research purposes. The data started rolling in and have yet to stop — as of September 2017, more than 13 million players have collectively contributed over 3 billion trials. This is the largest visual search dataset available for research, and can be used to reveal novel facts about search (e.g. [1]) and to better understand the workings of research in general (e.g. [2•]).

The sheer size of the *Airport Scanner* dataset is impressive, but what makes it 'big data' [3] is the intricacy of the saved information — each trial consists of a single 'bag' (see Figure 2) and data are saved about all aspects of the trial (target identity, distractor item identities, item locations, time to respond, response accuracy, etc.). Each trial contains a complex set of data, collectively providing more than *250 trillion pieces of data*. This partnership has been successful, producing multiple academic papers (e.g. [1,4,5,6•,7••]), over a dozen academic presentations, and Dr. Mitroff has obtained grant funding to use the data for research purposes from the US Army Research Office and the US Transportation Security Administration (TSA).

> "Continuing on this path of collaboration with "unlikely" partners provides us a new way of solving problems – as we move forward and have enormous amounts of data available to us, we need people who understand how to use these data to solve emerging problems. Dr. Mitroff and Sharpe have the uncanny ability to insightfully explore vast amounts of data to offer astute, remarkable recommendations and solutions."
>
> **Bonnie Kudrick**
> **Program Manager, US Transportation Security Administration**

Building on the academic progress, Dr. Mitroff and Mr. Sharpe have used this partnership to directly inform real-world operations. Kedlin Company received a contract from the TSA to use a modified version of Airport Scanner to assess performance of TSA Officers. Through this contract, Dr. Mitroff and Mr. Sharpe were able to demonstrate that their new tool, *XRAY Screener*, could reliably identify professional screeners who were more accurate and quicker when working behind a real XRAY machine at the actual airport passenger checkpoint. This success led to the creation of a new company that hopes to help inform personnel selection, assessment, and training for a wide variety of organizations that depend on visual search performance.

We briefly reviewed our personal case study here with the hope of highlighting how this sort of partnership endeavor can benefit academic researchers (novel datasets, papers, grants, etc.), industry (publicity, helping science, contracts, informing product development, etc.), and the broader public interest (here using already created data to inform real world operations to the benefit of millions of air travelers). In the remainder of this chapter we provide broader suggestions for navigating such endeavors.

Key issues for using big data in an academic–industry partnership

Finding the data

The first major hurdle in using big data in an academic–industry partnership is identifying an appropriate dataset. Fortunately, the past decade has given rise to a number of technological advances that provide ideal platforms for data collection. Some platforms gather data as a side effect of the technology's primary goal (e.g. the *Airport Scanner* dataset was a by-product of people playing and enjoying the game). Other platforms gather data as the core function of the technology's purpose (e.g. exercise trackers gather data to provide feedback to the user). An academic looking to access big data has multiple options, and we highlight a few here.

First, some researchers have collected data via apps created specifically for research purposes. For example, Sea Hero Quest (www.seaheroquest.com) and Axon (axon.wellcomeapps.com) are interactive games that provide data to research teams. Second, community websites allow some researchers to gather large amount of data from individuals who voluntarily participate in research projects. For example, Test My Brain (www.testmybrain.org) and Project Implicit (www.projectimplicit.net) have successfully produced a large number of publications from crowdsourced data collection (e.g. [8•,9–11]). Third, some companies that gather big data can partner with academics to make the data available for research. For example, studies have resulted from data collected by Facebook [12•] and Fit Brains (e.g. [13]). Fourth, some groups have looked to create new communities where large datasets can be openly shared for ongoing research purposes (e.g. www.dataonthemind.org). Finally, there are specific programs that are designed as 'brain training' tools that purposefully gather large amounts of data with the hopes of helping the users (e.g. Fit Brains, Lumosity, Ultimeyes). While there are open questions about if and how brain-training tools affect cognition [14•], the gathered datasets have potential for research purposes.

> "As performance specialists, therapists, applied sport scientists, nutritionists, etc., we must listen to the voice of the athletes we train (and those that we don't, but that we can learn from listening to). Said best by Matt Nurse at Nike's Sport Research Lab, one key voice to listen to is their biometric data. This is a voice that tells a story at a depth specificity that no verbalization or visualization can provide. By themselves, data are useless. But added to the other voices of the athlete (subjective and objective) we have the best ability to provide impactful solutions and management guidance to reach full athletic potential. Sometimes these data speak instantly, and other times retrospectively. Either way, without it we are truly flying the high performance plane blind."
> **Lance Walker, MS, PT**
> **Director, Michael Johnson Performance**

Finding the right partnership

Academics have successfully accessed large datasets for research purposes by making their own programs and gathering their own data (e.g. Space Fortress [15,16]), but there are many challenges to consider — to get really 'big data' numbers the program needs to go 'viral,' creating your own program can be costly, and creating your own program often requires skills not all academics possess. Alternatively, academics can find an appropriate industry partner that leverages the strengths of the partner while allowing academics to focus on their areas of expertise.

Our suggestion is to partner with a company or institution that already has an established program that is proven to be successful and can be a source of data, which helps guarantee the critical mass necessary for an influx of big data and can potentially save months or years of program development where the probability of viral success is likely low. While we feel this path is ideal, there are clear challenges that must be identified and navigated to forge a successful partnership. For example, the industry partner's program was likely not designed as an experiment, so it may not be appropriately controlled without (significant) adjustments. Likewise, the company may be hesitant to 'rock the boat' as the changes could affect their successful product or could create public backlash [17,18]. These are just two simple examples of the challenges one might face — while such challenges are likely to arise, the potential payoff of forging this path can be well worth the effort.

What problem are you solving?

While the lure of big data can be enticing, it is vital that academic and industry partners keep an eye toward the end result and what they are looking to accomplish. Academically, it is key to make sure the data can speak to the underlying theories and mechanisms that are at the heart of your research. Practically, it is key to make sure the efforts can help address a real problem. See Box 1 for broader insight into this issue.

> "Advanced Analytics and Big Data are the motors of digital innovation, disruption and modern business transformations. Successfully competing in business is quickly becoming an entirely different ballgame: Winning requires mastery of data, of creating original insights, and of navigating transformative change when applying such insights to large organizations. My clients are using Big Data to improve their decision making, their processes, products and services. They develop new offerings, and the best find ways to out-innovate their competition and disrupt their respective industries. As a Bain Partner and facilitator of business transformations these are truly exciting and energizing times. We are leaders in advanced analytics and are investing heavily into Bain's talent, our tools and techniques as we harness the power of Big Data to help our clients create such high levels of economic value that together we set new standards of excellence in our respective industries."
> **Rasmus Wegener, Ph.D.**
> **Partner, Bain & Company**

Possible funding sources

Obtaining funding is often a critical step for a successful project. While research can get off the ground with minimal external support, ultimately resources will likely be needed to support the research team in their efforts. There are many paths to take. First, several US government agencies have funding mechanisms that are geared toward small businesses and/or academic–industry partnerships (e.g. Small Business Innovative Research (SBIR) awards[3]). Second, other funding mechanisms are specifically aimed at funding big data projects (e.g. NSF's Big Data Initiative[4]). Third, many government agencies recognize the important and power of big data (e.g. the National Geospatial-Intelligence Agency's big data challenge[5]), and they can support research directly. Finally, there are many private organizations that want tangible answers and can fund research to help them find those answers. Our primary suggestion is to be entrepreneurial and seek out creative solutions.

> "Big data is already reshaping the field of cancer control. The great promise over the next 5–10 years is the ability to get intensive longitudinal data on individuals. Anyone with a smartphone will be able to easily volunteer to contribute data on their cognitive function, activity levels, sleep habits, and a wide range of other behavioral variables. This will allow us to break the logjam on many outstanding problems in cancer prevention and control."
>
> **Todd Horowitz, Ph.D.**
> **Program Director, National Cancer Institute, National Institutes of Health**

Conclusions

Big data has the potential to provide exciting solutions to difficult problems. From an academic point of view, gaining access to big data can be an amazing windfall; the *Airport Scanner* dataset can be used to advance vision science research for (at least) another decade — assuming a well-functioning cognitive psychology lab can test ~500 participants a year, the *Airport Scanner* dataset already accounts for ~25 000 years worth of lab-based data. While there are risks, challenges, and investments of time and money, we believe the benefits make the risks and challenges worthwhile and the opportunities and rewards are exciting. From an industry partner point of view, there is potential for great gains from working with academics to use big data to advance real world issues and in many cases at minimal expense to the industry partner. Each situation needs to be evaluated on its own merits, but in the end, without aiming high, potential gains will never be realized.

Conflict of interest statement

The current article outlines the possible benefits and drawbacks of using big data for research purposes. While neither SRM nor BS see a conflict of interest related to any of the opinions stated in this chapter, they are co-partners of a company (Kedlin Screening International) that has used big data.

Acknowledgements

The authors thank Thomas Liljetoft, Alex Hartley, and the Duke and GW Visual Cognition Labs, for thoughtful input, and they thank the players of Airport Scanner for creating the dataset. Dr. Mitroff's research using the Airport Scanner dataset has been supported by the Army Research Office (Award #s W911NF-14-1-0361 and W911NF-16-1-0274) and the TSA (#HSTS0415CCT7031). Mr. Sharpe and Dr. Mitroff have been supported by TSA contract #HSTS0415CCT7031.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of special interest

1. Mitroff SR, Biggs AT: **The ultra-rare-item effect: visual search for exceedingly rare items is highly susceptible to error**. *Psychol Sci* 2014, **25**:284-289 http://dx.doi.org/10.1177/0956797613504221.

2. • Kravitz DJ, Mitroff SR: **Estimates of a priori power and false discovery rates induced by post-hoc changes from thousands of independent replications**. *J Vis* 2017, **17**:223 http://dx.doi.org/10.1167/17.10.223.

Kravitz *et al.* (2017) [2•] is an example of how a massive dataset can be used to address a wide variety of topics. This project used the data from Airport Scanner to examine how simple design choices made by experimenters (e.g., how to use pilot data) can impact false discovery rates. By using real-data for 10,000s of independent replications, this project was able to calculate the impact of such choices and their implications for scientific integrity.

3. Wegener R, Sinha V: *Navigating the "Big Data" Challenge*. Bain & Company; 2012 http://www.bain.com/publications/articles/navigating-the-big-data-challenge.aspx.

4. Biggs AT, Adamo SH, Dowd EW, Mitroff SR: **Examining perceptual and conceptual set biases in multiple-target visual search**. *Atten Percept Psychophys* 2015, **77**:844-855 http://dx.doi.org/10.3758/s13414-014-0822-0.

5. Biggs AT, Adamo SH, Mitroff SR: **Rare, but obviously there: effects of target frequency and salience on visual search accuracy**. *Acta Psychol (Amst)* 2014, **152**:158-165 http://dx.doi.org/10.1016/j.actpsy.2014.08.005.

6. • Ericson JM, Kravitz DJ, Mitroff SR: **Visual search: you are who you are (+ a learning curve)**. *Perception* 2017 http://dx.doi.org/10.1177/0301006617721091.

Ericson *et al.* [6•] used the Airport Scanner data to examine if early performance could reliably predict later success. It was found that accuracy and response time on even the very first trial was related to, and predictive of, later success in Airport Scanner. This suggests that some individuals might just be "better" at visual search, and they can be identified from early performance in a simple task.

7. •• Mitroff SR, Biggs AT, Adamo SH, Dowd EW, Winkle J, Clark K: **What can 1 billion trials tell us about visual search?** *J Exp Psychol: Hum Percept Perform* 2015, **41**:1-5 http://dx.doi.org/10.1037/xhp0000012.

Mitroff *et al.* [7••] is a review/opinion piece about how a massive dataset of over 1 billion trials can be used to advance academic theories. This paper is about the Airport Scanner data discussed in this chapter (which is now

[3] https://en.wikipedia.org/wiki/Small_Business_Innovation_Research.

[4] https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767.

[5] https://www.nga.mil/MediaRoom/PressReleases/Pages/NGA-challenge-offers-up-to-$200K-for-disparate-data-solutions.aspx.

at over 3 billion trials). Mitroff *et al.* [7••] highlights a few examples of using the data to inform visual search theories.

8. • Fortenbaugh FC, DeGutis J, Germine L, Wilmer JB, Grosso M, Russo K, Esterman M: **Sustained attention across the life span in a sample of 10,000 dissociating ability and strategy**. *Psychol Sci* 2015, **26**:1497-1510.

Fortenbaugh *et al.* [8•] gathered data from a large sample so that they could address specific questions about attention varies with age. It is a nice example of using large datasets to address questions of interest.

9. Germine L, Duchaine B, Nakayama K: **Where cognitive development and aging meet: face learning ability peaks after age 30**. *Cognition* 2011, **118**:201-210.

10. Halberda J, Ly R, Wilmer JB, Naiman DQ, Germine L: **Number sense across the lifespan as revealed by a massive Internet-based sample**. *Proc Natl Acad Sci U S A* 2012, **109**:11116-111120.

11. Nosek BA, Banaji MR, Greenwald AG: **Harvesting implicit group attitudes and beliefs from a demonstration website**. *Group Dyn* 2002, **6**:101-115.

12. • Kramer ADI, Guillory JE, Hancock JT: **Experimental evidence of massive-scale emotional contagion through social networks**. *Proc Natl Acad Sci U S A* 2014, **111**:8788-8790.

Kramer *et al.* [12•] introduced a small change into Facebook's algorithm for determining which news items appear in a user's feed for a one-week period. This was done for a small subset of users, with some seeing more positive items and some seeing fewer positive items. The study received public backlash for 'manipulating' users without their knowledge. It is an interesting study both for its use of a massive social media platform and for the public's reaction.

13. Bowles AR, Harper D, Lin CH, Amer L, Linck JA: *Cognitive ability profiles, brain training gains, and foreign language learning*. Boston, MA: Poster presented at the 57th Annual Meeting of the Psychonomic Society; 2016, November.

14. • Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, Stine-Morrow EAL: **Do 'brain training' programs work?** *Psychol Sci Public Interest* 2016, **17**:103-186.

Simons *et al.* [14•] is a comprehensive and critical evaluation of commercial brain training programs. Brain training programs offer a potentially exciting source of big data (some collected billions of trials across millions of individuals), but Simons *et al.* highlights that the data may need to be considering in a broader context.

15. Blumen HM, Gopher D, Steinerman JR, Stern Y: **Training cognitive control in older adults with the Space Fortress game: the role of training instructions and basic motor ability**. *Front Aging Neurosci* 2010, **2**:1-12.

16. Mané AM, Donchin E: **The Space Fortress game**. *Acta Psychol (Amst)* 1989, **71**:17-22 http://dx.doi.org/10.1016/0001-6918(89)90003-6.

17. Fiske ST, Hauser RM: **Protecting human research participants in the age of big data**. *Proc Natl Acad Sci U S A* 2014, **111**:13675-13676.

18. Kahn JP, Vayena E, Mastroianni AC: **Opinion: learning as we go: lessons from the publication of Facebook's social-computing research**. *Proc Natl Acad Sci U S A* 2014, **111**:13677-13679 http://dx.doi.org/10.1073/pnas.1416405111.