

# OIDD 989. Explaining Explanation

## Instructor: Duncan J. Watts

Time: Wed, 3-6PM<sup>1</sup>

Location: TBD

## Course Overview

### Description

In the social sciences we often use the word “explanation” as if (a) we know what we mean by it, and (b) we mean the same thing that other people do. In this course we will critically examine these assumptions and their consequences for scientific progress. In part 1 of the course we will examine how, in practice, researchers invoke at least three logically and conceptually distinct meanings of “explanation:” identification of causal mechanisms; ability to predict (account for variance in) some outcome; and ability to make subjective sense of something. In part 2 we will examine how and when these different meanings are invoked across a variety of domains, focusing on social science, history, business, and machine learning, and will explore how conflation of these distinct concepts may have created confusion about the goals of science and how we evaluate its progress. Finally, in part 3 we will discuss some related topics such as null hypothesis testing and the replication crisis. We will also discuss specific practices that could help researchers clarify exactly what they mean when they claim to have “explained” something, and how adoption of such practices may help social science be more useful and relevant to society.

### Structure of the course

Class will be discussion based and will meet once per week for 3 hours. Students will be expected to have read all the mandatory readings for each week prior to attending class and will be required to submit weekly “reading reports” prior to each class.

### Evaluation

30% Class attendance and participation.

30% Weekly reading reports (to be submitted prior to class)

40% Final project (due final class).

---

<sup>1</sup> First class will be Wed Jan 20. There will be no class the week of March 10 (Spring Break). There will be no class either the week of Feb 17 or Feb 24 (TBD). Last class will be Wed Apr 28

### *Class attendance and participation.*

This course, by its nature, is dealing with an imprecisely defined topic with blurry boundaries and ambiguous connections among numerous other topics. For this reason, it is essential for students to engage actively with the readings and, via in-class discussions, with each other. Students are therefore expected to attend all classes where exceptions will be made for medical illness (all other absences should be approved in advance by the instructor). In order to facilitate broad participation the instructor will appoint weekly discussants (schedule TBA).

### *Reading reports*

To ensure that students come to class prepared, a weekly reading report that briefly summarizes the main arguments of the required readings.

### *Final project (15-20 pages double spaced, excluding references)*

Choose a domain (e.g. your research area, a literature review of a field, something else that catches your interest such as history or contemporary events) and analyze how explanations in that domain are deployed in both clarifying and misleading ways. Your approach may be quantitative or qualitative, broad or narrow, and may focus on any of the subtopics of the class. The objective is to demonstrate understanding of the material and an ability to apply it “in the wild.”

## **PART 1**

### **Week 1: Introduction**

Dienes, Zoltan. 2008. *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Macmillan International Higher Education.

#### **Optional**

1. Watts, Duncan J. 2011. *Everything Is Obvious: \* Once You Know the Answer*. Crown Business.
2. Watts, Duncan J., Emorie D. Beck, Elisa J. Bienenstock, Jake Bowers, Aaron Frank, Anthony Grubestic, Jake M. Hofman, Julia M. Rohrer, and Matthew Salganik. 2018. “Explanation, Prediction, and Causality: Three Sides of the Same Coin?” <https://doi.org/10.31219/osf.io/u6vz5>.

### **Week 2: Explanation as Causality**

1. Pearl, Judea. 2009. *Causality*. Cambridge University Press. Epilogue only.
2. Gelman, Andrew. 2011. “Causality and Statistical Learning.” *The American Journal of Sociology* 117 (3): 955–66
3. Hedström, Peter, and Petri Ylikoski. 2010. “Causal Mechanisms in the Social Sciences.” *Annual Review of Sociology* 36: 49–67.
4. Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. “Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies.” *The American Political Science Review* 105 (4): 765–89.

## Optional

1. Gelman, Andrew, and Guido Imbens. 2013. "Why Ask Why? Forward Causal Inference and Reverse Causal Questions." National Bureau of Economic Research
2. Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books. (esp. Chapter 1 and Chapter 8)
3. Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference*. Cambridge University Press.
4. Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, USA.
5. Small, Mario Luis. 2013. "Causal Thinking and Ethnographic Research." *The American Journal of Sociology* 119 (3): 597–601.

## Week 3: Explanation as Prediction

1. Hofman, Jake M., Amit Sharma, and Duncan J. Watts. 2017. "Prediction and Explanation in Social Systems." *Science* 355 (6324): 486–88.
2. Yarkoni, Tal, and Jacob Westfall. 2017. "Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 12 (6): 1100–1122.
3. Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems." *The American Economic Review* 105 (5): 491–95.
4. Salganik, Matthew J., Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, et al. 2020. "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration." *Proceedings of the National Academy of Sciences of the United States of America* 117 (15): 8398–8403.

## Optional

1. Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 16 (3): 199–231.
2. Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.
3. Athey, Susan. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science* 355 (6324): 483–85.
4. Shmueli, Galit, and Others. 2010. "To Explain or to Predict?" *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25 (3): 289–310.
5. Sanders, Nathan. 2019. "A Balanced Perspective on Prediction and Inference for Data Science in Industry." *Harvard Data Science Review* 1 (1).

## Week 4: Explanation as Sensemaking

1. Gopnik, Alison. 1998. "Explanation as Orgasm." *Minds and Machines* 8 (1): 101–18.
2. Shanton, Karen, and Alvin Goldman. 2010. "Simulation Theory." *Wiley Interdisciplinary Reviews*.

*Cognitive Science* 1 (4): 527–38.

3. Lombrozo, Tania. 2016. "Explanatory Preferences Shape Learning and Inference." *Trends in Cognitive Sciences* 20 (10): 748–59.
4. Tilly, Charles. 2004. "Reasons Why." *Sociological Theory* 22 (3): 445–54.

## Optional

1. Gelman, Andrew, and Thomas Basbøll. 2014. "When Do Stories Work? Evidence and Illustration in the Social Sciences." *Sociological Methods & Research* 43 (4): 547–70.
2. Madsbjerg, Christian. 2017. *Sensemaking: What Makes Human Intelligence Essential in the Age of the Algorithm*. Little, Brown Book Group.
3. Becker, Howard S. 1998. *Tricks of the Trade: How to Think about Your Research While You're Doing It*. Chicago: University of Chicago Press. (Chapter 3)
4. Freeman, Mark. 2010. "Hindsight." Oxford, England: Oxford University Press.
5. Lombrozo, Tanya. 2007. "Simplicity and Probability in Causal Explanation." *Cognitive Psychology* 55 (3): 232–57.
6. Lombrozo, T. 2006. "The Structure and Function of Explanations." *Trends in Cognitive Sciences* 10 (10): 464–70.
7. Freling, Traci H., Zhiyong Yang, Ritesh Saini, Omar S. Itani, and Ryan Rashad Abualsamh. 2020. "When Poignant Stories Outweigh Cold Hard Facts: A Meta-Analysis of the Anecdotal Bias." *Organizational Behavior and Human Decision Processes* 160 (September): 51–67.

# PART 2: Examples

## Week 5: Explanations in Social Science

1. Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47 (4): 363–75.
2. Watts, Duncan J. 2014. "Common Sense and Sociological Explanations." *The American Journal of Sociology* 120 (2): 313–51.
3. Turco, Catherine J., and Ezra W. Zuckerman. 2017. "Verstehen for Sociology: Comment on Watts." *The American Journal of Sociology* 122 (4): 1272–91.
4. Watts, Duncan. 2017. "Response to Turco and Zuckerman's 'Verstehen for Sociology.'" *The American Journal of Sociology* 122 (4): 1292–99.

## Optional

1. Debrouwere, Stijn. 2020. "The Conceptual, Cunning and Conclusive Experiment in Psychology." <https://users.ugent.be/~stdebrouw/2020-02-19-stijn-debrouwere-conceptual-cunning-and-conclusive-experiment.pdf>.
2. DeJesus, Jasmine M., Maureen A. Callanan, Graciela Solis, and Susan A. Gelman. 2019. "Generic Language in Scientific Communication." *Proceedings of the National Academy of Sciences of the United States of America* 116 (37): 18370–77.
3. Elster, Jon. 2015. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge University Press
4. Lieberman, Stanley, and Freda B. Lynn. 2002. "Barking up the Wrong Branch: Scientific

- Alternatives to the Current Model of Sociological Science.” *Annual Review of Sociology*, 1–19.
5. Stafford, Tom. 2014. “The Perspectival Shift: How Experiments on Unconscious Processing Don’t Justify the Claims Made for Them.” *Frontiers in Psychology* 5 (September): 1067.
  6. Vancouver, Jeffrey B. 2012. “Rhetorical Reckoning: A Response to Bandura.” *Journal of Management* 38 (2): 465–74.

## Week 6: Explanations in History

1. Kreiswirth, M. 2000. “Merely Telling Stories? Narrative and Knowledge in the Human Sciences.” *Poetics Today*. <https://read.dukeupress.edu/poetics-today/article-abstract/21/2/293/74627>.
2. Risi, Joseph, Amit Sharma, Rohan Shah, Matthew Connelly, and Duncan J. Watts. 2019. “Predicting History.” *Nature Human Behaviour* 3 (9): 906–12.
3. Stueber, Karsten R. 2008. “2. REASONS, GENERALIZATIONS, EMPATHY, AND NARRATIVES: THE EPISTEMIC STRUCTURE OF ACTION EXPLANATION.” *History and Theory* 47 (1): 31–43.
4. Sunstein, Cass R. 2016. “Historical Explanations Always Involve Counterfactual History.” *Journal of the Philosophy of History* 10 (3): 433–40.

### Optional

1. Berlin, Isaiah. 2013. *The Hedgehog and the Fox: An Essay on Tolstoy’s View of History - Second Edition*. Princeton University Press.
2. Danto, Arthur C. 1965. *Analytical Philosophy of History*. Cambridge, UK: Cambridge University Press.
3. Ferguson, Niall. 2008. *Virtual History: Alternatives and Counterfactuals*. Hachette UK. (pp. 1-90)
4. Gaddis, John Lewis. 2002. *The Landscape of History: How Historians Map the Past*. Oxford, UK: Oxford University Press.
5. MacMullen, Ramsay. 2012. *Feelings in History: Ancient and Modern*. CreateSpace Independent Publishing Platform.
6. Rosenberg, Alexander. 2018. *How History Gets Things Wrong: The Neuroscience of Our Addiction to Stories*. MIT Press.

## Week 7: Explanations in Business

1. Rosenzweig, Phil. 2007. *The Halo Effect*. New York: Free Press.
2. Mitchell, Gregory. 2004. “Case Studies, Counterfactuals, and Causal Explanations.” *University of Pennsylvania Law Review* 152 (5): 1517–1608.

### Optional

- Raynor, Michael. 2007. *The Strategy Paradox: Why Committing to Success Leads to Failure*. New York: Doubleday.
- Niendorf, Bruce, and Kristine Beck. 2008. “Good to Great, or Just Good?” *Academy of Management Perspectives* 22 (4): 13–20.

## Week 8: Explanations in Machine Learning

1. Domingos, Pedro. 1999. "The Role of Occam's Razor in Knowledge Discovery." *Data Mining and Knowledge Discovery* 3 (4): 409–25.
2. Lipton, Zachary C. 2018. "The Mythos of Model Interpretability." *Queueing Systems. Theory and Applications* 16 (3): 31–57.
3. Selbst, Andrew, and Solon Barocas. 2017. "Regulating Inscrutable Systems." In *WeRobot 2017*.
4. Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *The Journal of Economic Perspectives: A Journal of the American Economic Association* 31 (2): 87–106.

### Optional

1. Coveney, Peter V., Edward R. Dougherty, and Roger R. Highfield. 2016. "Big Data Need Big Theory Too." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 374 (2080). <https://doi.org/10.1098/rsta.2016.0153>.
2. Agrawal, Mayank, Joshua C. Peterson, and Thomas L. Griffiths. 2020. "Scaling up Psychology via Scientific Regret Minimization." *Proceedings of the National Academy of Sciences of the United States of America* 117 (16): 8825–35.
3. Mothilal, R. K., A. Sharma, and C. Tan. 2020. "Explaining Machine Learning Classifiers through Diverse Counterfactual Explanations." *Proceedings of the 2020 Conference on*. <https://dl.acm.org/doi/abs/10.1145/3351095.3372850>.
4. Fudenberg, Drew, Jon Kleinberg, Annie Liang, and Sendhil Mullainathan. 2019. "Measuring the Completeness of Theories." <https://doi.org/10.2139/ssrn.3018785>.

## PART 3: Improving Scientific Explanations

### Week 9. Null hypothesis testing

1. Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52 (3): 647–74.
2. Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11): 1359–66.
3. Gelman, Andrew, and Eric Loken. 2014. "The Statistical Crisis in Science Data-Dependent Analysis—a 'garden of Forking Paths'—explains Why Many Statistically Significant Comparisons Don't Hold up." *American Scientist* 102 (6): 460.
4. Schneider, J. 2018. "Data-Dependent Analytical Choices Relying on NHST Should Not Be Trusted!" In *23rd International Conference on Science and Technology Indicators (STI 2018), September 12-14, 2018, Leiden, The Netherlands*. Centre for Science and Technology Studies (CWTS). <https://openaccess.leidenuniv.nl/handle/1887/65352>.

## Optional

1. Johnson, D. H. 1999. "The Insignificance of Statistical Significance Testing." *The Journal of Wildlife Management*.
2. Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): e124.
3. Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. 2016. "Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology* 31 (4): 337–50.
4. Amrhein, Valentin, Fränzi Korner-Nievergelt, and Tobias Roth. 2017. "The Earth Is Flat ( $p > 0.05$ ): Significance Thresholds and the Crisis of Unreplicable Research." *PeerJ* 5: e3544.
5. Gelman, Andrew, and John Carlin. 2017. "Some Natural Solutions to the P-Value Communication Problem—and Why They Won't Work." *Journal of the American Statistical Association* 112 (519): 899–901.

## Week 10. Reproducibility and Replication

1. Freese, Jeremy, and David Peterson. n.d. "Replication in Social Science." <https://doi.org/10.31235/osf.io/5bck9>.
2. Billheimer, Dean. 2019. "Predictive Inference and Scientific Reproducibility." *The American Statistician* 73 (sup1): 291–95.
3. Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences of the United States of America* 115 (11): 2600–2606.
4. Dwork, Cynthia, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. 2015. "The Reusable Holdout: Preserving Validity in Adaptive Data Analysis." *Science* 349 (6248): 636–38.

## Optional

1. Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. "A Manifesto for Reproducible Science." *Nature Human Behaviour* 1: 0021.
2. King, Gary. 1995. "Replication, Replication." *PS, Political Science & Politics* 28 (3): 444–52.
3. National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Committee on Science, Engineering, Medicine, and Public Policy, Board on Research Data and Information, Division on Engineering and Physical Sciences, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences and Analytics, et al. 2019. *Reproducibility and Replicability in Science*. National Academies Press.
4. Miller, Jeff. 2009. "What Is the Probability of Replicating a Statistically Significant Effect?" *Psychonomic Bulletin & Review* 16 (4): 617–40.
5. Coffman, Lucas C., and Muriel Niederle. 2015. "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible." *Journal of Economic Perspectives*. <https://doi.org/10.1257/jep.29.3.81>.

## Week 11: Generalization

1. Yarkoni, Tal. 2019. "The Generalizability Crisis." <https://doi.org/10.31234/osf.io/jqw35>
2. Takens, Daniel. 2020. "Review of 'The Generalizability Crisis' by Tal Yarkoni" <http://daniellakens.blogspot.com/2020/01/review-of-generalizability-crisis-by.html>
3. Yarkoni, Tal. 2020. "Induction is not optional if you're using inferential statistics." <https://www.talyarkoni.org/blog/2020/05/06/induction-is-not-optional-if-youre-using-inferential-statistics-reply-to-lakens/>
4. Domingos, Pedro. 2012. "A Few Useful Things to Know about Machine Learning." *Communications of the ACM* 55 (10): 78–87.

### Optional

1. Gelman, Andrew. 2020. Comment on Yarkoni. <https://statmodeling.stat.columbia.edu/2020/04/07/the-generalizability-crisis-in-the-human-sciences/>
2. Hand, David J. 2006. "Classifier Technology and the Illusion of Progress." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 21 (1): 1–14.

## Week 12. Experiments

Manzi, Jim. 2012. "Uncontrolled: The Surprising Payoff of Trial-and-Error for Business." *Politics, and Society. Basic Books*, 1–320.

### Optional

1. Luca, Michael, and Max H. Bazerman. 2020. *The Power of Experiments: Decision Making in a Data-Driven World*. MIT Press.
2. Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press.
3. Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.
4. Gordon, Brett R., Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. 2019. "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook." *Marketing Science* 38 (2): 193–225.

## Week 13. Other Ideas

1. DellaVigna, Stefano, Devin Pope, and Eva Vivaldi. 2019. "Predict Science to Improve Science." *Science* 366 (6464): 428–29.
2. Watts, Duncan J. 2017. "Should Social Science Be More Solution-Oriented?" *Nature Human Behaviour* 1: 0015.
3. Griffiths, Thomas L. 2015. "Manifesto for a New (computational) Cognitive Revolution." *Cognition* 135 (February): 21–23.



4. Baribault, Beth, Chris Donkin, Daniel R. Little, Jennifer Trueblood, Zita Oravec, Don van Ravenzwaaij, Corey White, Paul De Boeck, and Joachim Vandekerckhove. 2017. "Meta-Studies for Robust Tests of Theory." <https://doi.org/10.31219/osf.io/g84py>.

## Optional

1. Coyne, James C. 2016. "Replication Initiatives Will Not Salvage the Trustworthiness of Psychology." *BMC Psychology* 4 (1): 28.
2. Muthukrishna, M., and J. Henrich. n.d. 2019. "A Problem in Theory." *Nature Human Behaviour*.
3. Oberauer, Klaus, and Stephan Lewandowsky. 2019. "Addressing the Theory Crisis in Psychology." *Psychonomic Bulletin & Review* 26 (5): 1596–1618.
4. LeBel, Etienne P., Randy J. McCarthy, Brian D. Earp, Malte Elson, and Wolf Vanpaemel. 2018. "A Unified Framework to Quantify the Credibility of Scientific Findings." *Advances in Methods and Practices in Psychological Science* 1 (3): 389–402.
5. Baumeister, Roy F. 2016. "Charting the Future of Social Psychology on Stormy Seas: Winners, Losers, and Recommendations." *Journal of Experimental Social Psychology* 66 (September): 153–58.
6. Forscher, B. K. 1963. "Chaos in the Brickyard." *Science* 142 (3590): 339.
7. Morling, Beth, and Robert Calin-Jageman. 2019. "What Psychology Teachers Should Know about Open Science and the New Statistics (Morling & Calin-Jageman, 2020)." <https://doi.org/10.31234/osf.io/qxwb7>.