

EDITORIAL

Open Access

Novel predictions arise from contradictions



Itai Yanai^{1*} and Martin Lercher^{2*}

* Correspondence: itai.yanai@nyulangone.org; martin.lercher@hhu.de

¹Institute for Computational Medicine, NYU Langone Health, New York, NY 10016, USA

²Institute for Computer Science & Department of Biology, Heinrich Heine University, 40225 Düsseldorf, Germany

“In formal logic, a contradiction is the signal of defeat, but in the evolution of real knowledge, it marks the first step in progress toward a victory.” Alfred North Whitehead

From the outside, science seems to be the epitome of order, with its careful logical process, white lab coats, the methodical analyses of data, and, at its core, the formal testing of hypotheses. While that image may capture the “day science” aspect of science, it ignores the creative “night science” part, which generates the hypotheses in the first place. When we are in night science mode, we recognize facts that do not sit quite right against our clear and precise mental representation of the state of knowledge. While such contradictions arise from the data generated by day science, it is night science that revels in them, as these are the first, faint glimpses of new concepts. Depending on our state of mind, contradictions might appear as nuisances; embracing them helps us to counteract our natural human tendency for confirmation bias, a well-documented phenomenon in psychology. To explore the interaction of confirmation bias with a contradiction present in a dataset, we devised a simple experiment: individuals with different expectations examined the same data plot, which showed a superposition of two conflicting trends. We found that participants who expected a positive correlation between the two variables in the plot were more than twice as likely to report detecting one than those expecting a negative correlation. We posit that night science’s exploratory mode counteracts such cognitive biases, opening the door to new insights and predictions that can profoundly alter the course of a project. Thus, while science’s practitioners may be biased, the cyclical process of day science and night science allows us to spiral ever closer to the truth.

Data is not transparent

Science prides itself on being above the “idols of the tribe and the cave” [1], unperturbed by assumptions and unproven theories. Data is objective, after all, and scientists commit to “letting the data do the talking.” But data, of course, cannot speak for itself. It must be interpreted against an extensive conceptual, theoretical, and methodological background—and this background is unlikely to be bias-free. Thus, to venture that a dataset makes a particular statement hides the degree to which potential biases may have influenced our conclusions.



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

To study how our biases influence data interpretation, we performed an experiment with university students of computer science [2]. We showed them the plot in Fig. 1a, claiming that the data was collected to study the relationship between the personal wealth of individuals (x -axis) and their life satisfaction (“happiness,” y -axis). The data points indicating individuals were colored according to age groups. We then asked each participant: “Does the data suggest that wealth leads to happiness?”

The plot shows an overall negative correlation between wealth and happiness (Fig. 1b), while within each age group, a positive correlation is evident (Fig. 1c). This is an example of Simpson’s paradox [3, 4], where the correlation between two variables changes sign after controlling for another variable. The most parsimonious explanation of the pattern in the plot is that, all else being equal, more money makes you happier (thus the positive within-age group correlation). Across age groups, this effect is drowned out by a second, stronger effect from an underlying negative relationship between age and happiness. [We note that since this data was artificially created, no conclusions on real-life connections should be drawn.] In sum, while a first glimpse may indicate that wealthier individuals are less happy, the data indeed suggests that wealth leads to happiness. About a third of our participants acknowledged this by answering “yes” (49/171).

Regardless of what conclusion a particular participant made, we would like to assume that it arose from an analysis of the data plot, rather than resulting from preconceived notions. To test if this was indeed the case, we had inquired into the participants’ biases before showing them the data plot, asking the following question: “Imagine collecting data to study the relationship between the personal wealth of individuals and their life satisfaction (“happiness”). What overall general correlation do you expect?” Seventy percent of the students (119/171) expected a positive correlation, while the remaining students expected a negative one. Strikingly, the two groups saw the same data differently: those with an expectation of a positive correlation were more than twice as likely to conclude a positive correlation than those expecting a negative correlation (Table 1; 42/119 vs. 7/52, odds ratio 3.48, $P = 0.0024$ from one-sided Fisher’s exact test) [2].

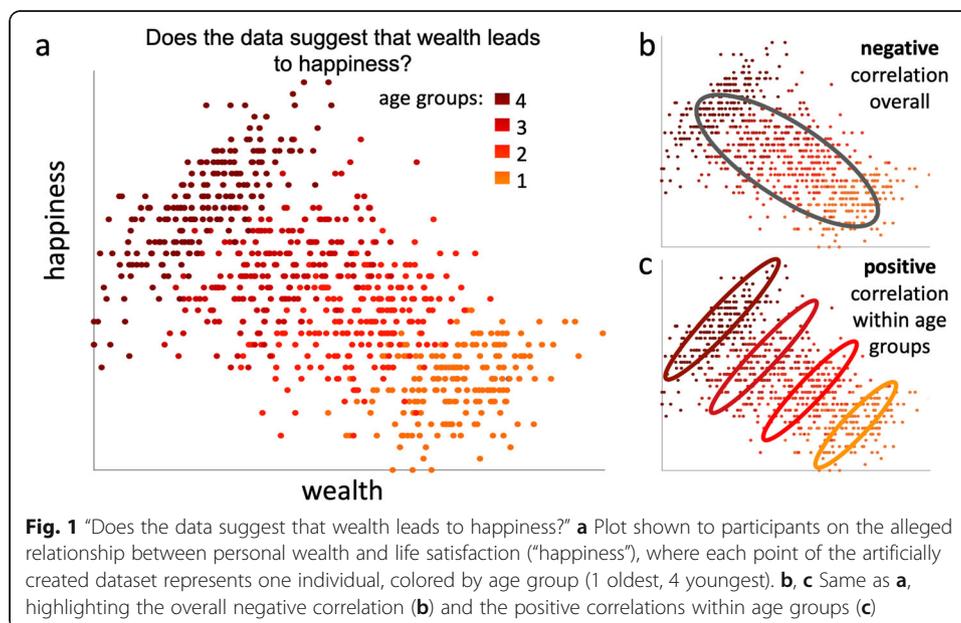


Table 1 The contingency table for students in the two groups (expecting either a positive or a negative correlation between wealth and happiness) and what they observed [2]

		Observed		Row sums
		Negative correlation	Positive correlation	
Expected	Negative correlation	45	7	52 (30%)
	Positive correlation	77	42	119 (70%)
	Column sums	122 (71%)	49 (29%)	N=171

Thus, the results suggest that when the participants looked at the plot, they came with a bias that framed their perception. This may reflect a general phenomenon in science. One famous episode stars Arthur Stanley Eddington, who in 1919 set out to test a prediction of Einstein’s theory of relativity. The data Eddington published could have been interpreted as either supporting Einstein’s or Newton’s theory of gravity (or, maybe most appropriately, as inconclusive). Yet in his paper, Eddington—who expected Einstein’s theory to prevail—framed the data as clearly supporting Einstein [5].

The extent to which the same data can lead to different conclusions has been the subject of recent studies that provide different scientific experts with a dataset together with a set of hypotheses to test. In one study, 70 independent teams were asked to analyze functional magnetic resonance images and to test 9 specific hypotheses [6]. Strikingly, no two teams chose identical work flows to analyze the dataset, and sets of teams reported contradictory, statistically significant effects based on the same dataset. In another study, 73 teams used the same data to test a single hypothesis; the “tremendous variation” in conclusions led the researchers to conclude a “vast universe of research design variability normally hidden from view” [7]. These and other studies [8–13] demonstrate that data is not transparent and that converting data into information through statistical analysis has a substantial subjective component. Our own experiment shows that even when studying the *same* plot, preconceived biases can result in different interpretations.

Scientists are biased—especially you

Psychologists refer to “confirmation bias” as the propensity of viewing new evidence as supporting one’s beliefs. Writing in the fifth century BC, the historian Thucydides put it this way: “It is a habit of mankind to entrust to careless hope what they long for, and to use sovereign reason to thrust aside what they do not fancy” [14]. In more recent times, experiments have shown that people demand a much higher standard of evidence for ideas they find disagreeable compared to those they hold themselves [15, 16]. As different individuals have different beliefs and experiences, each member of a community may end up perceiving reality differently, each through the prism of their own biased views. Confirmation bias may explain why harmful medieval medical practices were perpetuated over centuries, as only those patients that recovered (possibly despite rather than because of the treatment) were remembered [17]. The same phenomenon may underlie today’s widespread acceptance of “alternative” medicine [18–20].

While modern science appears to demand an objectivity that stands above such biases, confirmation bias is also well documented here. It shows itself, for example, in the peer review process of scientific papers. Studies whose findings are incompatible

with a reviewer's own assumptions are reviewed much more harshly than those that support the reviewer's beliefs [21–23]. Confirmation bias leads scientists to dismiss or misinterpret publications that contradict their own preconceptions, up to the point where contradicting papers are cited as if they in fact supported a favored notion [24].

Beyond publications, does the execution of the scientific method itself also fall prey to confirmation bias? In theory, it should not, as emphasized by Karl Popper's description of the scientific process [25]. For any given new idea, scientists should attempt with all of their experimental powers to falsify it, to prove it to be wrong. But, as philosopher Michael Strevens has argued, humans are not disciplined enough to strictly follow this method in their daily work [5]. Indeed, as much as we hold the philosophy of falsification in high regard, it simply is not very practical. First of all, hardly anyone actually excludes a beautiful and well evidenced hypothesis at the first sight of a falsifying result. If, for example, an experiment contradicts the conservation of energy, we do not simply throw out the first law of thermodynamics. Instead, we search for some flaw in the experiment or its interpretation [26]. But more than that, as we describe next, scientists—just like any other human beings—are hoping for and hence looking for evidence that confirms, not refutes, their favorite notions. If we make a prediction and the prediction seems to be borne out by the data on the surface, we tend to not dig deeper.

Do not stop just because you like the result

Many scientists and the public at large are concerned about a “reproducibility crisis” in science: the observation that many published results cannot be replicated [27]. While it may be tempting to lay the fault at the door of a select group of misbehaving individuals, the wide scope of the observed non-reproducibility suggests that it may be a pernicious aspect of the scientific process itself [27].

As a point of reference, it is useful to recognize one scientific framework that escaped the reproducibility crisis: clinical trials. The acknowledgement and correction of potential biases is built into the rigid process of clinical trials, where study designs are pre-registered and detailed protocols have to be followed to the letter. Confounding variables are clearly identified a priori, the data is blinded to exclude potential biases of the practitioners, and significance is tested only when the pre-specified dataset has been collected. Clinical trials are carefully designed to be exclusively in the mode of hypothesis-testing; if executed with a large enough sample size, they ought to be immune not only to confirmation bias, but also to other sources of systematic irreproducibility.

In sharp contrast to the purely hypothesis-testing mode of a clinical trial, basic research projects typically do not fully know what to expect from the data before analyzing it—after all, “If we knew what we were doing, it wouldn't be research, would it?” High-throughput datasets are especially likely to contain information unanticipated in their generation, observations one could not have predicted a priori; for this reason, we may often be better poised for a discovery if we conversed with a dataset without having formulated a concrete hypothesis [28]. The natural place for almost every dataset is right in the middle of the data-hypothesis conversation [29, 30], where it is used both in a hypothesis-testing way and in explorations that look for unexpected patterns [31], the seeds of future hypotheses.

Arguably, the central part of the scientific method is to challenge a hypothesis by contrasting its predictions with data. But when we do that, we are typically anxious for the falsification to fail: unless we are testing someone else's competing hypothesis, we

hope—and frequently expect—that our predictions will be borne out. If the results do not initially comply, we will think about problems with the experiment or with auxiliary assumptions (such as how we expect a genetic manipulation to perturb a cellular system). Sometimes, we will identify multiple such problems, and the results may converge to what we predicted them to be. There is, in principle, nothing wrong with this general approach: the scientific process is very much trial and error, and we cannot expect that an initial experiment and our first analysis were without fault. Obviously, we have to be careful not to selectively exclude contradictory data by making sure that similar errors have not equally affected other data points. There is a more subtle point here, though, a more discrete way in which confirmation bias may creep into our science. Our human instincts will lead to a sense of fulfillment once the expected pattern finally emerges. While this may mark the perfect time for a well-earned coffee break, it is not the time when we should stop analyzing the data. Instead, we should continue to think about possible biases and errors in our experiment, its analysis, and its interpretation. If we do not, we may be abandoning our efforts toward falsification too soon.

One extreme example, where confirmation bias is elevated to a guiding principle, is p-hacking [32]: one modifies the specifics of the analysis until the expected result emerges, subsequently reporting only the final configuration. It is important to realize, though, that in this case, it is the biased reporting that contributes to the reproducibility crisis, not the exploratory analysis itself [33]. Just the opposite: an exploration of how our results vary with changes in the specifics of the analysis, if communicated openly, provides important information on the robustness of our interpretation.

An elegant way to counter the drag toward self-fulfilling hypotheses is to test not one, but multiple alternative hypotheses, a core element of a method John Platt called “strong inference” [34]. Platt argued that the fastest scientific progress results from formulating a set of opposing hypotheses and then devising a test that can distinguish between them. While this is indeed a powerful approach, we often do not know initially what may be the best set of competing hypotheses. Forcing ourselves to look beyond one favored hypothesis in order to come up with such competing hypotheses is a serious—and non-fun—night science task, requiring hard and deliberate work.

Embrace the contradiction

Looking back at the wealth-happiness experiment described above, the main aspect was also that the participants were faced with a contradiction. The data could be interpreted in two ways: a positive correlation if one looks at the data in one way (within age groups) or a negative correlation if one looks at it another (overall). Presumably, given enough staring at the data, each participant would have reached the same conclusion that it is the positive correlation that best summarizes the underlying relationship. As we argued above, the problem is that contradictions often go unexplored, perhaps because they are confusing or inconsistent with prior notions, or simply because acknowledging them suggests tedious additional work. And yet, a contradiction should be reason for joy: it hints at an apparent discrepancy between the state of knowledge and reality—we might have stumbled upon something new and interesting [35].

In research, we frequently find that to make sense of contradictory data, we must identify a false, hidden assumption. As we highlighted in an earlier piece [36], Einstein’s path toward the special theory of relativity began when he noticed a contradiction

between Maxwell's equations and the idea of traveling at the speed of light. It took him years of hard work, though, to identify the false, hidden assumption: that time was absolute and independent of our frame of reference.

As an example from our own work, in a recent project we compared developmental gene expression across ten species, each from a different phylum (flies, fish, worms, etc.) [37]. Having previously compared sets of species from the same phylum, we expected to again find a shared pattern of expression occurring toward the middle of the development of all embryos (the "hourglass model"). This was an exciting prospect: it would have confirmed a specific pattern of signaling and transcription factor expression common to all animals. However, when comparing the gene expression across phyla, what emerged was the exact opposite: the greatest correspondence between phyla occurred in the early and late transcriptomes, bridged by a less conserved transitional state (an "inverse hourglass"). Confused by this contradictory signal, we retreated into more night science to attempt a resolution. We finally realized that when combining the contradictory patterns within and across phyla, a molecular definition of animal phyla emerged: early and late development are broadly conserved, while the transitional state—conserved within, but highly variable between phyla—is phylum-specific, distinguishing one phylum from another. We had to learn to apply the two modes, the hourglass and the inverse hourglass, to different evolutionary timescales.

In the course of any project, there may be points where we stumble upon more or less blatant contradictions. The way in which we choose to deal with them will define the project's fate. Embracing a contradiction requires us to be comfortable with uncertainty and will inevitably prolong the project. But it will provide space for unweaving the contradiction—arriving at the contradiction was not a signal of defeat, but rather the first sign of progress, as indicated by the Whitehead quote above.

In the absence of a contradiction, a common night science approach is to actually seek one out, playing devil's advocate. Adopting a contrary viewpoint "for the sake of argument" can help to counter confirmation bias [38]. More than once, in a discussion with a student or collaborator, most of us have probably started a sentence with "Well, a reviewer might say..." If that does not help, a more severe approach would be to imagine that at some point in the future, a competing lab would write a paper that criticizes the current project. What would that "anti-paper" be about? By deliberately challenging our assumptions and expectations, we may avoid cheating ourselves out of discoveries.

Acknowledgements

We thank Jacob Goldberg for discussions about contradictions, Arjun Raj for insights into the non-fun but necessary search for alternative hypotheses, and Ellen Rothenberg for a discussion on the relationship between discoveries and knowledge representation. We thank Veronica Maurino, Dalia Barkley, and Andrew Pountain for critical readings of the manuscript. We thank Michal Gilon-Yanai and Leon Anavy for suggestions on the design of the experiment. We thank the computer science students of Heinrich Heine University Düsseldorf that participated in our experiment.

Authors' contributions

IY and MJL wrote the manuscript together. The authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Published online: 11 May 2021

References

1. Bacon F. *The New Organon* (1620). Whithorn: Anodos Books; 2019.
2. Yanai I, Lercher MJ. Does the data do the talking? Confirmation bias in the interpretation of data visualizations. *PsyArXiv* 2021. <https://doi.org/10.31234/osf.io/q8ckd>.

3. Simpson EH. The interpretation of interaction in contingency tables. *J R Stat Soc Ser B (Methodological)*. 1951;13(2):238–41. <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>.
4. Wagner CH. Simpson's Paradox in real life. *Am Statistician* 1982;36:46. <https://doi.org/10.2307/2684093>.
5. Strevens M. *The knowledge machine: how irrationality created modern Science*. New York: Liveright Publishing; 2020.
6. Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*. 2020;582(7810):84–8. <https://doi.org/10.1038/s41586-020-2314-9>.
7. Breznau N, Rinke EM, Wuttke A, Adem M, Adriaans J, Alvarez-Benjumea A, et al. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *BITSS* 2021. <https://doi.org/10.31222/osf.io/cd5j9>.
8. Del Giudice M, Gangestad SW. A traveler's guide to the multiverse: promises, pitfalls, and a framework for the evaluation of analytic decisions. *Adv Methods Pract Psychol Sci*. 2021;4(1):251524592095492. <https://doi.org/10.1177/2515245920954925>.
9. Landy JF, Jia ML, Ding IL, Viganola D, Tierney W, Dreber A, et al. Crowdsourcing hypothesis tests: making transparent how design choices shape research results. *Psychol Bull*. 2020;146(5):451–79. <https://doi.org/10.1037/bul0000220>.
10. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011;22(11):1359–66. <https://doi.org/10.1177/0956797611417632>.
11. Dutilh G, Annis J, Brown SD, Cassey P, Evans NJ, Grasman RPPP, et al. The quality of response time data inference: a blinded, collaborative assessment of the validity of cognitive models. *Psychon Bull Rev*. 2019;26(4):1051–69. <https://doi.org/10.3758/s13423-017-1417-2>.
12. Bastiaansen JA, Kunkels YK, Blaauw FJ, Boker SM, Ceulemans E, Chen M, et al. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *J Psychosom Res*. 2020;137:110211.
13. Huntington-Klein N, Arenas A, Beam E, Bertoni M, Bloem JR, Burli P, et al. The influence of hidden researcher decisions in applied microeconomics. *Econ Inq*. 2021; <https://doi.org/10.1111/ecin.12992>.
14. Thucydides. *History of the Peloponnesian War*. New York: Courier Dover Publications; 2017.
15. Dawson E, Gilovich T, Regan DT. Motivated reasoning and performance on the was on Selection Task. *Pers Soc Psychol Bull*. 2002;28(10):1379–87. <https://doi.org/10.1177/014616702236869>.
16. Ditto PH, Lopez DF. Motivated skepticism: use of differential decision criteria for preferred and nonpreferred conclusions. *J Pers Soc Psychol*. 1992;63(4):568–84. <https://doi.org/10.1037/0022-3514.63.4.568>.
17. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol*. 1998;2(2):175–220. <https://doi.org/10.1037/1089-2680.2.2.175>.
18. Goldacre B. *Bad science*. New York: HarperPerennial; 2009.
19. Singh S, Ernst E. *Trick or treatment?: Alternative medicine on trial*. New York: Random House; 2009.
20. Atwood KC 4th. Naturopathy, pseudoscience, and medicine: myths and fallacies vs truth. *MedGenMed*. 2004;6:33.
21. Hergovich A, Schott R, Burger C. Biased evaluation of abstracts depending on topic and conclusion: further evidence of a confirmation bias within scientific psychology. *Curr Psychol*. 2010;29(3):188–209. <https://doi.org/10.1007/s12144-010-9087-5>.
22. Koehler JJ. The influence of prior beliefs on scientific judgments of evidence quality. *Organ Behav Hum Decis Process*. 1993;56(1):28–55. <https://doi.org/10.1006/obhd.1993.1044>.
23. Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognit Ther Res*. 1977;1(2):161–75. <https://doi.org/10.1007/BF01173636>.
24. Letrud K, Hernes S. Affirmative citation bias in scientific myth debunking: a three-in-one case study. *PLoS One*. 2019; 14(9):e0222213. <https://doi.org/10.1371/journal.pone.0222213>.
25. Popper K. *The logic of scientific discovery*: Routledge; 2005. <https://doi.org/10.4324/9780203994627>.
26. Chalmers AF. *What is this thing called science?* 4th ed. Maidenhead: Hackett Publishing; 2013.
27. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124>.
28. Yanai I, Lercher M. A hypothesis is a liability. *Genome Biol*. 2020;21(1):231. <https://doi.org/10.1186/s13059-020-02133-w>.
29. Yanai I, Lercher M. The data-hypothesis conversation. *Genome Biol*. 2021;22(1):58. <https://doi.org/10.1186/s13059-021-02277-3>.
30. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*. 2004;26(1):99–105. <https://doi.org/10.1002/bies.10385>.
31. Tukey JW. *Exploratory data analysis*: Addison-Wesley Publishing Company; 1977. <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02277-3>.
32. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. *PLoS Biol*. 2015;13(3):e1002106. <https://doi.org/10.1371/journal.pbio.1002106>.
33. Hollenbeck JR, Wright PM. Harking, sharking, and tharking: making the case for post hoc analysis of scientific data. *J Manage*. 2017;43:5–18.
34. Platt JR. Strong inference. *Science*. 1964;146(3642):347–53. <https://doi.org/10.1126/science.146.3642.347>.
35. Boyd D, Goldenberg J. *Inside the Box: A Proven System of Creativity for Breakthrough Results* Paperback. New York: Simon & Schuster; 2013.
36. Yanai I, Lercher M. What is the question? *Genome Biol*. 2019;20(1):289. <https://doi.org/10.1186/s13059-019-1902-1>.
37. Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, et al. The mid-developmental transition and the evolution of animal body plans. *Nature*. 2016;531(7596):637–41. <https://doi.org/10.1038/nature16994>.
38. Krueger D, Mann JD. *The Secret language of money: how to make smarter financial decisions and live a richer life*. New York: McGraw Hill Professional; 2009.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.