



یادداشت‌ها

مقدمه

۱. دانیل برنولی^۱، ریاضی‌دان سوئیسی، در جستاری راجع به مشکلات تخمین همین مثال را، با تیر و کمان به جای تفنگ، ارائه کرده بود.

Bernoulli, "The Most Probable Choice Between Several Discrepant Observations and the Formation Therefrom of the Most Likely Induction," *Biometrika* 48, no. 1–2 (June 1961): 3–18, <https://doi.org/10.1093/biomet/48.1-2.3>.

2. Joseph J. Doyle Jr., "Child Protection and Child Outcomes: Measuring the Effects of Foster Care," *American Economic Review* 95, no. 5 (December 2007): 1583–1610.

3. Stein Grimstad and Magne Jørgensen, "Inconsistency of Expert Judgment-Based Estimates of Software Development Effort," *Journal of Systems and Software* 80, no. 11 (2007): 1770–1777.

4. Andrew I. Schoenholtz, Jaya Ramji-Nogales, and Philip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication," *Stanford Law Review* 60, no. 2 (2007).

¹ Daniel Bernoulli

5. Mark A. Lemley and Bhaven Sampat, "Examiner Characteristics and Patent Office Outcomes," *Review of Economics and Statistics* 94, no. 3 (2012): 817–827. See also Iain Cockburn, Samuel Kortum, and Scott Stern, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics," working paper 8980, June 2002, www.nber.org/papers/w8980; and Michael D. Frakes and Melissa F. Wasserman, "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data," *Review of Economics and Statistics* 99, no. 3 (July 2017): 550–563.

بخش یک

فصل ۱

1. Marvin Frankel, *Criminal Sentences: Law Without Order*, 25 Inst. for Sci. Info. Current Contents / Soc. & Behavioral Scis.: This Week's Citation Classic 14, 2A-6 (June 23, 1986), available at <http://www.garfield.library.upenn.edu/classics1986/A1986C697400001.pdf>.
2. Marvin Frankel, *Criminal Sentences: Law Without Order (New York: Hill and Wang, 1973)*, 5.
3. Frankel, *Criminal Sentences*, 103.
4. Frankel, 5.
5. Frankel, 11.
6. Frankel, 114.
7. Frankel, 115.
8. Frankel, 119.
9. Anthony Partridge and William B. Eldridge, *The Second Circuit Sentence Study: A Report to the Judges of the Second Circuit August 1974 (Washington, DC: Federal Judicial Center, August 1974)*, 9.
10. US Senate, "Comprehensive Crime Control Act of 1983: Report of the Committee on the Judiciary, United States Senate, on S. 1762, Together with Additional and Minority Views" (Washington, DC: US Government Printing Office, 1983). Report No. 98-225.
11. Anthony Partridge and Eldridge, *Second Circuit Sentence Study*, A-11.
12. Partridge and Eldridge, *Second Circuit Sentence Study*, A-9.
13. Partridge and Eldridge, A-5-A-7

14. William Austin and Thomas A. Williams III, "A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity," *Journal of Criminal Law & Criminology* 68 (1977): 306.
15. John Bartolomeo et al., "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity," *Journal of Criminal Law and Criminology* 72, no. 2 (1981). (See chapter 6 for a full discussion.) See also Senate Report, 44.
16. Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous Factors in Judicial Decisions," *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 17 (2011): 6889-92.
17. Ozkan Eren and Naci Mocan, "Emotional Judges and Unlucky Juveniles," *American Economic Journal: Applied Economics* 10, no. 3 (2018): 171-205.
18. Daniel L. Chen and Markus Loecher, "Mood and the Malleability of Moral Reasoning: The Impact of Irrelevant Factors on Judicial Decisions," *SSRN Electronic Journal* (September 21, 2019): 1-70,
http://users.nber.org/dlchen/papers/Mood_and_the_Malleability_of_Moral_Reasoning.pdf.
19. Daniel L. Chen and Arnaud Philippe, "Clash of Norms: Judicial Leniency on Defendant Birthdays," (2020) available at SSRN: <https://ssrn.com/abstract=3203624>.
20. Anthony Heyes and Soodeh Saberian, "Temperature and Decisions: Evidence from 207,000 Court Cases," *American Economic Journal: Applied Economics* 11, no. 2 (2018): 238-265.
21. *Senate Report, 38.*
22. *Senate Report, 38.*
23. Justice Breyer is quoted in Jeffrey Rosen, "Breyer Restraint," *New Republic*, July 11, 1994, at 19, 25.
24. United States Sentencing Commission, Guidelines Manual (2018), www.ussc.gov/sites/default/files/pdf/guidelines-manual/2018/GLMFull.pdf.
25. James M. Anderson, Jeffrey R. Kling, and Kate Stith, "Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines," *Journal of Law and Economics* 42, no. S1 (April 1999): 271-308.
26. *US Sentencing Commission, The Federal Sentencing Guidelines: A Report on the Operation of the Guidelines System and Short-Term Impacts on Disparity in Sentencing, Use of Incarceration, and Prosecutorial Discretion and Plea Bargaining, vols. 1 & 2 (Washington, DC: US Sentencing Commission, 1991).*
27. Anderson, Kling, and Stith, "Interjudge Sentencing Disparity."

28. Paul J. Hofer, Kevin R. Blackwell, and R. Barry Ruback, "The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity," *Journal of Criminal Law and Criminology* 90 (1999): 239, 241.

29. Kate Stith and José Cabranes, *Fear of Judging: Sentencing Guidelines in the Federal Courts* (Chicago: University of Chicago Press, 1998), 79.

30. 543 U.S. 220 (2005).

31. US Sentencing Commission, "Results of Survey of United States District Judges, January 2010 through March 2010" (June 2010) (question 19, table 19), www.ussc.gov/sites/default/files/pdf/research-and-publications/research-projects-and-surveys/surveys/20100608_Judge_Survey.pdf.

32. Crystal Yang, "Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from Booker," *New York University Law Review* 89 (2014): 1268–1342; pp. 1278, 1334.

فصل ۲

۱. مدیران شرکت شرح تفصیلی نمونه‌های روشنگری را، مشابه ریسک‌ها و دعوی خسارت‌هایی که کارکنان به طور روزمره با آن‌ها درگیر بودند، آماده کردند. از کارکنان خواستند تا، مستقل از یکدیگر، هریک به ارزیابی دو یا سه نمونه بپردازند. برای کارشناسان خسارت شش نمونه و برای بیمه‌گران متخصص ریسک مالی چهار نمونه تهیه شده بود. کارکنان به مدت نیم‌روز از کارهای معمول خود آزاد بودند تا هرکدام دو یا سه مورد را ارزیابی کنند. آن‌ها دستور داشتند که مستقل کار کنند و به آن‌ها نگفته بودند که هدف از این مطالعه سنجش بی‌ثباتی قضاوت‌های آن‌هاست. در مجموع، از چهل‌وهشت بیمه‌گر تعداد هشتادوشش قضاوت و از شصت‌وهشت کارشناس خسارت تعداد صدوسیزده قضاوت به دست آوریم.

2. Dale W. Griffin and Lee Ross, "Subjective Construal, Social Inference, and Human Misunderstanding," *Advances in Experimental Social Psychology* 24 (1991): 319–359; Robert J. Robinson, Dacher Keltner, Andrew Ward, and Lee Ross, "Actual Versus Assumed Differences in Construal: 'Naive Realism' in Intergroup Perception and Conflict," *Journal of Personality and Social Psychology* 68, no. 3 (1995): 404; and Lee Ross and Andrew Ward, "Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding," *Values and Knowledge* (1997).

بخش دو

۱. انحراف معیار مجموعه‌ای از اعداد خود از یک کمیت آماری دیگر به دست می‌آید که به آن واریانس می‌گویند. برای محاسبه واریانس، ابتدا توزیع انحرافات از میانگین را به دست می‌آوریم و آن‌گاه توان‌دوم هریک از انحراف‌ها را می‌گیریم. واریانس عبارت است از میانگین مربعات این انحرافات، و انحراف معیار جذر واریانس است.

فصل ۴

1. R. T. Hodgson, "An Examination of Judge Reliability at a Major U.S. Wine Competition," *Journal of Wine Economics* 3, no. 2 (2008): 105–113.

۲. مطابق تعريف بعضی کسانی که دربارهٔ تصميم‌گيري تحقيق می‌کنند، تصميم عبارت است از انتخاب از بين گزینه‌ها و قضاوت‌های کمی عبارت‌اند از نوع خاصی از تصميم که در آن طیفی از انتخاب‌های ممکن وجود دارد. طبق آن دیدگاه، قضاوت‌ها نوع خاصی از تصميم هستند. رویکرد ما در اینجا متفاوت است: ما تصميم‌هایی را که می‌خواهند از بين چند گزینه یکی را انتخاب کنند برآمده از قضاوتی ارزشیابانه دربارهٔ هر گزینه می‌بینیم. به عبارت دیگر، تصميم‌ها را نوع خاصی از قضاوت در نظر می‌گیریم.

فصل ۵

۱. نخست آدرین-ماری لژاندر^۱ در سال ۱۹۰۵ روش کمترین مربعات را منتشر کرد. گاوس ادعا کرد که او اولین بار ده سال پیش از این روش استفاده کرده است و بعدها آن را به توسعهٔ یک نظریهٔ خطا و نمودار نرمال خطایی با نام خودش مرتبط ساخت. دربارهٔ این نزاع دربارهٔ تقدم و تأخر بحث‌های زیادی انجام شده است و مورخان اغلب به پذیرش ادعای گاوس گرایش دارند.

Stephen M. Stigler, "Gauss and the Invention of Least Squares," *Annals of Statistics* 9 [1981]: 465–474; and Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900* [Cambridge, MA: Belknap Press of Harvard University Press, 1986].

۲. نويز را همچون انحراف معيار خطا تعريف کردیم؛ بنابراین توان دوم نويز عبارت است از واریانس خطا. تعريف واریانس عبارت است «میانگین مربعات منهای مربع میانگین». چون میانگین همان سوگیری است، «مربع میانگین» همان توان دوم سوگیری است. بنابراین نويز به توان دوم برابر است با خطای میانگین مربعات منهای توان دوم سوگیری.

3. Berkeley J. Dietvorst and Soaham Bharti, "People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error," *Psychological Science* 31, no. 10 (2020): 1302–1314.

فصل ۶

1. Kevin Clancy, John Bartolomeo, David Richardson, and Charles Wellford, "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity," *Journal of Criminal Law and Criminology* 72, no. 2 (1981): 524–554; and INSLAW, Inc. et al., "Federal Sentencing: Towards a More Explicit Policy of Criminal Sanctions III-4," (1981).

¹ Adrien-Marie Legendre

۲. حکم کیفری می‌توانست شامل هر ترکیبی از حبس در زندان، دوران تحت نظارت، و جریمه باشد. برای سادگی در اینجا بیشتر تمرکزمان بر مؤلفه اصلی احکام کیفری - یعنی زندان - است و دو مؤلفه دیگر را کنار می‌گذاریم.

۳. در شرایطی که موارد و قضاوت‌کنندگان متعدد داریم، شکلی بسطیافته از معادله خطا که فصل پنج معرفی کردیم شامل تعبیری است که این تنوع یا واریانس را منعکس می‌کند. به طور خاص، اگر کلان‌سوگیری^۱ را میانگین خطا در تمام موارد در نظر بگیریم، و اگر این خطا در موارد مختلف یکسان نباشد، آنگاه با واریانس سوگیری‌های موردی مواجه هستیم. معادله به این صورت خواهد بود:

خطای میانگین مربعات = توان دوم کلان‌سوگیری + واریانس سوگیری‌های موردی + توان دوم نویز سیستمی.

۴. اعدادی که در این فصل به آن‌ها اشاره شده برگرفته از تحقیقی اصلی هستند که در ادامه خواهد آمد.

5. Bartolomeo et al., "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity," *Journal of Criminal Law and Criminology* 72, no. 2 [1981], table 6).

با این حال، در اینجا به طور کلی‌تری به تأثیرات هر پرونده، شامل همه مشخصاتی که به قاضی ارائه شده است، توجه داریم، شامل این که آیا متهم سابقه کیفری دارد و آیا در ارتکاب جرم از سلاحی استفاده شده است؟ طبق تعریف ما، همه این ویژگی‌ها بخشی از واریانس موردی حقیقی هستند، نه نویز. به همین ترتیب، ما همه برهم‌کنش‌ها میان ویژگی‌های هر پرونده را در واریانس پرونده ادغام کردیم (این امر ۱۱٪ کل واریانس را در برمی‌گیرد؛ ببینید:

Bartolomeo et al., table 10.

در نتیجه، از نو سهم واریانس موردی را ۵۶٪، تأثیر عمده قاضی (نویز ترازوی) را ۲۱٪، و برهم‌کنش‌های مجموع واریانس را ۲۳٪ در نظر می‌گیریم. بنابراین نویز سیستمی ۴۴٪ مجموع واریانس است.

6., Bartolomeo et al., 89.

در جدولی که میانگین مجازات برای هر پرونده فهرست شده است؛ واریانس ۱۵ است. اگر این ۵۶٪ از کل واریانس باشد، آنگاه کل واریانس ۲۶.۷۹ است و واریانس نویز سیستمی ۱۱.۷۹. مجذور آن واریانس بیانگر انحراف معیار یک پرونده انتخابی است، یعنی ۳.۴ سال.

۷. مجذور این واریانس معادل انحراف معیاری است که به نویز ترازوی قاضی نسبت داده شده است، یعنی ۲.۴ سال.

۸. این مقدار مجذور میانگین واریانس احکام کیفری برای شانزده پرونده است. آن را همان‌طور که در یادداشت قبلی ذکر شد محاسبه کرده‌ایم.

¹ Grand bias

۹. فرضیه افزایشی^۱ عملاً فرض می‌کند که سخت‌گیری قاضی مقدار ثابتی به زمان زندان اضافه می‌کند. نامحتمل است که این فرضیه درست باشد: محتمل‌تر آن است که سخت‌گیری قاضی مقداری متناسب با میانگین حکم را اضافه کند. در گزارش اصلی این مسئله مغفول مانده است و بنابراین راهی برای ارزیابی اهمیت آن نداریم.

10. "Sentence Decisionmaking," 23.

۱۱. این معادله برقرار است:

$$\text{(نویز سیستمی)} = ۲ = \text{(نویز ترازوی)} + ۲ \text{ (نویز الگویی)}$$

جدول نشان می‌دهد نویز سیستمی ۳.۴ سال و نویز ترازوی ۲.۴ سال است. نتیجه می‌شود که نویز الگویی نیز ۲.۴ سال است. محاسبات به‌عنوان مثال و توضیح هستند - به سبب خطاهای گردکردن مقادیرهای واقعی کمی متفاوت‌اند.

فصل ۷

1. http://www.iweblists.com/sports/basketball/FreeThrowPercent_c.html, consulted Dec. 27, 2020.

2. <https://www.basketball-reference.com/players/o/onealsh01.html>, consulted Dec. 27, 2020.

3. R. T. Hodgson, "An Examination of Judge Reliability at a Major U.S. Wine Competition," *Journal of Wine Economics* 3, no. 2 (2008): 105–113.

4. Stein Grimstad and Magne Jørgensen, "Inconsistency of Expert Judgment-Based Estimates of Software Development Effort," *Journal of Systems and Software* 80, no. 11 (2007): 1770–1777.

5. Robert H. Ashton, "A Review and Analysis of Research on the Test–Retest Reliability of Professional Judgment," *Journal of Behavioral Decision Making* 294, no. 3 (2000): 277–294. Incidentally, the author then noted that not a single one of the forty-one studies he reviewed was designed to evaluate occasion noise: "In all cases, the measurement of reliability was a by-product of some other research objectives" (Ashton, 279). This comment suggests that the interest in studying occasion noise is relatively recent.

6. Central Intelligence Agency, *The World Factbook* (Washington, DC: Central Intelligence Agency, 2020).

آنچه مورد استناد ماست شامل همه فرودگاه‌های بزرگ و کوچک مختلف قابل تشخیص از منظر هوایی است. باند یا باندهای پرواز ممکن است هموار یا ناهموار باشد. همچنین، این مقدار شامل موارد تعطیل یا متروک می‌شود.

¹ Additive hypothesis

7. Edward Vul and Harold Pashler, "Crowd Within: Probabilistic Representations Within Individuals,"

8. James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (New York: Doubleday, 2004).

۹. انحراف معیار قضاوت‌های میانگین (سنجۀ ما برای نویز) به نسبت مجذور تعداد قضاوت‌ها کاهش می‌یابد.

10. Vul and Pashler, "Crowd Within," 646.

11. Stefan M. Herzog and Ralph Hertwig, "Think Twice and Then: Combining or Choosing in Dialectical Bootstrapping?," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40, no. 1 (2014): 218–232.

12. Vul and Pashler, "Measuring the Crowd Within," 647.

13. Joseph P. Forgas, "Affective Influences on Interpersonal Behavior," *Psychological Inquiry* 13, no. 1 (2002): 1–28.

14. Forgas, "Affective Influences," 10.

15. A. Filipowicz, S. Barsade, and S. Melwani, "Understanding Emotional Transitions: The Interpersonal Consequences of Changing Emotions in Negotiations," *Journal of Personality and Social Psychology* 101, no. 3 (2011): 541–556.

16. Joseph P. Forgas, "She Just Doesn't Look like a Philosopher...? Affective Influences on the Halo Effect in Impression Formation," *European Journal of Social Psychology* 41, no. 7 (2011): 812–817.

17. Gordon Pennycook, James Allan Cheyne, Nathaniel Barr, Derek J. Koehler, and Jonathan A. Fugelsang, "On the Reception and Detection of Pseudo-Profound Bullshit," *Judgment and Decision Making* 10, no. 6 (2015): 549–563.

18. Harry Frankfurt, *On Bullshit* (Princeton, NJ: Princeton University Press, 2005).

19. Pennycook et al., "Pseudo-Profound Bullshit," 549.

20. Joseph P. Forgas, "Happy Believers and Sad Skeptics? Affective Influences on Gullibility," *Current Directions in Psychological Science* 28, no. 3 (2019): 306–313.

21. Joseph P. Forgas, "Mood Effects on Eyewitness Memory: Affective Influences on Susceptibility to Misinformation," *Journal of Experimental Social Psychology* 41, no. 6 (2005): 574–588.

22. Piercarlo Valdesolo and David Desteno, "Manipulations of Emotional Context Shape Moral Judgment," *Psychological Science* 17, no. 6 (2006): 476–477.

23. Hannah T. Nephrash and Michael L. Barnett, "Association of Primary Care Clinic Appointment Time with Opioid Prescribing," *JAMA Network Open* 2, no. 8 (2019);

Lindsey M. Philpot, Bushra A. Khokhar, Daniel L. Roellinger, Priya Ramar, and Jon O. Ebbert, "Time of Day Is Associated with Opioid Prescribing for Low Back Pain in Primary Care," *Journal of General Internal Medicine* 33 (2018): 1828.

24. Jeffrey A. Linder, Jason N. Doctor, Mark W. Friedberg, Harry Reyes Nieva, Caroline Birks, Daniella Meeker, and Craig R. Fox, "Time of Day and the Decision to Prescribe Antibiotics," *JAMA Internal Medicine* 174, no. 12 (2014): 2029–2031.

25. Rebecca H. Kim, Susan C. Day, Dylan S. Small, Christopher K. Snider, Charles A. L. Rareshide, and Mitesh S. Patel, "Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination," *JAMA Network Open* 1, no. 5 (2018): 1–10.

۲۶. برای بحث دربارهٔ حافظهٔ بهتر، ببینید:

Joseph P. Forgas, Liz Goldenberg, and Christian Unkelbach, "Can Bad Weather Improve Your Memory? An Obtrusive Field Study of Natural Mood Effects on Real-Life Memory," *Journal of Experimental Social Psychology* 45, no. 1 (2008): 254–257.

For comment on sunshine, see David Hirshleifer and Tyler Shumway, "Good Day Sunshine: Stock Returns and the Weather," *Journal of Finance* 58, no. 3 (2003): 1009–1032.

27. Uri Simonsohn, "Clouds Make Nerds Look Good: Field Evidence of the Impact of Incidental Factors on Decision Making," *Journal of Behavioral Decision Making* 20, no. 2 (2007): 143–152.

28. Daniel Chen et al., "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," *Quarterly Journal of Economics* 131, no. 3 (2016): 1181–1242.

29. Jaya Ramji-Nogales, Andrew I. Schoenholtz, and Philip Schrag, "Refugee Roulette: Disparities in Asylum Adjudication," *Stanford Law Review* 60, no. 2 (2007).

30. Michael J. Kahana et al., "The Variability Puzzle in Human Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44, no. 12 (2018): 1857–1863.

فصل ۸

1. Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts, "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science* 311 (2006): 854–856. See also Matthew Salganik and Duncan Watts, "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market," *Social Psychology Quarterly* 71 (2008): 338–355; and Matthew Salganik and Duncan Watts, "Web-Based Experiments for the Study of Collective Social Dynamics in Cultural Markets," *Topics in Cognitive Science* 1 (2009): 439–468.

2. Salganik and Watts, "Leading the Herd Astray."
3. Michael Macy et al., "Opinion Cascades and the Unpredictability of Partisan Polarization," *Science Advances* (2019): 1–8. See also Helen Margetts et al., *Political Turbulence* (Princeton: Princeton University Press, 2015).
4. Michael Macy et al., "Opinion Cascades."
5. Lev Muchnik et al., "Social Influence Bias: A Randomized Experiment," *Science* 341, no. 6146 (2013): 647–651.
6. Jan Lorenz et al., "How Social Influence Can Undermine the Wisdom of Crowd Effect," *Proceedings of the National Academy of Sciences* 108, no. 22 (2011): 9020–9025.
7. Daniel Kahneman, David Schkade, and Cass Sunstein, "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," *Journal of Risk and Uncertainty* 16 (1998): 49–86.
8. David Schkade, Cass R. Sunstein, and Daniel Kahneman, "Deliberating about Dollars: The Severity Shift," *Columbia Law Review* 100 (2000): 1139–1175.

بخش سه

۱. درصد هم‌نواپی (PC) ربط وثیقی به دابلوی کندال^۱ دارد که آن را با عنوان ضریب هم‌نواپی هم می‌شناسند.

2. Kanwal Kamboj et al., "A Study on the Correlation Between Foot Length and Height of an Individual and to Derive Regression Formulae to Estimate the Height from Foot Length of an Individual," *International Journal of Research in Medical Sciences* 6, no. 2 (2018): 528.

۳. درصد هم‌نواپی بر اساس این فرض محاسبه می‌شود که توزیع مشترک نرمال دومتغیره است. مقدارهای ذکرشده در جدول تقریب‌هایی بر اساس این فرض هستند. از جولیان پاریس بابت ساخت این جدول ممنونیم.

فصل ۹

1. Martin C. Yu and Nathan R. Kuncel, "Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments," *Personnel Assessment and Decisions* 6, no. 2 (2020): 1–10.

متخصصاً بر اساس میانگین بی‌وزن سه نمونه مطالعه‌شده، شامل مجموعاً ۸۴۷ مورد، به هم‌بستگی ۰.۱۵ رسیده‌اند. مطالعه واقعی از چند جهت با این وصف ساده‌شده تفاوت دارد.

¹ Kendall's W

۲. يك پيش‌شرط انجام ميانه‌گين وزني آن است كه بايد همه عوامل پيش‌بيني‌گر در واحدهاي قابل‌مقايسه اندازه‌گيري شوند. در مثال مقدماتي ما اين پيش‌شرط تأمين شده است، چون همه امتيازدهي‌ها بين ۰ تا ۱۰ است، اما هميشه چنين نيسـت. براي مثال، ممكن است عوامل پيش‌بيني‌گر عملكرد عبارت باشند از ارزيابي مصاحبه‌كننده در مقياس ۰ تا ۱۰، سال‌هاي تجربه بر حسب سال، و امتياز در آزمون مهارت. برنامه رگرسيون چندگانه قبل از تركيب اين پيش‌بيني‌كننده‌ها آن‌ها را به امتيازهاي استاندارد تبديل مي‌كند. امتياز استاندارد فاصله مشاهده را از ميانه‌گين جامعه اندازه مي‌گيرد و واحد آن انحراف معيار است. براي مثال، اگر ميانه‌گين آزمون مهارت ۵۵ و انحراف معيار ۸ باشد، امتياز استاندارد ۱.۵+ متناظر است با نتيجه آزمون ۶۷. به طور خاص، استانداردسازي داده‌هاي هر فرد باعث مي‌شود رد هرگونه خطا در ميانه‌گين يا در واريانس قضاوت‌هاي افراد حذف شود.

۳. يكي از ويژگي‌هاي مهم رگرسيون چندگانه آن است كه وزن بهينه براي هر عامل پيش‌بيني‌گر به ساير عوامل پيش‌بيني‌گر وابسته است. اگر عامل پيش‌بيني‌گري هم‌بستگي زيادي با ديگري داشته باشد، نبايد آن هم وزن زيادي داشته باشد - چون شكلي از «دوباره‌شماري»^۱ محسوب مي‌شود.

4. Robin M. Hogarth and Natalia Karelaia, "Heuristic and Linear Models of Judgment: Matching Rules and Environments," *Psychological Review* 114, no. 3 (2007): 734.

۵. مدل لنزي قضاوت^۲ يكي از چهارچوب‌هاي تحقيقي است كه در اين زمينه بسيار از آن استفاده شده و بحث حاضر بر آن استوار است. ببينيد:

Kenneth R. Hammond, "Probabilistic Functioning and the Clinical Method," *Psychological Review* 62, no. 4 (1955): 255-262; Natalia Karelaia and Robin M. Hogarth, "Determinants of Linear Judgment: A Meta-Analysis of Lens Model Studies," *Psychological Bulletin* 134, no. 3 (2008): 404-426.

6. Paul E. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Minneapolis: University of Minnesota Press, 1954).

7. Paul E. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Northvale, NJ: Aronson, 1996), preface.

8. "Paul E. Meehl," in Ed Lindzey (ed.), *A History of Psychology in Autobiography*, 1989.

9. "Paul E. Meehl," in *A History of Psychology in Autobiography*, ed. Ed Lindzey (Washington, DC: American Psychological Association, 1989), 362.

10. William M. Grove et al., "Clinical Versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment* 12, no. 1 (2000): 19-30.

11. William M. Grove and Paul E. Meehl, "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction

¹ Double counting

² Lens model of judgment

Procedures: The Clinical-Statistical Controversy," *Psychology, Public Policy, and Law* 2, no. 2 (1996): 293–323.

12. Lewis Goldberg, "Man Versus Model of Man: A Rationale, plus Some Evidence, for a Method of Improving on Clinical Inferences," *Psychological Bulletin* 73, no. 6 (1970): 422–432.

13. Milton Friedman and Leonard J. Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy* 56, no. 4 (1948): 279–304.

14. Karelaia and Hogarth, "Determinants of Linear Judgment," 411, table 1.

15. Nancy Wiggins and Eileen S. Kohen, "Man Versus Model of Man Revisited: The Forecasting of Graduate School Success," *Journal of Personality and Social Psychology* 19, no. 1 (1971): 100–106.

16. Karelaia and Hogarth, "Determinants of Linear Judgment."

۱۷. اصلاح ضریب هم‌بستگی برای اعتمادپذیری [یا پایایی] ناقص عامل پیش‌بینی‌گر را با عنوان اصلاح تضعیف^۱ می‌شناسند. فرمول آن به شرح زیر است:

$$\text{Corrected } r_{xy} = r_{xy} / \sqrt{r_{xx}}$$

با این توضیح که r_{xy} ضریب اعتمادپذیری است، یعنی نسبت واریانس حقیقی بر واریانس مشاهده‌شده عامل پیش‌بینی‌گر.

18. Yu and Kuncel, "Judgmental Consistency."

۱۹. در فصل آینده به تفصیل بیشتر از مدل‌های هم‌وزن و وزن‌تصادفی بحث می‌کنیم. این قید در مورد وزن وجود دارد که گستره آن‌ها اعداد کوچک است و سیگنال درستی بدهند.

فصل ۱۰

1. Robyn M. Dawes and Bernard Corrigan, "Linear Models in Decision Making," *Psychological Bulletin* 81, no. 2 (1974): 95–106.

داز و کوریگان همچنین استفاده از وزن‌های تصادفی را پیشنهاد کردند. مطالعه عملکرد مدیریتی که فصل ۹ شرح آن گذشت کاربرد همین ایده است.

2. Jason Dana, "What Makes Improper Linear Models Tick?," in *Rationality and Social Responsibility: Essays in Honor of Robyn M. Dawes*, ed. Joachim I. Krueger, 71–89 (New York: Psychology Press, 2008), 73.

3. Jason Dana and Robyn M. Dawes, "The Superiority of Simple Alternatives to Regression for Social Sciences Prediction," *Journal of Educational and Behavior Statistics* 29 (2004): 317–331; Dana, "What Makes Improper Linear Models Tick?"

¹ Correction for attenuation

4. Howard Wainer, "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," *Psychological Bulletin* 83, no. 2 (1976): 213–217.

5. Dana, "What Makes Improper Linear Models Tick?," 72.

6. Martin C. Yu and Nathan R. Kuncel, "Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments," *Personnel Assessment and Decisions* 6, no. 2 (2020): 1–10.

همانند فصل پیش، هم‌بستگی گزارش شده میانگین بی‌وزن سه نمونه مطالعه شده است. مقایسه در مورد هر سه نمونه صدق می‌کند: اعتبار قضاوت تخصصی بالینی به ترتیب ۰.۱۷، ۰.۱۶، و ۰.۱۳ بوده و اعتبار مدل‌های هم‌وزن ۰.۱۹، ۰.۳۳، و ۰.۲۲ بوده است.

7. Robyn M. Dawes, "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist* 34, no. 7 (1979): 571–582.

8. Dawes and Corrigan, "Linear Models in Decision Making," 105.

9. Jongbin Jung, Conner Concannon, Ravi Shroff, Sharad Goel, and Daniel G. Goldstein, "Simple Rules to Guide Expert Classifications," *Journal of the Royal Statistical Society, Statistics in Society*, no. 183 (2020): 771–800.

10. Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Science Advances* 4, no. 1 (2018): 1–6.

۱۱. این دو مثال مدل‌های خطی مبتنی بر مجموعه بسیار محدودی از متغیرها هستند (و در مورد مدل قرار وثیقه، مبتنی بر تقریبی از وزن‌های خطی حاصل از روش گردکردن که باعث می‌شود مدل به نوعی محاسبه مختصر پشت پاکت‌نامه‌ای تبدیل شود). نوع دیگری از «مدل ناسره» عبارت است از قاعده تک‌متغیری که فقط یک پیش‌بینی‌کننده را در نظر گرفته، باقی را نادیده می‌گیرند. ببینید:

Peter M. Todd and Gerd Gigerenzer, "Précis of Simple Heuristics That Make Us Smart," *Behavioral and Brain Sciences* 23, no. 5 (2000): 727–741.

12. P. Gendreau, T. Little, and C. Goggin, "A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!," *Criminology* 34 (1996).

۱۳. در این زمینه باید کلان‌بودن را همچون نسبت تعداد مشاهدات به عوامل پیش‌بینی‌گر در نظر گرفت. داز، در «زیبایی زمختی»، پیشنهاد کرده است تا پیش از آنکه وزن‌های بهینه در اعتبارسنجی متقابل عملکردی بهتر از وزن‌های واحدی داشته باشند، این مقدار باید حتی به ۱۵ یا ۲۰ به ۱ برسد. دانا و داز، در «برتری بدیل‌های ساده»، با استفاده از موردکاوی‌های بسیار بیشتر، این سطح را تا نسبت ۱۰۰ به ۱ افزایش می‌دهند.

14. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan, "Human Decisions and Machine Predictions," *Quarterly Journal of Economics* 133 (2018): 237–293.

۱۵. الگوریتم را با زیرمجموعه‌ای از داده‌های آموزشی تربیت و سپس توانایی آن در پیش‌بینی نتایج را با زیرمجموعه تصادفی دیگری ارزشیابی کردند.

16. Kleinberg et al., "Human Decisions," 16.
17. Gregory Stoddard, Jens Ludwig, and Sendhil Mullainathan, e-mail exchanges with the authors, June–July 2020.
18. B. Cowgill, "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Résumé Screening," paper presented at Smith Entrepreneurship Research Conference, College Park, MD, April 21, 2018.
19. William M. Grove and Paul E. Meehl, "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy," *Psychology, Public Policy, and Law* 2, no. 2 (1996): 293–323.
20. Jennifer M. Logg, Julia A. Minson, and Don A. Moore, "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment," *Organizational Behavior and Human Decision Processes* 151 (April 2018): 90–103.
21. B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err," *Journal of Experimental Psychology General* 144 (2015): 114–126. See also A. Prael and L. Van Swol, "Understanding Algorithm Aversion: When Is Advice from Automation Discounted?," *Journal of Forecasting* 36 (2017): 691–702.
22. M. T. Dzindolet, L. G. Pierce, H. P. Beck, and L. A. Dawe, "The Perceived Utility of Human and Automated Aids in a Visual Detection Task," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 44, no. 1 (2002): 79–94; K. A. Hoff and M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, no. 3 (2015): 407–434; and P. Madhavan and D. A. Wiegmann, "Similarities and Differences Between Human–Human and Human–Automation Trust: An Integrative Review," *Theoretical Issues in Ergonomics Science* 8, no. 4 (2007): 277–301.

فصل ۱۱

1. E. Dane and M. G. Pratt, "Exploring Intuition and Its Role in Managerial Decision Making," *Academy of Management Review* 32, no. 1 (2007): 33–54; Cinla Akinci and Eugene Sadler-Smith, "Intuition in Management Research: A Historical Review," *International Journal of Management Reviews* 14 (2012): 104–122; and Gerard P. Hodgkinson et al., "Intuition in Organizations: Implications for Strategic Management," *Long Range Planning* 42 (2009): 277–297.
2. Hodgkinson et al., "Intuition in Organizations," 279.
3. Nathan Kuncel et al., "Mechanical Versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis," *Journal of Applied Psychology* 98, no. 6

(2013): 1060–1072. See also chapter 24 for further discussion of personnel decisions.

4. Don A. Moore, *Perfectly Confident: How to Calibrate Your Decisions Wisely* (New York: HarperCollins, 2020).

5. Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton University Press, 2005), 239 and 233.

6. William M. Grove et al., "Clinical Versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment* 12, no. 1 (2000): 19–30.

7. Sendhil Mullainathan and Ziad Obermeyer, "Who Is Tested for Heart Attack and Who Should Be: Predicting Patient Risk and Physician Error," 2019. NBER Working Paper 26168, National Bureau of Economic Research.

8. Weston Agor, "The Logic of Intuition: How Top Executives Make Important Decisions," *Organizational Dynamics* 14, no. 3 (1986): 5–18; Lisa A. Burke and Monica K. Miller, "Taking the Mystery Out of Intuitive Decision Making," *Academy of Management Perspectives* 13, no. 4 (1999): 91–99.

9. Poornima Madhavan and Douglas A. Wiegmann, "Effects of Information Source, Pedigree, and Reliability on Operator Interaction with Decision Support Systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49, no. 5 (2007).

فصل ۱۲

1. Matthew J. Salganik et al., "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration," *Proceedings of the National Academy of Sciences* 117, no. 15 (2020): 8398–8403.

۲. شامل ۴۲۴۲ خانواده، چه این که بعضی خانواده‌های حاضر در مطالعه خانواده‌های شکننده به‌دلایل حریم خصوصی از این تحلیل حذف شدند.

۳. برگزارکنندگان این رقابت برای امتیازدهی به دقت از همان سنج‌های استفاده کردند که در بخش یک معرفی کردیم: خطای میانگین مربعات. برای سادگی مقایسه، آن‌ها همچنین خطای میانگین مربعات هریک از مدل‌ها را در مقابل استراتژی پیش‌بینی «بی‌حاصل» محک زدند. منظور از استراتژی بی‌حاصل این است: پیش‌بینی یکسان برای همه که طبق آن هیچ مورد فردی از میانگین داده‌های مورد استفاده برای تربیت الگوریتم متفاوت نیست. ما برای سادگی نتایج آن‌ها را به ضریب همبستگی تبدیل کرده‌ایم. ارتباط خطای میانگین مربعات و هم‌بستگی به این صورت است:

$$r^2 = (\text{Var}(Y) - \text{MSE}) / \text{Var}(Y)$$

به این صورت که $\text{Var}(Y)$ عبارت است از واریانس متغیر نتیجه و $(\text{Var}(Y) - \text{MSE})$ عبارت است از واریانس نتایج پیش‌بینی‌شده.

4. F. D. Richard et al., "One Hundred Years of Social Psychology Quantitatively Described," *Review of General Psychology* 7, no. 4 (2003): 331–363.

5. Gilles E. Gignac and Eva T. Szodorai, "Effect Size Guidelines for Individual Differences Researchers," *Personality and Individual Differences* 102 (2016): 74–78.

۶. در اینجا هشدار لازم است. آن‌طور که این مطالعه طراحی شده است، از داده‌های توصیفی موجودی استفاده می‌کند که بسیار کلان هستند، اما برای پیش‌بینی نتایج خاصی تهیه نشده‌اند. این تفاوتی مهم با مطالعه متخصصان از سوی تتلاک است که می‌توانند از هر اطلاعاتی که به نظرشان مناسب بود استفاده کنند. برای مثال، شاید ممکن باشد عوامل پیش‌بینی‌گر محکومیت را بیابیم که در این مجموعه از داده‌ها موجود نباشند، اما بتوان به نحوی آن‌ها را گردآوری کرد. لذا این مطالعه ثابت نمی‌کند که محکومیت و دیگر نتایج ذاتاً پیش‌بینی‌ناپذیرند، بلکه در این باره است که مبتنی بر این مجموعه داده‌ها، که از سوی کثیری از دانشمندان علوم اجتماعی مورد استفاده قرار گرفته است چقدر پیش‌بینی‌پذیرند.

7. Jake M. Hofman et al., "Prediction and Explanation in Social Systems," *Science* 355 (2017): 486–488; Duncan J. Watts et al., "Explanation, Prediction, and Causality: Three Sides of the Same Coin?," October 2018, 1–14, available through Center for Open Science, <https://osf.io/bgwjc>.

۸. این به یک تفکیک دیگر بسیار نزدیک است: تفکر مصداقی^۱ از تفکر غیرمصداقی یا مفهومی^۲ تفکیک شده است.

Amos Tversky and Daniel Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 4 (1983): 293–315.

9. Daniel Kahneman and Dale T. Miller, "Norm Theory: Comparing Reality to Its Alternatives," *Psychological Review* 93, no. 2 (1986): 136–153.

10. Baruch Fischhoff, "An Early History of Hindsight Research," *Social Cognition* 25, no. 1 (2007): 10–13, doi:10.1521/soco.2007.25.1.10; Baruch Fischhoff, "Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty," *Journal of Experimental Psychology: Human Perception and Performance* 1, no. 3 (1975): 288.

11. Daniel Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

¹ Extensional

² Intentional

بخش چهار

فصل ۱۳

1. *The first four decades*: Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011).

۲. در اینجا هشدار به جاست. روان‌شناسانی که به مطالعه سوگیری‌های قضاوت می‌پردازند اصلاً به پنج شرکت‌کننده در هر گروه که در شکل ۱۰ شاهد آن هستیم رضایت نمی‌دهند و دلیل بسیار خوبی هم دارند: چون قضاوت نویزی است، نتایج هر گروه آزمایش به‌ندرت چنان که در شکل ۱۱ دیده می‌شود خوشه‌زدیک به هم می‌سازد. افراد در میزان تأثیرپذیری از هر سوگیری متفاوت هستند و متغیرهای مربوط را به طور کامل نادیده نمی‌گیرند. برای مثال، با تعداد بسیار زیادی از شرکت‌کنندگان تقریباً با اطمینان می‌توانید تأیید کنید که نحاسی به دامنه در حد ناقص است؛ میانگین احتمال کناره‌گیری گامباردی از شغلش بعد از سه سال فقط اندکی بیش از دو سال است. باین‌حال، توصیف بی‌حسی به دامنه به‌جاست چون اختلاف فقط اندکی از چیزی است که باید.

3. Daniel Kahneman et al., eds., *Judgment Under Uncertainty: Heuristics and Biases* (New York: Cambridge University Press, 1982), chap. 6; Daniel Kahneman and Amos Tversky, "On the Psychology of Prediction," *Psychological Review* 80, no. 4 (1973): 237–251.

۴. برای مثال، ببینید:

Steven N. Kaplan and Bernadette A. Minton, "How Has CEO Turnover Changed?" *International Review of Finance* 12, no. 1 (2012): 57–87. See also Dirk Jenter and Katharina Lewellen, "Performance-Induced CEO Turnover," Harvard Law School Forum on Corporate Governance, September 2, 2020, <https://corpgov.law.harvard.edu/2020/09/02/performance-induced-ceo-turnover>.

5. J. W. Rinzler, *The Making of Star Wars: Return of the Jedi: The Definitive Story* (New York: Del Rey, 2013), 64.

6. Cass Sunstein, *The World According to Star Wars* (New York: HarperCollins, 2016).

۷. در اینجا تأکیدمان بر حالت ساده‌ای است که وقتی قضاوت آغاز می‌شود پیش‌قضاوتی موجود است. در واقع، حتی در نبود چنین پیش‌قضاوتی هم ممکن است با انباشت شواهد نوعی سوگیری به سمت نتیجه‌ای خاص شکل بگیرد و این به سبب گرایش ما به سادگی و انسجام است. وقتی نتیجه‌ای محتاطانه و اولیه شکل می‌گیرد، سوگیری تأیید جمع‌آوری و تفسیر شواهد جدید را به‌نفع خودش می‌گرداند.

۸. این پدیده را سوگیری باور^۱ نامیده‌اند. ببینید:

¹ Belief bias

- J. St. B. T. Evans, Julie L. Barson, and Paul Pollard, "On the Conflict between Logic and Belief in Syllogistic Reasoning," *Memory & Cognition* 11, no. 3 (1983): 295–306.
9. Dan Ariely, George Loewenstein, and Drazen Prelec, "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences," *Quarterly Journal of Economics* 118, no. 1 (2003): 73–105.
10. Adam D. Galinsky and T. Mussweiler, "First Offers as Anchors: The Role of Perspective-Taking and Negotiator Focus," *Journal of Personality and Social Psychology* 81, no. 4 (2001): 657–669.
11. Solomon E. Asch, "Forming Impressions of Personality," *Journal of Abnormal and Social Psychology* 41, no. 3 (1946): 258–290, first used a series of adjectives in different orders to illustrate this phenomenon.
12. Steven K. Dallas et al., "Don't Count Calorie Labeling Out: Calorie Counts on the Left Side of Menu Items Lead to Lower Calorie Food Choices," *Journal of Consumer Psychology* 29, no. 1 (2019): 60–69.

فصل ۱۴

1. S. S. Stevens, "On the Operation Known as Judgment," *American Scientist* 54, no. 4 (December 1966): 385–401.

استفاده ما از تعبیر «هماندسازی» موسع‌تر از کاربرد استونیز است. کاربرد او محدود به مقیاس‌های نسبی است که در فصل ۱۵ به آن‌ها می‌پردازیم.

۲. این مثال را اول بار دنیل کانمن طرح کرده است:

Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011).

3. Daniel Kahneman and Amos Tversky, "On the Psychology of Prediction," *Psychological Review* 80 (1973): 237–251.
4. G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review* (1956): 63–97.
5. R. D. Goffin and J. M. Olson, "Is It All Relative? Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others," *Perspectives on Psychological Science* 6 (2011): 48–60.

فصل ۱۵

1. Daniel Kahneman, David Schkade, and Cass Sunstein, "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," *Journal of Risk and Uncertainty* 16 (1998): 49–86, <https://link.springer.com/article/10.1023/A:1007710408413>; and Cass Sunstein, Daniel Kahneman, and David Schkade, "Assessing Punitive Damages (with Notes on

Cognition and Valuation in Law)," *Yale Law Journal* 107, no. 7 (May 1998): 2071–2153.

اکسان^۱ هزینه‌های تحقیق را یک‌باره و در یک مرحله پوشش داد، اما نه به این صورت که شرکت به محققان حقوق بدهد و نه به این صورت که کنترلی بر داده‌ها داشته باشد یا حتی پیش از انتشار آن‌ها در ژورنال‌های علمی از آن‌ها مطلع باشد.

2. A. Keane and P. McKeown, *The Modern Law of Evidence* (New York: Oxford University Press, 2014).

3. Andrew Mauboussin and Michael J. Mauboussin, "If You Say Something Is 'Likely,' How Likely Do People Think It Is?," *Harvard Business Review*, July 3, 2018.

4. *BMW v. Gore*, 517 U.S. 559 (1996),
<https://supreme.justia.com/cases/federal/us/517/559>.

۵. برای مباحثی درباره نقش عاطفه در قضاوت‌های اخلاقی، ببینید:

J. Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001): 814–834; Joshua Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (New York: Penguin Press, 2014).

۶. نظر به مقدار زیاد نويز در این امتیازدهی‌ها، شاید از هم‌بستگی بسیار بالای (۰.۹۸) بین قضاوت‌های انزجار و قصد تنبیهی که مؤید فرضیه انزجار هستند گیج شوید. وقتی به خاطر بیاورید که هم‌بستگی بین میانگین قضاوت‌ها محاسبه شده است این معما حل می‌شود. در میانگینی از ۱۰۰ قضاوت، نويز (انحراف معیار قضاوت‌ها) بخش بر ۱۰ می‌شود. وقتی قضاوت‌های زیادی تجمیع شوند دیگر نويز در کار نیست. فصل ۲۱ را ببینید.

7. S. S. Stevens, *Psychophysics: Introduction to Its Perceptual, Neural and Social Prospects* (New York: John Wiley & Sons, 1975).

8. Dan Ariely, George Loewenstein, and Drazen Prelec, "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences," *Quarterly Journal of Economics* 118, no. 1 (2003): 73–106.

۹. تبدیل به رتبه‌بندی مستلزم از دست دادن اطلاعات است، چون فاصله بین قضاوت‌ها حفظ نمی‌شود. فرض کنید فقط سه پرونده داریم و یک عضو هیئت منصفه خسارت‌های ۱۰ میلیون دلار، ۲ میلیون دلار، و ۱ میلیون دلار را توصیه می‌کند. به‌وضوح این فرد قصد دارد که تفاوت قصد تنبیهی بین دو پرونده اول بیشتر از دو پرونده دوم باشد. اما وقتی این قضاوت‌ها را به رتبه‌بندی تبدیل می‌کنیم، تفاوت آن‌ها یکسان است - فقط یک رتبه تفاوت. با تبدیل قضاوت‌ها به امتیازهای استاندارد می‌توان این مشکل را حل کرد.

¹ Exxon

فصل ۱۶

1. R. Blake and N. K. Logothetis, "Visual competition," *Nature Reviews Neuroscience* 3 (2002) 13–21; M. A. Gernsbacher and M. E. Faust, "The Mechanism of Suppression: A Component of General Comprehension Skill," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17 (March 1991): 245–262; and M. C. Stites and K. D. Federmeier, "Subsequent to Suppression: Downstream Comprehension Consequences of Noun/Verb Ambiguity in Natural Reading," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41 (September 2015): 1497–1515.
2. D. A. Moore and D. Schatz, "The three faces of overconfidence," *Social and Personality Psychology Compass* 11, no. 8 (2017), article e12331.
3. S. Highhouse, A. Broadfoot, J. E. Yugo, and S. A. Devendorf, "Examining Corporate Reputation Judgments with Generalizability Theory," *Journal of Applied Psychology* 94 (2009): 782–789.

از اسکات‌های‌هاوس و الیسن برودفوت بابت در اختیار گذاشتن اصل داده‌ها و از جولیان پاریس بابت بعضی تحلیل‌های تکمیلی سپاسگزاریم.

4. P. J. Lamberson and Scott Page, "Optimal forecasting groups," *Management Science* 58, no. 4 (2012): 805–10. We thank Scott Page for drawing our attention to this source of pattern noise.
5. The work of Allport and Odbert (1936) on English personality-relevant vocabulary is cited in Oliver P. John and Sanjay Srivastava, "The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives," in *Handbook of Personality: Theory and Research*, 2nd ed., ed. L. Pervin and Oliver P. John (New York: Guilford, 1999).
6. Ian W. Eisenberg, Patrick G. Bissett, A. Zeynep Enkavi et al., "Uncovering the structure of self-regulation through data-driven ontology discovery," *Nature Communications* 10 (2019): 2319.
7. Walter Mischel, "Toward an integrative science of the person," *Annual Review of Psychology* 55 (2004): 1–22.

فصل ۱۷

۱. گرچه درباره نحوه تقسیم سوگیری و نوین قاعده‌ای کلی خود ندارد، نسبت‌هایی که در این شکل می‌بینیم حدوداً بازنمای بعضی مثال‌های واقعی و ساختگی است که پیش از این مرور کردیم. به طور مشخص، در این شکل سوگیری و نوین مساوی هستند (چنان که در پیش‌بینی‌های فروش شرکت خوش‌فروش دیدیم). توان دوم نوین تراز با ۳۷٪ توان دوم نوین سیستمی است (چنان که در مطالعه خسارت تنبیهی چنین بود). چنان که در تصویر مشخص است، توان دوم نوین موقعیتی حدود ۳۵٪ توان دوم نوین الگویی است.
۲. ارجاعات فصل «مقدمه» را ببینید.

Mark A. Lemley and Bhaven Sampat, "Examiner Characteristics and Patent Office Outcomes," *Review of Economics and Statistics* 94, no. 3 (2012): 817–827.

همچنين ببينيد:

Iain Cockburn, Samuel Kortum, and Scott Stern, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics," working paper 8980, June 2002, www.nber.org/papers/w8980; and Michael D. Frakes and Melissa F. Wasserman, "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data," *Review of Economics and Statistics* 99, no. 3 (July 2017): 550–563.

3. Joseph J. Doyle Jr., "Child Protection and Child Outcomes: Measuring the Effects of Foster Care," *American Economic Review* 95, no. 5 (December 2007): 1583–1610.

4. Andrew I. Schoenholtz, Jaya Ramji-Nogales, and Philip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication," *Stanford Law Review* 60, no. 2 (2007).

۵. اين مقدار بر اساس محاسبات ارائه شده در فصل ۶ تخمين زده شده است. در فصل ۶ واريانس برهم کنش ۲۳٪ از کل واريانس است. با اين فرض که احکام کيفري توزيع نرمال دارند، ميادگين اختلاف مطلق ميان دو مشاهده تصادفي ۱.۱۲۸ انحراف معيار است.

6. J. E. Martinez, B. Labree, S. Uddenberg, and A. Todorov, "Meaningful 'noise': Comparative judgments contain stable idiosyncratic contributions" (unpublished ms.).

7. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan, "Human Decisions and Machine Predictions," *Quarterly Journal of Economics* 133 (2018): 237–293.

۸. اين مدل برای هر قاضي هم رتبه‌بندی ۱۳۱۸۳۳ پرونده را ارائه کرد و هم آستانه‌ای را که فراتر از آن درخواست قرار وثيقه پذيرفته شود. نويز ترازی بازتاب بی‌ثباتی آستانه‌هاست، درحالی‌که نويز الگویی بازتاب بی‌ثباتی در رتبه‌بندی پرونده‌هاست.

9. Gregory Stoddard, Jens Ludwig, and Sendhil Mullainathan.

مکاتبة ایمیلی با مؤلفان، ژوئن-جولای ۲۰۲۰.

10. *Phil Rosenzweig*. *Left Brain, Right Stuff: How Leaders Make Winning Decisions* (New York: PublicAffairs, 2014).

بخش پنجم

فصل ۱۸

1. Albert E. Mannes et al., "The Wisdom of Select Crowds," *Journal of Personality and Social Psychology* 107, no. 2 (2014): 276–299; Jason Dana et al., "The Composition of Optimally Wise Crowds," *Decision Analysis* 12, no. 3 (2015): 130–143.
2. Briony D. Pulford, Andrew M. Colmna, Eike K. Buabang, and Eva M. Krockow, "The Persuasive Power of Knowledge: Testing the Confidence Heuristic," *Journal of Experimental Psychology: General* 147, no. 10 (2018): 1431–1444.
3. Nathan R. Kuncel and Sarah A. Hezlett, "Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions," *Current Directions in Psychological Science* 19, no. 6 (2010): 339–345.
4. Kuncel and Hezlett, "Fact and Fiction."
5. Frank L. Schmidt and John Hunter, "General Mental Ability in the World of Work: Occupational Attainment and Job Performance," *Journal of Personality and Social Psychology* 86, no. 1 (2004): 162.
6. Angela L. Duckworth, David Weir, Eli Tsukayama, and David Kwok, "Who Does Well in Life? Conscientious Adults Excel in Both Objective and Subjective Success," *Frontiers in Psychology* 3 (September 2012). For grit, see Angela L. Duckworth, Christopher Peterson, Michael D. Matthews, and Dennis Kelly, "Grit: Perseverance and Passion for Long-Term Goals," *Journal of Personality and Social Psychology* 92, no. 6 (2007): 1087–1101.
7. Richard E. Nisbett et al., "Intelligence: New Findings and Theoretical Developments," *American Psychologist* 67, no. 2 (2012): 130–159.
8. Schmidt and Hunter, "Occupational Attainment," 162.
9. Kuncel and Hezlett, "Fact and Fiction."
۱۰. این هم‌بستگی‌ها از فراتحلیل‌هایی به دست آمده‌اند که هم‌بستگی‌های ظاهری ناشی از خطای اندازه‌گیری در معیار و حدود دامنه را اصلاح می‌کنند. در بین محققان نزاعی هست که آیا این اصلاح‌ها در ارزش پیش‌گویانه توانایی ذهنی عمومی اغراق می‌کنند یا خیر. با این حال، چون همین نزاع‌های روشی درباره سایر عوامل پیش‌بینی‌گر هم صدق می‌کند، محققان عموماً اتفاق نظر دارند که توانایی ذهنی عمومی (در کنار آزمون‌های نمونه کار؛ فصل ۲۴ را ببینید) بهترین عامل پیش‌بینی‌گر موجود برای موفقیت شغلی هستند. ببینید:
- Kuncel and Hezlett, "Fact and Fiction."
11. Schmidt and Hunter, "Occupational Attainment," 162.
12. David Lubinski, "Exceptional Cognitive Ability: The Phenotype," *Behavior Genetics* 39, no. 4 (2009): 350–358.

13. Jonathan Wai, "Investigating America's Elite: Cognitive Ability, Education, and Sex Differences," *Intelligence* 41, no. 4 (2013): 203–211.
14. Keela S. Thomson and Daniel M. Oppenheimer, "Investigating an Alternate Form of the Cognitive Reflection Test," *Judgment and Decision Making* 11, no. 1 (2016): 99–113.
15. Gordon Pennycook et al., "Everyday Consequences of Analytic Thinking," *Current Directions in Psychological Science* 24, no. 6 (2015): 425–432.
16. Gordon Pennycook and David G. Rand, "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning," *Cognition* 188 (June 2018): 39–50.
17. Nathaniel Barr et al., "The Brain in Your Pocket: Evidence That Smartphones Are Used to Supplant Thinking," *Computers in Human Behavior* 48 (2015): 473–480.
18. Niraj Patel, S. Glenn Baker, and Laura D. Scherer, "Evaluating the Cognitive Reflection Test as a Measure of Intuition/Reflection, Numeracy, and Insight Problem Solving, and the Implications for Understanding Real-World Judgments and Beliefs," *Journal of Experimental Psychology: General* 148, no. 12 (2019): 2129–2153.
19. John T. Cacioppo and Richard E. Petty, "The Need for Cognition," *Journal of Personality and Social Psychology* 42, no. 1 (1982): 116–131.
20. Stephen M. Smith and Irwin P. Levin, "Need for Cognition and Choice Framing Effects," *Journal of Behavioral Decision Making* 9, no. 4 (1996): 283–290.
21. Judith E. Rosenbaum and Benjamin K. Johnson, "Who's Afraid of Spoilers? Need for Cognition, Need for Affect, and Narrative Selection and Enjoyment," *Psychology of Popular Media Culture* 5, no. 3 (2016): 273–289.
22. Wandl Bruine De Bruin et al., "Individual Differences in Adult Decision-Making Competence," *Journal of Personality and Social Psychology* 92, no. 5 (2007): 938–956.
23. Heather A. Butler, "Halpern Critical Thinking Assessment Predicts Real-World Outcomes of Critical Thinking," *Applied Cognitive Psychology* 26, no. 5 (2012): 721–729.
24. Uriel Haran, Ilana Ritov, and Barbara Mellers, "The Role of Actively Open-Minded Thinking in Information Acquisition, Accuracy, and Calibration," *Judgment and Decision Making* 8, no. 3 (2013): 188–201.
25. Haran, Ritov, and Mellers, "Role of Actively Open-Minded Thinking."
26. J. Baron, "Why Teach Thinking? An Essay," *Applied Psychology: An International Review* 42 (1993): 191–214; J. Baron, *The Teaching of Thinking: Thinking and Deciding*, 2nd ed. (New York: Cambridge University Press, 1994), 127–148.

فصل ۱۹

۱. برای مروری فوق‌العاده، ببینید:

Jack B. Soll et al., "A User's Guide to Debiasing," in *The Wiley Blackwell Handbook of Judgment and Decision Making*, ed. Gideon Keren and George Wu, vol. 2 (New York: John Wiley & Sons, 2015), 684.

2. HM Treasury, *The Green Book: Central Government Guidance on Appraisal and Evaluation* (London: UK Crown, 2018),
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf.

3. Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New Haven, CT: Yale University Press, 2008).

4. Ralph Hertwig and Till Grüne-Yanoff, "Nudging and Boosting: Steering or Empowering Good Decisions," *Perspectives on Psychological Science* 12, no. 6 (2017).

5. Geoffrey T. Fong et al., "The Effects of Statistical Training on Thinking About Everyday Problems," *Cognitive Psychology* 18, no. 3 (1986): 253–292.

6. Willem A. Wagenaar and Gideon B. Keren, "Does the Expert Know? The Reliability of Predictions and Confidence Ratings of Experts," *Intelligent Decision Support in Process Environments* (1986): 87–103.

7. Carey K. Morewedge et al., "Debiasing Decisions: Improved Decision Making with a Single Training Intervention," *Policy Insights from the Behavioral and Brain Sciences* 2, no. 1 (2015): 129–140.

8. Anne-Laure Sellier et al., "Debiasing Training Transfers to Improve Decision Making in the Field," *Psychological Science* 30, no. 9 (2019): 1371–1379.

9. Emily Pronin et al., "The Bias Blind Spot: Perceptions of Bias in Self Versus Others," *Personality and Social Psychology Bulletin* 28, no. 3 (2002): 369–381.

10. Daniel Kahneman, Dan Lovallo, and Olivier Sibony, "Before You Make That Big Decision...," *Harvard Business Review* 89, no. 6 (June 2011): 50–60.

11. Atul Gawande, *Checklist Manifesto: How to Get Things Right* (New York: Metropolitan Books, 2010).

12. Office of Information and Regulatory Affairs, "Agency Checklist: Regulatory Impact Analysis," no date,
www.whitehouse.gov/sites/whitehouse.gov/files/omb/inforeg/inforeg/regpol/RIA_Checklist.pdf.

۱۳. این چک‌لیست تا حدی برگرفته از دنیل کانمن و دیگران است:

Daniel Kahneman et al., "Before You Make That Big Decision," *Harvard Business Review*.

۱۴. ببینید: کتاب چک لیست، نوشتهٔ آتول گوانده. نشر نوین.

فصل ۲۰

1. R. Stacey, "A Report on the Erroneous Fingerprint Individualisation in the Madrid Train Bombing Case," *Journal of Forensic Identification* 54 (2004): 707–718.
2. Michael Specter, "Do Fingerprints Lie?," *The New Yorker*, May 27, 2002. Emphasis added.
3. I. E. Dror and R. Rosenthal, "Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts," *Journal of Forensic Science* 53 (2008): 900–903.
4. I. E. Dror, D. Charlton, and A. E. Péron, "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications," *Forensic Science International* 156 (2006): 74–78.
5. I. E. Dror and D. Charlton, "Why Experts Make Errors," *Journal of Forensic Identification* 56 (2006): 600–616.
6. I. E. Dror and S. A. Cole, "The Vision in 'Blind' Justice: Expert Perception, Judgment, and Visual Cognition in Forensic Pattern Recognition," *Psychonomic Bulletin and Review* 17 (2010): 161–167, 165. See also I. E. Dror, "A Hierarchy of Expert Performance (HEP)," *Journal of Applied Research in Memory and Cognition* (2016): 1–6.
7. I. E. Dror et al., "Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison," *Forensic Science International* 208 (2011): 10–17.
8. B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. A. Buscaglia, "Changes in Latent Fingerprint Examiners' Markup Between Analysis and Comparison," *Forensic Science International* 247 (2015): 54–61.
9. I. E. Dror and G. Hampikian, "Subjectivity and Bias in Forensic DNA Mixture Interpretation," *Science and Justice* 51 (2011): 204–208.
10. M. J. Saks, D. M. Risinger, R. Rosenthal, and W. C. Thompson, "Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States," *Science Justice Journal of Forensic Science Society* 43 (2003): 77–90.
11. President's Council of Advisors on Science and Technology (PCAST), Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods (Washington, DC: Executive Office of the President, PCAST, 2016).

12. Stacey, "Erroneous Fingerprint."
13. Dror and Cole, "Vision in 'Blind' Justice."
14. I. E. Dror, "Biases in Forensic Experts," *Science* 360 (2018): 243.
15. Dror and Charlton, "Why Experts Make Errors."
16. B. T. Ulery, R. A. Hicklin, J. A. Buscaglia, and M. A. Roberts, "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners," *PLoS One* 7 (2012).
17. Innocence Project, "Overturning Wrongful Convictions Involving Misapplied Forensics," *Misapplication of Forensic Science* (2018): 1–7, www.innocenceproject.org/causes/misapplication-forensic-science. See also S. M. Kassin, I. E. Dror, J. Kukucka, and L. Butt, "The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions," *Journal of Applied Research in Memory and Cognition* 2 (2013): 42–52.
18. PCAST, Report to the President .
19. B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Accuracy and Reliability of Forensic Latent Fingerprint Decisions," *Proceedings of the National Academy of Sciences* 108 (2011): 7733–7738.
20. (PCAST), Report to the President , p. 95. Emphasis in original.
21. Igor Pacheco, Brian Cerchiai, and Stephanie Stoiloff, "Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Examinations," final report, Miami-Dade Police Department Forensic Services Bureau, 2014, www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf.
22. B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. A. Buscaglia, "Factors Associated with Latent Fingerprint Exclusion Determinations," *Forensic Science International* 275 (2017): 65–75.
23. R. N. Haber and I. Haber, "Experimental Results of Fingerprint Comparison Validity and Reliability: A Review and Critical Analysis," *Science & Justice* 54 (2014): 375–389.
24. Dror, "Hierarchy of Expert Performance," 3.
25. M. Leadbetter, letter to the editor, *Fingerprint World* 33 (2007): 231.
26. L. Butt, "The Forensic Confirmation Bias: Problems, Perspectives and Proposed Solutions—Commentary by a Forensic Examiner," *Journal of Applied Research in Memory and Cognition* 2 (2013): 59–60. Emphasis added.
27. Stacey, "Erroneous Fingerprint," 713. Emphasis added.

28. J. Kukucka, S. M. Kassir, P. A. Zapf, and I. E. Dror, "Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners," *Journal of Applied Research in Memory and Cognition* 6 (2017).

29. I. E. Dror et al., letter to the editor: "Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making," *Journal of Forensic Science* 60 (2015): 1111–1112.

فصل ۲۱

1. Jeffrey A. Frankel, "Over-optimism in Forecasts by Official Budget Agencies and Its Implications," working paper 17239, National Bureau of Economic Research, December 2011, www.nber.org/papers/w17239.

2. H. R. Arkes, "Overconfidence in Judgmental Forecasting," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, ed. Jon Scott Armstrong, vol. 30, International Series in Operations Research & Management Science (Boston: Springer, 2001).

3. Itzhak Ben-David, John Graham, and Campbell Harvey, "Managerial Miscalibration," *The Quarterly Journal of Economics* 128, no. 4 (November 2013): 1547–1584.

4. T. R. Stewart, "Improving Reliability of Judgmental Forecasts," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, ed. Jon Scott Armstrong, vol. 30, International Series in Operations Research & Management Science (Boston: Springer, 2001) (hereafter cited as *Principles of Forecasting*), 82.

5. Theodore W. Ruger, Pauline T. Kim, Andrew D. Martin, and Kevin M. Quinn, "The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decision-Making," *Columbia Law Review* 104 (2004): 1150–1209.

6. Cass Sunstein, "Maximin," *Yale Journal of Regulation* (draft; May 3, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3476250.

۷. برای مثال‌های متنوع، ببینید:

Armstrong, *Principles of Forecasting*.

8. Jon Scott Armstrong, "Combining Forecasts," in *Principles of Forecasting*, 417–439.

9. T. R. Stewart, "Improving Reliability of Judgmental Forecasts," in *Principles of Forecasting*, 95.

10. Armstrong, "Combining Forecasts."

11. Albert E. Mannes et al., "The Wisdom of Select Crowds," *Journal of Personality and Social Psychology* 107, no. 2 (2014): 276–299.

12. Justin Wolfers and Eric Zitzewitz, "Prediction Markets," *Journal of Economic Perspectives* 18 (2004): 107–126.
13. Cass R. Sunstein and Reid Hastie, *Wiser: Getting Beyond Groupthink to Make Groups Smarter* (Boston: Harvard Business Review Press, 2014).
14. Gene Rowe and George Wright, "The Delphi Technique as a Forecasting Tool: Issues and Analysis," *International Journal of Forecasting* 15 (1999): 353–375. See also Dan Bang and Chris D. Frith, "Making Better Decisions in Groups," *Royal Society Open Science* 4, no. 8 (2017).
15. R. Hastie, "Review Essay: Experimental Evidence on Group Accuracy," in B. Grofman and G. Guillermo, eds., *Information Pooling and Group Decision Making* (Greenwich, CT: JAI Press, 1986), 129–157.
16. Andrew H. Van De Ven and André L. Delbecq, "The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes," *Academy of Management Journal* 17, no. 4 (2017).
17. Superforecasting, 95.
18. Superforecasting, 231.
19. Superforecasting, 273.
20. Ville A. Satopää, Marat Salikhov, Philip E. Tetlock, and Barb Mellers, "Bias, Information, Noise: The BIN Model of Forecasting," February 19, 2020, 23, <https://dx.doi.org/10.2139/ssrn.3540864>.
21. Satopää et al., "Bias, Information, Noise," 23.
22. Satopää et al., 22.
23. Satopää et al., 24.
24. Clinton P. Davis-Stober, David V. Budescu, Stephen B. Broomell, and Jason Dana. "The composition of optimally wise crowds." *Decision Analysis* 12, no. 3 (2015): 130–143.

فصل ۲۲

1. Laura Horton et al., "Development and Assessment of Inter- and Intra-Rater Reliability of a Novel Ultrasound Tool for Scoring Tendon and Sheath Disease: A Pilot Study," *Ultrasound* 24, no. 3 (2016): 134, www.ncbi.nlm.nih.gov/pmc/articles/PMC5105362.
2. Laura C. Collins et al., "Diagnostic Agreement in the Evaluation of Image-guided Breast Core Needle Biopsies," *American Journal of Surgical Pathology* 28 (2004): 126, https://journals.lww.com/ajsp/Abstract/2004/01000/Diagnostic_Agreement_in_the_Evaluation_of.15.aspx.

3. Julie L. Fierro et al., "Variability in the Diagnosis and Treatment of Group A Streptococcal Pharyngitis by Primary Care Pediatricians," *Infection Control and Hospital Epidemiology* 35, no. S3 (2014): S79, www.jstor.org/stable/10.1086/677820.
4. Diabetes Tests, Centers for Disease Control and Prevention, <https://www.cdc.gov/diabetes/basics/getting-tested.html> (last accessed January 15, 2020).
5. Joseph D. Kronz et al., "Mandatory Second Opinion Surgical Pathology at a Large Referral Hospital," *Cancer* 86 (1999): 2426, [https://onlinelibrary.wiley.com/doi/full/10.1002/\(SICI\)1097-0142\(19991201\)86:11%3C2426::AID-CNCR34%3E3.0.CO;2-3](https://onlinelibrary.wiley.com/doi/full/10.1002/(SICI)1097-0142(19991201)86:11%3C2426::AID-CNCR34%3E3.0.CO;2-3).

۶. بیشتر مطالب آن را می‌توان آنلاین یافت. کلیات آن در حجم کتاب نیز در دسترس است:

Dartmouth Medical School, The Quality of Medical Care in the United States: A Report on the Medicare Program; the Dartmouth Atlas of Health Care 1999 (American Hospital Publishers, 1999).

۷. برای مثال، ببینید:

OECD, *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance?* (Paris: OECD Publishing, 2014), 137–169; Michael P. Hurley et al., "Geographic Variation in Surgical Outcomes and Cost Between the United States and Japan," *American Journal of Managed Care* 22 (2016): 600, www.ajmc.com/journals/issue/2016/2016-vol22-n9/geographic-variation-in-surgical-outcomes-and-cost-between-the-united-states-and-japan; and John Appleby, Veena Raleigh, Francesca Frosini, Gwyn Bevan, Haiyan Gao, and Tom Lyscom, *Variations in Health Care: The Good, the Bad and the Inexplicable* (London: The King's Fund, 2011), www.kingsfund.org.uk/sites/default/files/Variations-in-health-care-good-bad-inexplicable-report-The-Kings-Fund-April-2011.pdf.

8. David C. Chan Jr. et al., "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," National Bureau of Economic Research, NBER Working Paper No. 26467, November 2019, www.nber.org/papers/w26467.
9. P. J. Robinson, "Radiology's Achilles' Heel: Error and Variation in the Interpretation of the Röntgen Image," *British Journal of Radiology* 70 (1997): 1085, www.ncbi.nlm.nih.gov/pubmed/9536897. A relevant study is Yusuke Tsugawa et al., "Physician Age and Outcomes in Elderly Patients in Hospital in the US: Observational Study," *BMJ* 357 (2017), www.bmj.com/content/357/bmj.j1797.

طبق نتیجه حاصل از این تحقیق، هرچه پزشکان از دوره آموزشی خود فاصله می‌گیرند نتایجشان بدتر می‌شود. نتیجه می‌شود که با نوعی بده‌بستان مواجه هستیم: یک سو کسب تجربه که به‌مرور سال‌ها کار

عملی به دست می‌آید و سوی دیگر آشنایی با آخرین یافته‌ها و خطوط راهنما. این تحقیق نتیجه می‌گیرد که بهترین نتایج از پزشکانی است که تازه چند سالی است از دوره رزیدنسی گذشته‌اند و این زمانی است که هنوز شواهد را در خاطر دارند که در کسب تجربه نوعی بدهستان وجود دارد.

10. Robinson, "Radiology's Achilles' Heel."

۱۱. کاپا نیز مثل ضریب هم‌بستگی می‌تواند منفی باشد، گرچه در عمل به ندرت رخ می‌دهد. در ادامه یک وصف از معنای آمارهای مختلف کاپا را می‌بینید: «ضعیف ($\kappa = 0.00$ to 0.20)، قابل قبول ($\kappa = 0.21$ to 0.40)، متوسط ($\kappa = 0.41$ to 0.60)، اساسی ($\kappa = 0.61$ to 0.80)، و نزدیک به کامل ($\kappa > 0.80$)».

(Ron Wald, Chaim M. Bell, Rosane Nisenbaum, Samuel Perrone, Orfeas Liangos, Andreas Laupacis, and Bertrand L. Jaber, "Interobserver Reliability of Urine Sediment Interpretation," *Clinical Journal of the American Society of Nephrology* 4, no. 3 [March 2009]: 567–571, <https://cjasn.asnjournals.org/content/4/3/567>.)

12. Howard R. Strasberg et al., "Inter-Rater Agreement Among Physicians on the Clinical Significance of Drug-Drug Interactions," *AMIA Annual Symposium Proceedings* (2013): 1325, www.ncbi.nlm.nih.gov/pmc/articles/PMC3900147.

13. Wald et al., "Interobserver Reliability of Urine Sediment Interpretation," <https://cjasn.asnjournals.org/content/4/3/567>.

14. Juan P. Palazzo et al., "Hyperplastic Ductal and Lobular Lesions and Carcinomas in Situ of the Breast: Reproducibility of Current Diagnostic Criteria Among Community- and Academic-Based Pathologists," *Breast Journal* 4 (2003): 230, www.ncbi.nlm.nih.gov/pubmed/21223441.

15. Rohit K. Jain et al., "Atypical Ductal Hyperplasia: Interobserver and Intraobserver Variability," *Modern Pathology* 24 (2011): 917, www.nature.com/articles/modpathol201166.

16. Alex C. Speciale et al., "Observer Variability in Assessing Lumbar Spinal Stenosis Severity on Magnetic Resonance Imaging and Its Relation to Cross-Sectional Spinal Canal Area," *Spine* 27 (2002): 1082, www.ncbi.nlm.nih.gov/pubmed/12004176.

17. Centers for Disease Control and Prevention, "Heart Disease Facts," accessed June 16, 2020, www.cdc.gov/heartdisease/facts.htm.

18. Timothy A. DeRouen et al., "Variability in the Analysis of Coronary Arteriograms," *Circulation* 55 (1977): 324, www.ncbi.nlm.nih.gov/pubmed/832349.

19. Olaf Buchwoltz et al., "Interobserver Variability in the Diagnosis of Minimal and Mild Endometriosis," *European Journal of Obstetrics & Gynecology and Reproductive Biology* 122 (2005): 213, [www.ejog.org/article/S0301-2115\(05\)00059-X/pdf](http://www.ejog.org/article/S0301-2115(05)00059-X/pdf).

20. Jean-Pierre Zellweger et al., "Intra-observer and Overall Agreement in the Radiological Assessment of Tuberculosis," *International Journal of Tuberculosis &*

Lung Disease 10 (2006): 1123, www.ncbi.nlm.nih.gov/pubmed/17044205. For "fair" interrater agreement, see Yanina Balabanova et al., "Variability in Interpretation of Chest Radiographs Among Russian Clinicians and Implications for Screening Programmes: Observational Study," *BMJ* 331 (2005): 379, www.bmj.com/content/331/7513/379.short.

21. Shinsaku Sakurada et al., "Inter-Rater Agreement in the Assessment of Abnormal Chest X-Ray Findings for Tuberculosis Between Two Asian Countries," *BMC Infectious Diseases* 12, article 31 (2012), <https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-12-31>.

22. Evan R. Farmer et al., "Discordance in the Histopathologic Diagnosis of Melanoma and Melanocytic Nevi Between Expert Pathologists," *Human Pathology* 27 (1996): 528, www.ncbi.nlm.nih.gov/pubmed/8666360.

23. Alfred W. Kopf, M. Mintzis, and R. S. Bart, "Diagnostic Accuracy in Malignant Melanoma," *Archives of Dermatology* 111 (1975): 1291, www.ncbi.nlm.nih.gov/pubmed/1190800.

24. Maria Miller and A. Bernard Ackerman, "How Accurate Are Dermatologists in the Diagnosis of Melanoma? Degree of Accuracy and Implications," *Archives of Dermatology* 128 (1992): 559, <https://jamanetwork.com/journals/jamadermatology/fullarticle/554024>.

25. Craig A. Beam et al., "Variability in the Interpretation of Screening Mammograms by US Radiologists," *Archives of Internal Medicine* 156 (1996): 209, www.ncbi.nlm.nih.gov/pubmed/8546556.

26. P. J. Robinson et al., "Variation Between Experienced Observers in the Interpretation of Accident and Emergency Radiographs," *British Journal of Radiology* 72 (1999): 323, www.birpublications.org/doi/pdf/10.1259/bjr.72.856.10474490.

27. Katherine M. Detre et al., "Observer Agreement in Evaluating Coronary Angiograms," *Circulation* 52 (1975): 979, www.ncbi.nlm.nih.gov/pubmed/1102142.

28. Horton et al., "Inter- and Intra-Rater Reliability"; and Megan Banky et al., "Inter- and Intra-Rater Variability of Testing Velocity When Assessing Lower Limb Spasticity," *Journal of Rehabilitation Medicine* 51 (2019), www.medicaljournals.se/jrm/content/abstract/10.2340/16501977-2496.

29. Esther Y. Hsiang et al., "Association of Primary Care Clinic Appointment Time with Clinician Ordering and Patient Completion of Breast and Colorectal Cancer Screening," *JAMA Network Open* 51 (2019), <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2733171>.

30. Hengchen Dai et al., "The Impact of Time at Work and Time Off from Work on Rule Compliance: The Case of Hand Hygiene in Health Care," *Journal of Applied Psychology* 100 (2015): 846, www.ncbi.nlm.nih.gov/pubmed/25365728.
31. Ali S. Raja, "The HEART Score Has Substantial Interrater Reliability," *NEJM J Watch*, December 5, 2018, www.jwatch.org/na47998/2018/12/05/heart-score-has-substantial-interrater-reliability (reviewing Colin A. Gershon et al., "Inter-rater Reliability of the HEART Score," *Academic Emergency Medicine* 26 [2019]: 552).
32. Jean-Pierre Zellweger et al., "Intra-observer and Overall Agreement in the Radiological Assessment of Tuberculosis," *International Journal of Tuberculosis & Lung Disease* 10 (2006): 1123, www.ncbi.nlm.nih.gov/pubmed/17044205; Ibrahim Abubakar et al., "Diagnostic Accuracy of Digital Chest Radiography for Pulmonary Tuberculosis in a UK Urban Population," *European Respiratory Journal* 35 (2010): 689, <https://erj.ersjournals.com/content/35/3/689.short>.
33. Michael L. Barnett et al., "Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians," *JAMA Network Open* 2 (2019): e19009, <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2726709>; Kimberly H. Allison et al., "Understanding Diagnostic Variability in Breast Pathology: Lessons Learned from an Expert Consensus Review Panel," *Histopathology* 65 (2014): 240, <https://onlinelibrary.wiley.com/doi/abs/10.1111/his.12387>.
34. Babak Ehteshami Bejnordi et al., "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer," *JAMA* 318 (2017): 2199, <https://jamanetwork.com/journals/jama/fullarticle/2665774>.
35. Varun Gulshan et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA* 316 (2016): 2402, <https://jamanetwork.com/journals/jama/fullarticle/2588763>.
36. Mary Beth Massat, "A Promising Future for AI in Breast Cancer Screening," *Applied Radiology* 47 (2018): 22, www.appliedradiology.com/articles/a-promising-future-for-ai-in-breast-cancer-screening; Alejandro Rodriguez-Ruiz et al., "Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison with 101 Radiologists," *Journal of the National Cancer Institute* 111 (2019): 916, <https://academic.oup.com/jnci/advance-article-abstract/doi/10.1093/jnci/djy222/5307077>.
37. Apgar Score, Medline Plus, <https://medlineplus.gov/ency/article/003402.htm> (last accessed February 4, 2020).
38. L. R. Foster et al., "The Interrater Reliability of Apgar Scores at 1 and 5 Minutes," *Journal of Investigative Medicine* 54, no. 1 (2006): 293, <https://jim.bmj.com/content/54/1/S308.4>.

39. Warren J. McIsaac et al., "Empirical Validation of Guidelines for the Management of Pharyngitis in Children and Adults," *JAMA* 291 (2004): 1587, www.ncbi.nlm.nih.gov/pubmed/15069046.
40. Emilie A. Ooms et al., "Mammography: Interobserver Variability in Breast Density Assessment" *Breast* 16 (2007): 568, www.sciencedirect.com/science/article/abs/pii/S0960977607000793.
41. Frances P. O'Malley et al., "Interobserver Reproducibility in the Diagnosis of Flat Epithelial Atypia of the Breast," *Modern Pathology* 19 (2006): 172, www.nature.com/articles/3800514.

۴۲. ببینید:

- Ahmed Aboraya et al., "The Reliability of Psychiatric Diagnosis Revisited," *Psychiatry (Edgmont)* 3 (2006): 41, www.ncbi.nlm.nih.gov/pmc/articles/PMC2990547. For an overview, see N. Kreitman, "The Reliability of Psychiatric Diagnosis," *Journal of Mental Science* 107 (1961): 876–886, www.cambridge.org/core/journals/journal-of-mental-science/article/reliability-of-psychiatric-diagnosis/92832FFA170F4FF41189428C6A3E6394.
43. Aboraya et al., "Reliability of Psychiatric Diagnosis Revisited," 43.
44. C. H. Ward et al., "The Psychiatric Nomenclature: Reasons for Diagnostic Disagreement," *Archives of General Psychiatry* 7 (1962): 198.
45. Aboraya et al., "Reliability of Psychiatric Diagnosis Revisited."
46. Samuel M. Lieblich, David J. Castle, Christos Pantelis, Malcolm Hopwood, Allan Hunter Young, and Ian P. Everall, "High Heterogeneity and Low Reliability in the Diagnosis of Major Depression Will Impair the Development of New Drugs," *British Journal of Psychiatry Open* 1 (2015): e5–e7, www.ncbi.nlm.nih.gov/pmc/articles/PMC5000492/pdf/bjporcpsych_1_2_e5.pdf.
47. Lieblich et al., "High Heterogeneity."

۴۸. ببینید:

- Elie Chéniaux et al., "The Diagnoses of Schizophrenia, Schizoaffective Disorder, Bipolar Disorder and Unipolar Depression: Interrater Reliability and Congruence Between DSM-IV and ICD-10," *Psychopathology* 42 (2009): 296–298, especially 293; and Michael Chmielewski et al., "Method Matters: Understanding Diagnostic Reliability in DSM-IV and DSM-5," *Journal of Abnormal Psychology* 124 (2015): 764, 768–769.
49. Aboraya et al., "Reliability of Psychiatric Diagnosis Revisited," 47.
50. Aboraya et al., 47.

۵۱. ببینید:

Chmielewski et al., "Method Matters."

۵۲. برای مثال، ببینید:

Helena Chmura Kraemer et al., "DSM-5: How Reliable Is Reliable Enough?," *American Journal of Psychiatry* 169 (2012): 13–15.

53. Lieblich et al., "High Heterogeneity."

54. Lieblich et al., "High Heterogeneity," e-5.

55. Lieblich et al., e-5.

56. Lieblich et al., e-6.

57. Aboraya et al., "Reliability of Psychiatric Diagnosis Revisited," 47.

58. Aboraya et al.

59. Aboraya et al.

۶۰. در اینجا می‌توان چند نکته عبرت‌آمیز ارزشمند را یافت:

Christopher Worsham and Anupam B. Jena, "The Art of Evidence-Based Medicine," *Harvard Business Review*, January 30, 2019, <https://hbr.org/2019/01/the-art-of-evidence-based-medicine>.

فصل ۲۳

1. Jena McGregor, "Study Finds That Basically Every Single Person Hates Performance Reviews," *Washington Post*, January 27, 2014.

۲. تحول دیجیتالی که در بسیاری از سازمان‌ها در جریان است ممکن است در این باره امکان‌های جدیدی ایجاد کند. در سطح نظری، شرکت‌ها می‌توانند بسیاری اطلاعات ریزریز و در لحظه درباره عملکرد همه کارکنان گردآوری کنند. ممکن است با استفاده از این داده‌ها ارزشیابی عملکرد کاملاً الگوریتمیک برای بعضی موقعیت‌های شغلی فراهم شود. با این حال، در اینجا تمرکز ما بر قضاوت‌هایی است که نمی‌توان آن‌ها را به طور کامل از سنجش عملکرد حذف کرد. ببینید:

E. D. Pulakos, R. Mueller-Hanson, and S. Arad, "The Evolution of Performance Management: Searching for Value," *Annual Review of Organizational Psychology and Organizational Behavior* 6 (2018): 249–271.

3. S. E. Scullen, M. K. Mount, and M. Goff, "Understanding the Latent Structure of Job Performance Ratings," *Journal of Applied Psychology* 85 (2000): 956–970.

۴. مؤلفه کوچکی - در بعضی مطالعات ۱۰٪ از مجموع واریانس - چیزی است که محققان آن را تأثیر منظر امتیازدهنده یا سطح می‌نامند. در اینجا سطح یا تراز به معنای سطح و تراز سازمان است. نه نوبت تراز که ما تعریف کردیم. منظر امتیازدهنده نشان می‌دهد که در امتیازدهی به شخص رییس به نحو نظام‌مندی متفاوت از همتایان است و همتایان هم متفاوت از زیردستان. بنا بر تفسیری همدلانه از نتایج سیستم‌های امتیازدهی ۳۶۰ درجه، می‌توان استدلال کرد که این نوبت محسوب نمی‌شود. اگر افرادی که در سطوح مختلف

سازمان هستند ووجه مختلفی از عملکرد شخص را می‌بینند، قضاوت‌هایشان درباره آن شخص باید به نحو نظام‌مندی متفاوت باشد و امتیازدهی‌هایشان باید این مطلب را منعکس کند.

5. Scullen, Mount, and Goff, "Latent Structure"; C. Viswesvaran, D. S. Ones, and F. L. Schmidt, "Comparative Analysis of the Reliability of Job Performance Ratings," *Journal of Applied Psychology* 81 (1996): 557–574. G. J. Greguras and C. Robie, "A New Look at Within-Source Interrater Reliability of 360-Degree Feedback Ratings," *Journal of Applied Psychology* 83 (1998): 960–968; G. J. Greguras, C. Robie, D. J. Schleicher, and M. A. Goff, "A Field Study of the Effects of Rating Purpose on the Quality of Multisource Ratings," *Personnel Psychology* 56 (2003): 1–21; C. Viswesvaran, F. L. Schmidt, and D. S. Ones, "Is There a General Factor in Ratings of Job Performance? A Meta-Analytic Framework for Disentangling Substantive and Error Influences," *Journal of Applied Psychology* 90 (2005): 108–131; and B. Hoffman, C. E. Lance, B. Bynum, and W. A. Gentry, "Rater Source Effects Are Alive and Well After All," *Personnel Psychology* 63 (2010): 119–151.

6. K. R. Murphy, "Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance," *Industrial and Organizational Psychology* 1 (2008): 148–160, especially 151.

۷. در بحث از منشأهای نويز، این امکان را نادیده گرفتیم که نويز موردی ناشی از سوگیری‌های نظام‌مند در امتیازدهی کارکنان خاص یا گروه‌های خاصی از کارکنان باشد. هیچ‌یک از تحقیقاتی که توانستیم در باب بی‌ثباتی امتیازدهی‌ها به عملکرد بیایم امتیازات را با عملکرد «درست» که به نحو بیرونی ارزیابی شده است مقایسه نکرده‌اند.

8. E. D. Pulakos and R. S. O'Leary, "Why Is Performance Management Broken?," *Industrial and Organizational Psychology* 4 (2011): 146–164; M. M. Harris, "Rater Motivation in the Performance Appraisal Context: A Theoretical Framework," *Journal of Management* 20 (1994): 737–756; and K. R. Murphy and J. N. Cleveland, *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives* (Thousand Oaks, CA: Sage, 1995).

9. Greguras et al., "Field Study."

10. P. W. Atkins and R. E. Wood, "Self- Versus Others' Ratings as Predictors of Assessment Center Ratings: Validation Evidence for 360-Degree Feedback Programs," *Personnel Psychology* (2002).

11. Atkins and Wood, "Self- Versus Others' Ratings."

12. Olson and Davis, cited in Peter G. Dominick, "Forced Ranking: Pros, Cons and Practices," in *Performance Management: Putting Research into Action*, ed. James W. Smither and Manuel London (San Francisco: Jossey-Bass, 2009), 411–443.

13. Dominick, "Forced Ranking."

14. Barry R. Nathan and Ralph A. Alexander, "A Comparison of Criteria for Test Validation: A Meta-Analytic Investigation," *Personnel Psychology* 41, no. 3 (1988): 517–535.

۱۵. برگرفته از:

Richard D. Goffin and James M. Olson, "Is It All Relative? Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others," *Perspectives on Psychological Science* 6, no. 1 (2011): 48–60.

16. M. Buckingham and A. Goodall, "Reinventing Performance Management," *Harvard Business Review*, April 1, 2015, 1–16, doi:ISSN: 0017-8012.

17. Corporate Leadership Council, cited in S. Adler et al., "Getting Rid of Performance Ratings: Genius or Folly? A Debate," *Industrial and Organizational Psychology* 9 (2016): 219–252.

18. Pulakos, Mueller-Hanson, and Arad, "Evolution of Performance Management," 250.

19. A. Tavis and P. Cappelli, "The Performance Management Revolution," *Harvard Business Review*, October 2016, 1–17.

20. Frank J. Landy and James L. Farr, "Performance Rating," *Psychological Bulletin* 87, no. 1 (1980): 72–107.

21. D. J. Woehr and A. I. Huffcutt, "Rater Training for Performance Appraisal: A Quantitative Review," *Journal of Occupational and Organizational Psychology* 67 (1994): 189–205; S. G. Roch, D. J. Woehr, V. Mishra, and U. Kieszczynska, "Rater Training Revisited: An Updated Meta-Analytic Review of Frame-of-Reference Training," *Journal of Occupational and Organizational Psychology* 85 (2012): 370–395; and M. H. Tsai, S. Wee, and B. Koh, "Restructured Frame-of-Reference Training Improves Rating Accuracy," *Journal of Organizational Behavior* (2019): 1–18, doi:10.1002/job.2368.

۲۲. سمت چپ برگرفته از:

Richard Goffin and James M. Olson, "Is It All Relative? Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others," *Perspectives on Psychological Science* 6, no. 1 (2011): 48–60.

23. Roch et al., "Rater Training Revisited."

24. Ernest O'Boyle and Herman Aguinis, "The Best and the Rest: Revisiting the Norm of Normality of Individual Performance," *Personnel Psychology* 65, no. 1 (2012): 79–119; and Herman Aguinis and Ernest O'Boyle, "Star Performers in Twenty-First Century Organizations," *Personnel Psychology* 67, no. 2 (2014): 313–350.

فصل ۲۴

1. A. I. Huffcutt and S. S. Culbertson, "Interviews," in S. Zedeck, ed., *APA Handbook of Industrial and Organizational Psychology* (Washington, DC: American Psychological Association, 2010), 185–203.
2. N. R. Kuncel, D. M. Klieger, and D. S. Ones, "In Hiring, Algorithms Beat Instinct," *Harvard Business Review* 92, no. 5 (2014): 32.
3. R. E. Ployhart, N. Schmitt, and N. T. Tippins, "Solving the Supreme Problem: 100 Years of Selection and Recruitment at the *Journal of Applied Psychology*," *Journal of Applied Psychology* 102 (2017): 291–304.
4. M. McDaniel, D. Whetzel, F. L. Schmidt, and S. Maurer, "Meta Analysis of the Validity of Employment Interviews," *Journal of Applied Psychology* 79 (1994): 599–616; A. Huffcutt and W. Arthur, "Hunter and Hunter (1984) Revisited: Interview Validity for Entry-Level Jobs," *Journal of Applied Psychology* 79 (1994): 2; F. L. Schmidt and J. E. Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," *Psychology Bulletin* 124 (1998): 262–274; and F. L. Schmidt and R. D. Zimmerman, "A Counterintuitive Hypothesis About Employment Interview Validity and Some Supporting Evidence," *Journal of Applied Psychology* 89 (2004): 553–561. Note that validities are higher when certain subsets of studies are considered, especially if research uses performance ratings specifically created for this purpose, rather than existing administrative ratings.
5. S. Highhouse, "Stubborn Reliance on Intuition and Subjectivity in Employee Selection," *Industrial and Organizational Psychology* 1 (2008): 333–342; D. A. Moore, "How to Improve the Accuracy and Reduce the Cost of Personnel Selection," *California Management Review* 60 (2017): 8–17.
6. L. A. Rivera, "Hiring as Cultural Matching: The Case of Elite Professional Service Firms," *American Sociology Review* 77 (2012): 999–1022.
7. Schmidt and Zimmerman, "Counterintuitive Hypothesis"; Timothy A. Judge, Chad A. Higgins, and Daniel M. Cable, "The Employment Interview: A Review of Recent Research and Recommendations for Future Research," *Human Resource Management Review* 10 (2000): 383–406; and A. I. Huffcutt, S. S. Culbertson, and W. S. Weyhrauch, "Employment Interview Reliability: New Meta-Analytic Estimates by Structure and Format," *International Journal of Selection and Assessment* 21 (2013): 264–276.
8. M. R. Barrick et al., "Candidate Characteristics Driving Initial Impressions During Rapport Building: Implications for Employment Interview Validity," *Journal of Occupational and Organizational Psychology* 85 (2012): 330–352; M. R. Barrick, B. W. Swider, and G. L. Stewart, "Initial Evaluations in the Interview: Relationships with

Subsequent Interviewer Evaluations and Employment Offers," *Journal of Applied Psychology* 95 (2010): 1163.

9. G. L. Stewart, S. L. Dustin, M. R. Barrick, and T. C. Darnold, "Exploring the Handshake in Employment Interviews," *Journal of Applied Psychology* 93 (2008): 1139–1146.

10. T. W. Dougherty, D. B. Turban, and J. C. Callender, "Confirming First Impressions in the Employment Interview: A Field Study of Interviewer Behavior," *Journal of Applied Psychology* 79 (1994): 659–665.

11. J. Dana, R. Dawes, and N. Peterson, "Belief in the Unstructured Interview: The Persistence of an Illusion," *Judgment and Decision Making* 8 (2013): 512–520.

12. Nathan R. Kuncel et al., "Mechanical versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis," *Journal of Applied Psychology* 98, no. 6 (2013): 1060–1072.

13. Laszlo Bock, *interview with Adam Bryant*, The New York Times, June 19, 2013. See also Laszlo Bock, *Work Rules!: Insights from Inside Google That Will Transform How You Live and Lead* (New York: Hachette, 2015).

14. C. Fernández-Aráoz, "Hiring Without Firing," *Harvard Business Review*, July 1, 1999.

۱۵. برای راهنمایی ساده و سرراست به مصاحبه‌ساختارمند، ببینید:

Michael A. Campion, David K. Palmer, and James E. Campion, "Structuring Employment Interviews to Improve Reliability, Validity and Users' Reactions," *Current Directions in Psychological Science* 7, no. 3 (1998): 77–82.

16. J. Levashina, C. J. Hartwell, F. P. Morgeson, and M. A. Campion, "The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature," *Personnel Psychology* 67 (2014): 241–293.

17. McDaniel et al., "Meta Analysis"; Huffcutt and Arthur, "Hunter and Hunter (1984) Revisited"; Schmidt and Hunter, "Validity and Utility"; and Schmidt and Zimmerman, "Counterintuitive Hypothesis."

18. Schmidt and Hunter, "Validity and Utility."

19. Kahneman, *Thinking, Fast and Slow*, 229.

20. Kuncel, Klieger, and Ones, "Algorithms Beat Instinct." See also Campion, Palmer, and Campion, "Structuring Employment Interviews."

21. Dana, Dawes, and Peterson, "Belief in the Unstructured Interview."

فصل ۲۵

1. Daniel Kahneman, Dan Lovallo, and Olivier Sibony, "A Structured Approach to Strategic Decisions: Reducing Errors in Judgment Requires a Disciplined Process," *MIT Sloan Management Review* 60 (2019): 67–73.
2. Andrew H. Van De Ven and André Delbecq, "The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes," *Academy of Management Journal* 17, no. 4 (1974): 605–621. See also chapter 21.

بخش شش

1. Kate Stith and José A. Cabranes, *Fear of Judging: Sentencing Guidelines in the Federal Courts* (Chicago: University of Chicago Press, 1998), 177.

فصل ۲۶

1. Albert O. Hirschman, *The Rhetoric of Reaction: Perversity, Futility, Jeopardy* (Cambridge, MA: Belknap Press, 1991).
2. Stith and Cabranes, *Fear of Judging*.

۳. برای مثال، ببینید:

Three Strikes Basics, Stanford Law School, <https://law.stanford.edu/stanford-justice-advocacy-project/three-strikes-basics/>.

4. *428 U.S. 280 (1976)*.

5. Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).

6. Will Knight, "Biased Algorithms Are Everywhere, and No One Seems to Care," *MIT Technology Review*, July 12, 2017.

7. Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, May 23, 2016, www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. The claim of bias in this example is disputed, and different definitions of bias may lead to opposite conclusions. For views on this case and more broadly on the definition and measurement of algorithmic bias, see later note, "Exactly how to test."

8. Aaron Shapiro, "Reform Predictive Policing," *Nature* 541, no. 7638 (2017): 458–460.

۹. گرچه این نگرانی بیشتر در زمینه مدل‌های مبتنی بر هوش مصنوعی در صدر قرار گرفته است، مختص هوش مصنوعی نیست. از همان ۱۹۷۲، پل اسلوویک اشاره کرده بود که مدل‌کردن شهودها سوگیری‌های شناختی موجود در آن‌ها را حفظ و تحکیم و حتی تشدید می‌کند.

Paul Slovic, "Psychological Study of Human Judgment: Implications for Investment Decision Making," *Journal of Finance* 27 (1972): 779.

۱۰. برای مقدمه‌ای به این نزاع در زمینهٔ مشاجرات راجع به الگوریتم پیش‌بینی تکرار جرم COMPAS، ببینید:

Larson et al., "COMPAS Recidivism Algorithm"; William Dieterich et al., "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity," Northpointe, Inc., July 8, 2016, http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf; Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Science Advances* 4, no. 1 (2018): 1–6; Sam Corbett-Davies et al., "A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear," *Washington Post*, October 17, 2016, www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas; Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," *Big Data* 153 (2017): 5; and Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores," Leibniz International Proceedings in Informatics, January 2017.

فصل ۲۷

1. Tom R. Tyler, *Why People Obey the Law*, 2nd ed. (New Haven, CT: Yale University Press, 2020).
2. Cleveland Bd. of Educ. v. LaFleur, 414 U.S. 632 (1974).
3. Laurence H. Tribe, "Structural Due Process," *Harvard Civil Rights–Civil Liberties Law Review* 10, no. 2 (spring 1975): 269.
4. *Stith and Cabranes*, Fear of Judging, 177.

۵. برای مثال، ببینید:

Philip K. Howard, *The Death of Common Sense: How Law Is Suffocating America* (New York: Random House, 1995); and Philip K. Howard, *Try Common Sense: Replacing the Failed Ideologies of Right and Left* (New York: W. W. Norton & Company, 2019).

فصل ۲۸

1. *Facebook's Community Standards in 2020* 12. Hate Speech, Facebook: Community Standards, www.facebook.com/communitystandards/hate_speech.
2. Andrew Marantz, "Why Facebook Can't Fix Itself," *The New Yorker*, October 12, 2020.

منابع و یادداشت‌های کتاب نویز | نشر نوین

3. Jerry L. Mashaw, *Bureaucratic Justice* (New Haven, CT: Yale University Press, 1983).

4. David M. Trubek, "Max Weber on Law and the Rise of Capitalism," *Wisconsin Law Review* 720 (1972): 733, n. 22 (quoting Max Weber, *The Religion of China* [1951], 149).

[لینک کتاب نویز در سایت نشر نوین](#)

