Primary Sources
WP 1.0 – 8.0

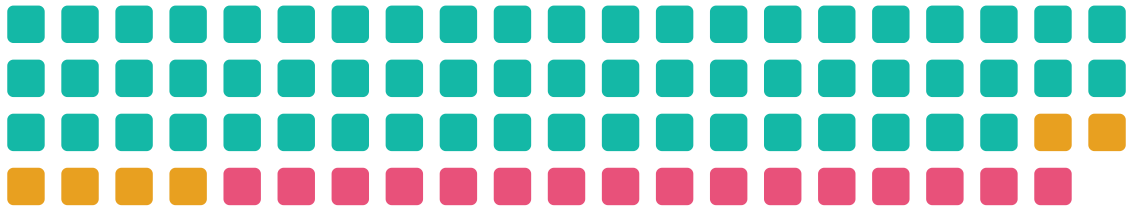
James E. Coleman

STRATEGIC DESIGN · WHITE PAPER SERIES · 2025

ABSTRACT *The consolidated bibliography for the white-paper series. Every externally sourced figure and framework in WP 1.0–8.0 is traced here to its primary source, with the methodology and caveats stated in full. Figures that could not be grounded in published research are labeled as modeled estimates rather than cited.*

FIG. — THE RECEIPTS, ONE SQUARE PER FIGURE — PRIMARY SOURCES: 61

Every externally-sourced figure in the **8 papers** traces to one of **61 primary sources** (teal). Where it couldn't, we said so: **6 figures were corrected** to match their source (amber), and **16 claims are labeled modeled estimates** (rose) — never dressed up as fact.



PRIMARY SOURCE · FIGURE CORRECTED · MODELED ESTIMATE

§ 01 Sourcing Methodology

How these sources *were handled.*

This bibliography accompanies a sourcing pass over the white-paper series. The method followed three rules, applied to every quantitative or empirical claim across the eight papers (WP 1.0–8.0):

> *"Cite every figure to its primary source. Where a figure did not match the source it came from, correct the figure and note the correction. Where no credible primary source exists, do not manufacture one — label the claim as a modeled estimate so the reader can tell the series' own models from externally verified facts."*
>
> — SOURCING PROTOCOL

The consequence is visible on the pages themselves. Six headline figures were corrected to align with their documented sources — for example, a "$12.9T" figure restated as the $12.9M per-organization figure Gartner actually reports, and a "300ms" latency threshold restated as the canonical 100ms limit. Across the series, sixteen claims for which no primary source could be found — among them WP 3.0's trust-formation timeline, transaction-loss rate, and fraud-concentration rate — are labeled "modeled estimate" rather than cited.

Frameworks original to the series — among them the Intelligence Architecture, the Translational Pipeline model, the Transaction-Trust Matrix, the Dual-Layer Architecture, the Augmentation Architecture, and the compounding-capability flywheel — are presented as original contributions, grounded where possible in the seminal literature of their fields. The sources below are organized by paper; each entry notes what it supports and, where relevant, its limitations.

§ 02 WP 1.0 — The Intelligence Gap

Organizational intelligence — *sources.*

Seven sources. Three are industry/analyst figures on data usage, data-quality cost, and data literacy (the first two secondary or self-reported, as noted); four are peer-reviewed or foundational works grounding the paper's thesis that converting data into decisions is an organizational capability.

1. [1]Forrester Research (Mike Gualtieri), analyst estimate, c. 2016 — between 60% and 73% of data within an enterprise goes unused for analytics. Widely cited; original report not publicly available. As reported, ITProToday.

   Secondary attribution — treat as an analyst estimate, not a measured figure.

2. [2]Gartner, “How to Improve Your Data Quality” (2021), figure attributed to Melody Chien, Sr. Director Analyst — poor data quality costs organizations an average of $12.9 million per year. gartner.com.

   Self-reported average from a Gartner customer-reference survey (154 organizations); a per-organization figure, not an economy-wide total.

3. [3]Accenture & Qlik, *The Human Impact of Data Literacy* (2020), The Data Literacy Project — only 21% of workers reported being confident in their data-literacy skills (i.e., ~79% are not). Qlik / Accenture press release.

4. [4]Hilbert M, López P. “The World’s Technological Capacity to Store, Communicate, and Compute Information.” *Science* 2011; 332(6025): 60–65. Empirical basis for the exponential growth of stored and transmitted data. doi:10.1126/science.1200970.

5. [5]Brynjolfsson E, Hitt LM, Kim HH. “Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance?” (2011). Firms adopting data-driven decision-making showed output and productivity 5–6% higher than expected. SSRN 1819486.

6. [6]Davenport TH, Harris JG. *Competing on Analytics: The New Science of Winning*. Harvard Business School Press, 2007. Foundational treatment of analytical capability as a source of competitive advantage.

7. [7]Pfeffer J, Sutton RI. “Evidence-Based Management.” *Harvard Business Review* 2006; 84(1): 62–74. On organizational cultures that privilege intuition and ideology over evidence. hbr.org.

§ 03 WP 2.0 — The Valley of Death

Translational research — *sources.*

Eleven sources, the most heavily evidenced paper in the series. They establish the translation lag and failure rate, the cost and funding structure of the pipeline, and the replication problem that degrades what enters it — including the NIH portfolio data that replaced the paper's original unsourced "one-tenth funding" estimate. The final two are the author's own translational-research work with practicing clinician-scientists, which supplies the series thruline its clinical anchor.

1. [1]Balas EA, Boren SA. “Managing Clinical Knowledge for Health Care Improvement.” *Yearbook of Medical Informatics* 2000; 9: 65–70. Origin of the “17 years / 14% of original research reaches patient benefit” estimate (hence ~86% does not). doi:10.1055/s-0038-1637943.

2. [2]Morris ZS, Wooding S, Grant J. “The answer is 17 years, what is the question: understanding time lags in translational research.” *Journal of the Royal Society of Medicine* 2011; 104(12): 510–520. Corroborates the ~17-year translation lag. doi:10.1258/jrsm.2011.110180.

3. [3]DiMasi JA, Grabowski HG, Hansen RW. “Innovation in the pharmaceutical industry: New estimates of R&D costs.” *Journal of Health Economics* 2016; 47: 20–33 (Tufts CSDD, announced 2014). The ~$2.6B figure is the cost to develop and win approval for a new drug. doi:10.1016/j.jhealeco.2016.01.012.

   Applied as a proxy for translational cost; the original study measures pharmaceutical R&D specifically, and the estimate is debated.

4. [4]Ioannidis JPA. “Why Most Published Research Findings Are False.” *PLoS Medicine* 2005; 2(8): e124. doi:10.1371/journal.pmed.0020124.

5. [5]Open Science Collaboration. “Estimating the reproducibility of psychological science.” *Science* 2015; 349(6251): aac4716. doi:10.1126/science.aac4716.

6. [6]Butler D. “Translational research: crossing the valley of death.” *Nature* 2008; 453(7197): 840–842. doi:10.1038/453840a.

7. [7]Contopoulos-Ioannidis DG, Ntzani EE, Ioannidis JPA. “Translation of highly promising basic science research into clinical applications.” *The American Journal of Medicine* 2003; 114(6): 477–484. Of 101 basic-science findings claiming clinical promise, 5 were licensed for clinical use ~20 years later. PMID 12731504.

8. [8]Galkina Cleary E, Jackson MJ, Ledley FD. “Government as the First Investor in Biopharmaceutical Innovation: Evidence from New Drug Approvals 2010–2019.” Institute for New Economic Thinking, Working Paper No. 133 (2021). ~83% of NIH funding went to basic research on targets vs. ~17% to applied research on the drugs. ineteconomics.org.

9. [9]National Center for Advancing Translational Sciences (NCATS), FY2021 Congressional Justification; NIH Office of Budget. NCATS FY2020 enacted funding was $832.9M of NIH's ~$40B total (≈2%). NIH Office of Budget.

The ~2% share is computed from NCATS's enacted appropriation against NIH's total budget authority.

10. [10]Ghuman H, Perry N, Grice L, Gerwig M, Moorhead J Jr, Nitzsche F, Poplawsky AJ, Ambrosio F, Modo M. "Physical therapy exerts sub-additive and suppressive effects on intracerebral neural stem cell implantation in a rat model of stroke." *Journal of Cerebral Blood Flow & Metabolism* 2022; 42(5): 826–843. A collectively authored preclinical study (Dr. Jeffrey Moorhead, MD, DPT, is a co-author), cited as a translational-research example. PMID 34826373 · doi:10.1177/0271678X211062955.

11. [11]Dr. Jeffrey Moorhead (MD, DPT), Moore J, Brooks R, Ghuman H, Tramontini F, Mayo E, Coleman JE. "Transforming Translational Research: Leveraging Cloud Infrastructure for Enhanced Collaboration and Efficiency to Sustain Innovations through the Valley of Death." Working manuscript, University of Arizona / HumanData. The clinical origin of the series thruline; co-authored by the series author.

Unpublished working manuscript; quotations are from its introduction.

§ 04 WP 3.0 — Speed vs. Trust

Transaction architecture — *sources.*

Five sources: two on the perceptual latency threshold that defines "instant," and three foundational works in the organizational-trust literature that ground the paper's relational and institutional layers. The 3–5 yr trust timeline, 71% transaction-loss, and 89% fraud figures are not cited — they remain labeled modeled estimates, as the trust literature supports multi-year formation but does not fix those specific values.

1. [1]Nielsen J. "Response Times: The 3 Important Limits." Nielsen Norman Group, 1993 (from *Usability Engineering*). The ~0.1 second (100 ms) limit is the threshold at which a system feels instantaneous. nngroup.com.

2. [2]Miller RB. "Response time in man–computer conversational transactions." *Proceedings of the AFIPS Fall Joint Computer Conference* 1968; 267–277. Original basis for the 0.1 second perceptual threshold. doi:10.1145/1476589.1476628.

3. [3]Mayer RC, Davis JH, Schoorman FD. "An Integrative Model of Organizational Trust." *Academy of Management Review* 1995; 20(3): 709–734. Foundational model of trust as willingness to be vulnerable, built from ability, benevolence, and integrity. doi:10.2307/258792.

4. [4]McAllister DJ. "Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations." *Academy of Management Journal* 1995; 38(1): 24–59. Distinguishes fast, cognition-based trust from slower, affect-based trust that accrues through interaction. doi:10.2307/256727.

5. [5]Zucker LG. "Production of Trust: Institutional Sources of Economic Structure, 1840–1920." *Research in Organizational Behavior* 1986; 8: 53–111. On institution-based trust — the systemic infrastructure (guarantees, intermediaries, reputation) that substitutes for relational history.

Judgment under uncertainty — *sources.*

Eight sources on judgment error and its correction: the heuristics-and-biases tradition, the noise/consistency evidence, clinical-versus-mechanical prediction, reference-class forecasting, and the forecasting-tournament evidence behind calibration discipline.

1. [1]Kahneman D, Rosenfield AM, Gandhi L, Blaser T. “Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making.” *Harvard Business Review* 2016; Oct. Reports the underwriting noise audit — a median 55% difference between underwriters pricing the same risk, against an expected ~10% — and a 43% difference among claims adjusters. hbr.org/2016/10/noise.

2. [2]Kahneman D, Sibony O, Sunstein CR. *Noise: A Flaw in Human Judgment*. Little, Brown Spark, 2021. Book-length treatment distinguishing bias from noise and setting out the noise audit and decision-hygiene practices. hachettebookgroup.com.

3. [3]Flyvbjerg B, Holm MS, Buhl S. “Underestimating Costs in Public Works Projects: Error or Lie?” *Journal of the American Planning Association* 2002; 68(3): 279–295. Cost underestimation found in ~9 of 10 large transport projects; actual rail costs average ~45% above estimate. Empirical basis for reference-class forecasting. doi:10.1080/01944360208976273.

4. [4]Grove WM, Meehl PE. “Clinical Versus Mechanical Prediction: A Meta-Analysis.” *Psychological Assessment* 2000; 12(1): 19–30. Across 136 studies, mechanical prediction matched or exceeded clinical judgment in 128; clinical judgment was clearly superior in only 8. doi:10.1037/1040-3590.12.1.19.

5. [5]Tversky A, Kahneman D. “Judgment under Uncertainty: Heuristics and Biases.” *Science* 1974; 185(4157): 1124–1131. Foundational account of the availability, representativeness, and anchoring heuristics and their systematic biases. doi:10.1126/science.185.4157.1124.

6. [6]Kahneman D, Tversky A. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 1979; 47(2): 263–291. Establishes loss aversion and framing effects — biases in risk preference itself. doi:10.2307/1914185.

7. [7]Dawes RM, Faust D, Meehl PE. “Clinical Versus Actuarial Judgment.” *Science* 1989; 243(4899): 1668–1674. Reviews the consistent advantage of simple actuarial models over unaided expert judgment, attributing much of the gap to the noisiness of human discretion. doi:10.1126/science.2648573.

8. [8]Mellers B, Ungar L, Baron J, et al. “Psychological Strategies for Winning a Geopolitical Forecasting Tournament.” *Psychological Science* 2014; 25(5): 1106–1115. Good Judgment Project evidence that probabilistic training, teaming, and tracking measurably improve forecasting accuracy — the empirical core of calibration discipline. doi:10.1177/0956797614524255.

Human + machine intelligence — *sources.*

Eight sources: the Prediction Machines framing and the Turing Trap, the automation-trust literature, and the recent field and experimental evidence on generative-AI productivity — including its uneven "jagged frontier."

1. [1]Noy S, Zhang W. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science* 2023; 381(6654): 187–192. Professional writing tasks completed ~40% faster with ~18% higher rated quality; gains largest for lower-ability writers. doi:10.1126/science.adh2586.

2. [2]Brynjolfsson E, Li D, Raymond LR. "Generative AI at Work." *NBER Working Paper* 31161, 2023; forthcoming *Quarterly Journal of Economics*. Access to an AI assistant raised issues resolved per hour ~14% on average and ~34% for novice agents. nber.org/papers/w31161.

3. [3]Peng S, Kalliamvakou E, Cihon P, Demirer M. "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot." *arXiv preprint* arXiv:2302.06590, 2023. Developers with Copilot completed the programming task 55.8% faster than the control group. arxiv.org/abs/2302.06590.

4. [4]Dell'Acqua F, McFowland III E, Mollick ER, et al. "Navigating the Jagged Technological Frontier." *Harvard Business School Working Paper* 24-013, 2023; published in *Organization Science* 2025. Large within-frontier gains; on a task beyond the frontier, AI-assisted consultants were ~19% less likely to produce correct solutions. doi:10.1287/orsc.2025.21838.

5. [5]Brynjolfsson E. "The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence." *Daedalus* 2022; 151(2): 272–287. Argues that benchmarking AI against human-like performance biases design toward substitution, concentrating wealth and power, rather than augmentation. doi:10.1162/daed_a_01915.

6. [6]Agrawal A, Gans J, Goldfarb A. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press, 2018. Frames machine intelligence as a fall in the cost of prediction that raises the value of its complement, human judgment.

7. [7]Agrawal A, Gans JS, Goldfarb A. "Prediction, Judgment and Complexity: A Theory of Decision Making and Artificial Intelligence." *NBER Working Paper* 24243, 2018. Formalizes the decomposition of decisions into prediction (machine) and judgment (human). nber.org/papers/w24243.

8. [8]Lee JD, See KA. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors* 2004; 46(1): 50–80. Defines calibrated trust and appropriate reliance; links miscalibration to automation misuse (over-reliance) and disuse (under-reliance). doi:10.1518/hfes.46.1.50_30392.

§ 07 WP 6.0 — The Measurement Trap

Designing what counts — *sources.*

Seven sources — the classic performance-measurement pathologies (Ridgway, Kerr, Goodhart, Campbell, Strathern), the experimental surrogation evidence, and Muller's survey of metric fixation.

1. [1]Ridgway VF. “Dysfunctional Consequences of Performance Measurements.” *Administrative Science Quarterly* 1956; 1(2): 240–247. The first controlled account of how single, composite, and multiple performance measures each generate dysfunctional side effects. doi:10.2307/2390989.

2. [2]Goodhart CAE. “Problems of Monetary Management: The UK Experience.” In *Monetary Theory and Practice: The UK Experience*, pp. 91–121. Macmillan, 1984 (presented 1975). The origin of “Goodhart's Law”: a statistical regularity breaks down once relied upon for control. doi:10.1007/978-1-349-17295-5_4.

3. [3]Kerr S. “On the Folly of Rewarding A, While Hoping for B.” *Academy of Management Journal* 1975; 18(4): 769–783. Classic analysis of reward systems that pay for one behavior while hoping for another. doi:10.2307/255378.

4. [4]Strathern M. “‘Improving ratings’: audit in the British University system.” *European Review* 1997; 5(3): 305–321. Source of the widely quoted formulation: “when a measure becomes a target, it ceases to be a good measure.” doi:10.1017/S1062798700002660.

5. [5]Campbell DT. “Assessing the Impact of Planned Social Change.” *Evaluation and Program Planning* 1979; 2(1): 67–90. “Campbell's Law”: the more a quantitative indicator is used for decision-making, the more it corrupts and distorts the process it monitors. doi:10.1016/0149-7189(79)90048-X.

6. [6]Choi J, Hecht GW, Tayler WB. “Lost in Translation: The Effects of Incentive Compensation on Strategy Surrogation.” *The Accounting Review* 2012; 87(4): 1135–1163. Experimental evidence that tying pay to a measure causes managers to substitute the measure for the strategy it represents. doi:10.2308/accr-10273.

7. [7]Muller JZ. *The Tyranny of Metrics*. Princeton University Press, 2018. Survey of “metric fixation” across medicine, education, policing, and business, and of the recurring costs of mistaking the measurable for the valuable. doi:10.1515/9780691191263.

§ 08 WP 7.0 — The Compounding Organization

Learning systems & capability — *sources.*

Seven sources: the learning/experience curve, single- versus double-loop learning, absorptive capacity, and psychological safety as the condition for organizational learning.

1. [1]Boston Consulting Group. “BCG Classics Revisited: The Experience Curve.” 2013 (restating “The Experience Curve,” BCG *Perspectives*, 1968). Real unit costs decline ~20–30% for each doubling of accumulated experience. bcg.com.

2. [2]Wright TP. “Factors Affecting the Cost of Airplanes.” *Journal of the Aeronautical Sciences* 1936; 3(4): 122–128. The original learning curve: unit labor falls a stable percentage with each doubling of cumulative output. doi:10.2514/8.155.

3. [3]Argyris C. “Double Loop Learning in Organizations.” *Harvard Business Review* 1977; Sept–Oct. Distinguishes single-loop (correcting actions) from double-loop (correcting the governing variables behind them). hbr.org.

4. [4]Argyris C, Schön DA. *Organizational Learning: A Theory of Action Perspective*. Addison-Wesley, 1978. Defines single-loop, double-loop, and deutero-learning (learning to learn), and the defensive routines that keep organizations single-loop.

5. [5]Cohen WM, Levinthal DA. “Absorptive Capacity: A New Perspective on Learning and Innovation.” *Administrative Science Quarterly* 1990; 35(1): 128–152. The capacity to absorb new knowledge is a cumulative function of prior related knowledge. doi:10.2307/2393553.

6. [6]Edmondson A. “Psychological Safety and Learning Behavior in Work Teams.” *Administrative Science Quarterly* 1999; 44(2): 350–383. Shows team psychological safety as the condition enabling the learning behaviors — error reporting, inquiry, challenge — that improvement requires. doi:10.2307/2666999.

7. [7]Senge PM. *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday/Currency, 1990. The disciplines — systems thinking, shared mental models, team learning — that let an organization continually expand its capacity to create.

§ 09 WP 8.0 — The Trust Infrastructure

Institutions, platforms & scale — *sources.*

Eight sources spanning the social limits of relational trust, institution-based trust production, commons governance, transaction-cost economics, embeddedness, reputation-system design, and the economic payoff of generalized trust.

1. [1]Dunbar RIM. “Coevolution of neocortical size, group size and language in humans.” *Behavioral and Brain Sciences* 1993; 16(4): 681–694. Source of the ~150 limit on stable social relationships, derived from the neocortex-ratio / group-size relationship. doi:10.1017/S0140525X00032325.

2. [2]Ostrom E. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990 — the eight design principles characterizing long-enduring common-pool-resource institutions. See also Ostrom E. “Beyond Markets and States: Polycentric Governance of Complex Economic Systems.” *American Economic Review* 2010; 100(3): 641–672. doi:10.1257/aer.100.3.641.

3. [3]Knack S, Keefer P. “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation.” *The Quarterly Journal of Economics* 1997; 112(4): 1251–1288. Associates generalized trust and civic norms with economic performance across 29 market economies. doi:10.1162/003355300555475.

4. [4]Zucker LG. “Production of Trust: Institutional Sources of Economic Structure, 1840–1920.” *Research in Organizational Behavior* 1986; 8: 53–111. Foundational account of institution-based trust — certification, intermediaries, and guarantees that substitute for personal and community history.

Book-series chapter; no DOI issued. Widely indexed.

5. [5]Williamson OE. “The Economics of Organization: The Transaction Cost Approach.” *American Journal of Sociology* 1981; 87(3): 548–577. Transaction-cost logic for the make-versus-market boundary and for governing against opportunism. doi:10.1086/227496.

6. [6]Granovetter M. “Economic Action and Social Structure: The Problem of Embeddedness.” *American Journal of Sociology* 1985; 91(3): 481–510. Argues economic action is embedded in concrete, ongoing structures of social relations. doi:10.1086/228311.

7. [7]Resnick P, Kuwabara K, Zeckhauser R, Friedman E. “Reputation Systems.” *Communications of the ACM* 2000; 43(12): 45–48. Seminal framing of online reputation systems as engineered trust infrastructure for exchange among strangers. doi:10.1145/355112.355122.

8. [8]Fukuyama F. *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, 1995. On the capacity to form trusting associations beyond the family as a determinant of economic organization and prosperity.

Monograph; no DOI issued. ISBN 978-0-02-910976-7.

SERIES CLOSING STATEMENT

"Intelligence and trust are both accumulated assets. They compound when invested consistently and depreciate rapidly when neglected. The organizations that will define the next era of economic activity are those that treat both as core infrastructure — not as emergent properties of scale, but as designed systems, maintained with intention, evaluated against outcomes."