Syllabus: Cultivating Your Judgment Skills: The Art and Science of Confidence Calibration in Business, Politics and Life

Prof. Philip E. Tetlock (Steinberg-Dietrich 3203: tetlock@wharton.upenn.edu)

Management 776, Q3 2013, Wednesdays, 4-7 pm (1/9/13 to 2/27/13)

Overview.
A world-class poker player defined the difference between serious players and amateurs this way: serious players know the difference between 40-60 and 60-40 propositions. In other words, serious players are well calibrated (skilled at distinguishing what they know from what they don't). It is a good idea to get your calibration workouts in low-risk settings, such as Wharton.

This course gives you opportunities to test your wits against competitors in a global-forecasting tournament sponsored by the Intelligence Advanced Research Projects Agency (IARPA)—as well as opportunities to design forecasting exercises around your careers and business plans. Key challenges include: (a) learning to translate "interesting questions" into "testable propositions;" (b) getting into the habit of translating vague hunches into probability metrics and using accuracy feedback and other tools to become better calibrated; (c) making solo forecasts—and exploring methods of making team forecasts that are more than the sum of their individual-member parts.

The goal is to bridge the gap between "private" and "public" knowledge—a gap captured by the distinction between knowing how to ride a bicycle and knowing Newton's laws of motion. In other classes, you have learned about formal models for guiding business decisions. In this class, you can use whatever models you want. The focus will be on your skills in sizing up real-world problems. You will learn how much trust you place in your advisers vs. your instincts—and how much trust you should place in your judgments of trustworthiness.

Bio Blurb:
Phil Tetlock is Annenberg University Professor at the University of Pennsylvania (bridging Wharton and the School of Arts and Sciences). He has written numerous books and articles in peer-reviewed journals—and received many scientific awards. In 2005, he wrote "Expert political judgment: How good is it? How can we know?" which tracked the accuracy of leading political experts between 1985 and 2005. He is currently Principal Investigator for Team Good Judgment, which is competing in a forecasting tournament sponsored by IARPA. The goal is to assign accurate probability estimates to a wide range of geopolitical and economic outcomes.

Grading (pass-fail permitted):
Grades will be based on (a) class participation, including your end-of-term presentation due February 27 (25%); (b) a term paper, *due 1 week after last class*, March 6, 5 pm, that shows you understand key concepts and have applied them (50%); (c) a *choice* between being evaluated on the accuracy of your probabilistic predictions (25% on pure "outcome accountability") or the quality of your rationales for those predictions (25% on pure "process accountability") or a hybrid system (12.5% based on accuracy metrics and 12.5% on quality-of-rationale metrics).

Week 1 (Jan. 9). Overview: Luck vs. Skill.

This session explores the challenges of disentangling luck vs. skill. To paraphrase Mauboussin (2012)'s definitive book:

> A basketball player's shot before the final buzzer bounces out of the basket and his team loses the championship. A pharmaceutical company develops a drug for hypertension that ends up a blockbuster seller for erectile dysfunction. An investor earns a windfall when he buys the stock of a company shortly before it gets acquired at a premium. Shifting mixes of skill and luck shape our lives. But we aren't adept at distinguishing the two.

This session has four additional goals:

(a) exploring whether your implicit beliefs about the roles of luck and skill are holding you back from cultivating career-enhancing skills;
(b) learning to translate interesting questions into testable propositions—and treating your beliefs like probability judgments that can be scored for accuracy;
(c) introducing the IARPA forecasting tournament—and adapting it to our purposes. Half of the questions will come from IARPA (questions of special relevance to business); the other half, nominated by the class;
(d) avoiding the "narrativist fallacy" (just because pundits can offer colorful explanations for an outcome does not mean they could have predicted it). What credibility cues do you use in choosing advisers? What do you *really know* about their track records?

Reading Assignments

Mauboussin, M. (2012). *The success equation: Untangling success and luck in business, sports, and investing*. Harvard Business Review Press. Chapters 1-3.

Rosenzweig, P. (2006). *The halo effect and eight other business delusions that deceive managers. New York; Free Press*. Chapters 1 and 2.

Week 2 (Jan. 16): Individual-level interventions for improving probability judgments.

This session has two goals: first, to introduce you to the warehouse of research findings, stocked up over 4 decades, that documents systematic (not just random) errors in human judgment ; second, introduce you to techniques and tools for improving the "calibration" and "discrimination" of your individual probability judgments.

Tentatively schedule guest: Michael Mauboussin, Chief Investment Officer, Legg-Mason Capital Management.

Reading Assignments

Lovallo, D. & Kahneman, D. (2003). *Delusions of success: How optimism undermines executives' decisions*. Harvard Business Review.

Makridakis, S. et al. (2010). *Why forecasts fail. What to do instead*. MIT Sloan Management Review.

Mauboussin, M. (2012). *The success equation: Untangling success and luck in business, sports, and investing*. Harvard Business Review Press. Chapters 4-7.

Week 3 (January 23). Team-level interventions for improving probability judgments.

Whereas the techniques in Week 2 can be practiced in private, in total isolation from others, those in Week 3 require working with other human beings. We will examine targeted-questioning tools that you can use to hold yourselves and each other accountable for rigorous, "evidence-based" reasoning about tough problems. And we will explore the classic concepts of "process gain" vs. "process loss" in teamwork: when can teams become more than the sum of their parts—and when are they likely to wind up being less?

We also address the nagging question: If keeping score is such a great idea, why do so few companies do it? The short answer is "tradition" (see "Fiddler on the Roof"): we are locked into the comfortable habit of relying on vague verbiage in expressing uncertainty. Vague verbiage serves valuable functions. We can claim credit when we are vaguely right (I said it could happen—and look!) and dodge blame when vaguely wrong (I only said it could happen). Translating hunches into metrics and then keeping score are hard work—and expose you to career risk (who wants to be caught on the wrong side of close calls?). To overcome this resistance, teams have to shift from a blame culture to an exploration culture: reward those willing to go out on a limb, even if they sometimes fall.

Reading Assignments

Duhigg, C. (2011). *The power of habit*. New York: Random House. Prologue, Chapters 1, 4, 6, and Appendix).

Hackman, J. R. (2009). *Why teams don't work*. Harvard Business Review.

Heath, C., Larrick, R., and Klayman, J. (1998). *Cognitive repairs: How organizational practices can compensate for individual shortcomings*. Research in organizational behavior.

Week 4 (Jan. 30). Miracles of Aggregation: Algorithmic solutions to improving probabilistic forecasts.

This session looks at ways of improving human predictions that require no cooperation from human forecasters. We explore why consensus forecasts are often more accurate than most, sometimes even all, of the individual forecasters from whom the consensus was derived. We also

explore new methods—statistical and psychological—of distilling wisdom from the crowd—and of applying these insights to business problems that the class cares about.

Tentatively scheduled guest speaker: Lyle Ungar, Professor of Computer Science, University of Pennsylvania, and Winning Algorithm Developer in IARPA tournament, 2011-2012.

Reading Assignments

Bonabeau, E. (2009). *Decisions 2.0: The power of collective intelligence*. MIT Sloan Management Review.

Schreiber, E. (2011). *Prediction markets*. In Jon Elster (ed). Crowd Sourcing.

Surowiecki, J. (2004). *The wisdom of the crowd.* Chapters 1-2.

Week 5 (Feb. 6): Optimal forecasting frontier: How to tell when you have hit the point of diminishing marginal predictive returns in efforts to improve foresight.

Thoughtful observers often hold different views of how much room there is for improving foresight. Skeptics stress the radical unpredictability of key outcomes (the Black Swan thesis) whereas "meliorists" stress the power of various tools—statistical, game-theoretic, etc—to boost overall performance (and not just shift the ratios of false-positive and false-negative mistakes that forecasters make as they bounce back and forth in blame games of accountability ping-pong in which the object is "avoid the most recent type of mistake").

Tentatively scheduled guest: Nassim Taleb (author of The Black Swan)

Reading Assignments

Bueno de Mesquita, B. (2009). *The predictioneer's game*. New York: Random House. Introduction-Chapter 3 and Chapters 8-11.

Taleb, N. et al. (2009). *Six mistakes executives make in risk management*. Harvard Business Review.

Tetlock, P.E. & Mellers, B. (2011). *The intelligent management of intelligence agencies: Beyond accountability ping-pong*. American Psychologist.

Week 6 (Feb. 13): Trade-Offs in Efforts to Improve Foresight. The Case of Morgan Stanley.

Executives routinely confront tough judgment calls about: (a) how close they are to the optimal forecasting frontier (beyond which improvements in foresight are impossible); (b) insofar as improvements are possible, whether existing offerings in the marketplace of ideas -- consultants,

technologies, etc. -- can help;(c) the pros and cons of evaluating staff using either process-accountability or outcome-accountability performance metrics.

Executives also face even deeper trade-offs. One is the stability-volatility trade-off: tools that improve foresight in stable worlds may degrade sensitivity to discontinuities-- and tools that sensitize us to black swans may degrade foresight in stable worlds. Another is the private-public trade-off: modes of thinking most conducive to making accurate probability judgments are not necessarily most conducive to sealing deals with clients or rising up the hierarchy.

Tentatively scheduled guest: Juan-Luis Perez, Global Head of Research, Morgan Stanley

Reading Assignments

Edmondson, A. (2011). *Strategies for learning from failure*. Harvard Business Review.

Srinivasan, S. & Lane, D. (2011). *The risk-reward framework at Morgan Stanley research.* Harvard Business School Case Study N9-111-011.

Schoemaker, P. & Day, G. (in press). *Integrating organizational networks, weak signals, strategic radars and scenario planning.* Technological Forecasting and Social Change.

Week 7 (Feb. 20): Resolving Debates Over How to Balance Process and Outcome Accountability in a Rapidly Changing World.

Some scholars (e.g., Nobel Laureate Daniel Kahneman) suggest that the best way of improving judgment is to institutionalize process safeguards that pressure people to override their often flawed (System 1) intuitions. In this view, managers should rein in discretion by creating process accountability guidelines that are grounded in rigorous statistical models of reality and permit few exceptions (in situation x, give these weights to these cues in prrediction). Other scholars concede the value of process but worry that it can ossify into bureaucratic ritualism—so we also need forms of outcome accountability that incentivize people to look aggressively for situations that require challenging process rules. (Which grading system did you choose for this class?)

Reading Assignments

Kahneman, D. (2011). *Thinking: Fast and slow*. New York: Farrar, Straus and Giroux. Chapters 15 and 16.

Simons, R. (2010). *Accountability and control as catalysts for strategic exploration and exploitation: Field study results*. Harvard Business School Working Paper 10-051.

Tetlock, P.E., & Mellers, B.A. (2011). *Structuring accountability systems in organizations: Key trade-offs and critical unknowns.* In Intelligence Analysis: Behavioral and Social Scientific Foundations. National Academy of Science Press. Washington, DC (pp. 249-270).

Week 8 (Feb. 27): Closing Summary—and Class Presentations.

We close with my integrative overview of course themes—and with students' presentations on how they have incorporated key class concepts into their business plans and work lives. I stress value of getting into the habit of testing one's judgments -- and creating a feedback-loop platform for lifetime learning.

Optional Readings

Week 1
Silver, N. (2012). *The signal and the noise*. Chapters 1 (A catastrophic failure of prediction) and 2 (are you smarter than a TV pundit?).

Terwiesch, C. & Ulrich, K. T. (2009). *Innovation tournaments*. Harvard Business School Press.

Tetlock, P. E. (2009). *Playing tarot on K Street*. The National Interest

Week 2: Optional (for those with special interests in political risk analysis for business):
Bremmer, I. & Keat, P. (2009). The fat tail: The power of political knowledge in an uncertain world. Oxford University Press, chapters 1, 2, and 3.

Gilovich, T. (1998). *How we know what is not so: The fallibility of human reason in everyday life*. Chapters 2-4.

Week 3 Optional
Page, S. (2007). *The difference: How the power of diversity creates better groups, firms, schools and societies.* Princeton University Press. Chapters 4 and 8.

Week 4 Optional
Watts, D. (2011). *Everything is obvious: How common sense fails us*. New York: Crown. Chapters 3-7.

Wolfers, J. and Zitzewitz, E. (2004). *Prediction markets*. Journal of Economic Perspectives

Week 5 Optional
Silver, N. (2012). *The signal and the noise*. Chapters 5-7.

Taleb, N. (2008). *The Black Swan*. Chapters, 1 and 7 and 9.

Week 6 Optional
Schoemaker, P. & Tetlock, P. E. (2012). *Taboo Scenarios: How to think about the unthinkable*. California Management Review.

Week 7 Optional
Kahneman, D. (2011). *Thinking: Fast and slow*. New York: Farrar, Straus and Giroux. Chapters 1, 2.