This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

SBE 2020: A Complete Theory of Human Behavior

Andrew W. Lo*, September 30, 2010

Abstract

I propose the following grand challenge question for SBE 2020: can we develop a complete theory of human behavior that is predictive in all contexts? The motivation for this question is the fact that the different disciplines within SBE do have a common subject: *Homo sapiens*. Therefore, psychological, sociological, neuroscientific, and economic implications of human behavior should be mutually consistent. When they contradict each other—as they have in the context of financial decisions—this signals important learning opportunities. By confronting and attempting to reconcile inconsistencies across disciplines, we develop a more complete understanding of human behavior than any single discipline can provide. The National Science Foundation can foster this process of "consilience" in at least four ways: (1) issuing RFPs around aspects of human behavior, not around disciplines; (2) holding annual conferences where PI's across NSF directorates present their latest research and their most challenging open questions; (3) organizing "summer camps" for NSF graduate fellowship recipients at the start of their graduate careers, where they are exposed to a broad array of research through introductory lectures by NSF PI's; and (4) broadening the NSF grant review process to include referees from multiple disciplines.

If economists could manage to get themselves thought of as humble, competent people on a level with dentists, that would be splendid.

— John Maynard Keynes, 1931

* Harris & Harris Group Professor, MIT Sloan School of Management, and Chairman and Chief Investment Strategist, AlphaSimplex Group, LLC. Please direct all correspondence to: Andrew W. Lo, MIT Sloan School of Management, 100 Main Street, E62–618, Cambridge, MA 02142. I would like to thank Dr. Myron Gutmann of the National Science Foundation's Social, Behavioral, and Economic Sciences Division for soliciting white papers on the grand challenge questions facing our respective disciplines, and Dr. Daniel Newlon for encouraging me to submit a response. By nature, such an exercise is meant to be speculative, not rigorous; nevertheless, I feel an obligation to apologize in advance to my academic colleagues for the informal and discursive nature of this essay. Research support from the National Science Foundation (SES–0624351) and the MIT Laboratory for Financial Engineering is gratefully acknowledged. This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit:
http://creativecommons.org/licenses/by-nc-sa/3.0/
or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

I believe the most important grand challenge question facing the NSF's Social, Behavioral, and Economic Sciences Directorate is relatively easy to state, but extraordinarily difficult, if not impossible, to achieve by 2020. *Can we develop a complete theory of human behavior that is predictive in all contexts?*

That this should be *the* grand challenge question for SBE 2020 is by no means clear. But before attempting to defend this proposal, let me explain more fully what the question asks. By "all contexts", I mean all situations in which humans may find themselves, including economic, social, cultural, political, and physical. By "predictive", I mean an empirically validated and repeatable cause-and-effect relation. And by "complete theory", I mean a theory that is consistent with all known facts of human behavior, and which is sufficient for making correct predictions of human behavior in novel contexts.

The motivation for seeking an answer to this ambitious question is the simple observation that the social, behavioral, and economic sciences have a single common focus: *Homo sapiens*. Because these disparate fields share the same object of study, their respective theories must be mutually consistent when there is any overlap in their implications. For example, anthropological theories of mating rituals must be consistent with the biology of human reproduction; otherwise flaws exist in one or both of these bodies of knowledge. Of course, in many cases, implications may not overlap. The particular mechanisms of genetic mutation have no direct bearing on the sources of time-varying stock market volatility, so checking for consistency between the former and the latter is unlikely to yield new insights. But because all SBE disciplines involve the study of the very same *human* behaviors and institutions, opportunities for consistency checks should arise often.

One of the most prominent inconsistencies among the SBE disciplines is the rational expectations paradigm of economics and the many behavioral biases documented by psychologists, behavioral economists, sociologists, and neuroscientists. Rational expectations, and its close cousin, the efficient markets hypothesis, have come under fire recently because of their apparent failure in predicting and explaining the current financial crisis. Some of this criticism is undoubtedly unwarranted populist reactions to the life-altering economic consequences of the national decline in U.S. residential real-estate prices from 2006 to 2009. In such an emotionally charged atmosphere, it is easy to forget the many genuine breakthroughs

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

that have occurred in economics over the last half-century such as general equilibrium theory, game theory, growth theory, econometrics, portfolio theory, and option-pricing models. But any virtue can become a vice when taken to an extreme. The fact that the 2,319-page Dodd-Frank financial reform bill was signed into law on July 21, 2010, six months before the Financial Crisis Inquiry Commission is scheduled to report its findings, and well before economists have developed any consensus on the crisis, underscores the relatively minor scientific role that economics apparently plays in policymaking. Imagine the FDA approving a drug before its clinical trials are concluded, or the FAA adopting new regulations in response to an airplane crash before the NTSB has completed its accident investigation.

There are legitimate arguments that the rigorous and internally consistent economic models of rational self-interest—models used implicitly and explicitly by policymakers, central bankers, and regulators to formulate laws, manage leverage, and rein in risk-taking in the economy—have failed us in important ways over the past decade. Even the most sophisticated stochastic dynamic general equilibrium models did not account for the U.S. housing market boom and bust, nor were they rich enough to capture the consequences of securitization, credit default insurance, financial globalization, and the political pressures influencing Fannie Mae and Freddie Mac.

But rather than discarding rationality altogether, a more productive response is to confront the inconsistencies between economic models of behavior and those from other disciplines, and attempt to reconcile them and improve our models in the process. While frustrating, contradictions often present opportunities for developing a deeper understanding of the phenomena in question.

Consider the example of probability matching: an experimenter asks a subject to guess the outcome of a coin toss, where, unknown to the subject, the coin is biased—75% heads and 25% tails—and the experimenter agrees to pay the subject $1 if she guesses correctly, but will expect the subject to pay $1 if she guesses incorrectly. This experiment is then repeated many times with the same subject and coin (and the tosses are statistically independent). After a sufficiently long sample of tosses, it should be possible for the subject to observe that the coin is biased toward heads, at which point the subject should always guess heads so as to maximize her cumulative expected winnings. However, the vast majority of subjects do not follow this

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

expected-wealth-maximizing strategy; instead, they appear to randomize, guessing heads 75% of the time and tails 25% of the time!

This strange and well-known example of irrationality in human judgment may not be so irrational after all when viewed from the perspective of evolutionary biology (Lo and Brennan, 2009). To see why, consider the hypothetical case of animal deciding whether to build its nest in a valley or on a plateau. If the weather is sunny, nesting in the valley will provide shade, leading to many offspring, whereas nesting on the plateau provides no cover from the sun, leading to no offspring. However, the opposite is true if the weather is rainy: the valley floods, hence any offspring will drown in their nests, but nests on the plateau survive, yielding many offspring. Now suppose the probability of sunshine is 75% and the probability of rain is 25%. The "rational" behavior for all individuals to follow is to build their nests in the valley, for this maximizes the expected number of each individual's offspring. Suppose the entire population exhibits such individually optimal behavior—the first time there is rain, the entire population will cease to reproduce, leading to extinction. Similarly, if the entire population behaves in the opposite manner, always choosing the plateau, the first time sunshine occurs, extinction also follows. Lo and Brennan (2009) show that the behavior that maximizes the growth of the population is for individuals to randomize their nesting choice by choosing the valley with probability 75% and the plateau with 25% probability. Matching probabilities confers an evolutionary advantage, not for the individual, but rather for the population as a whole. And since, by definition, the current population consists of the survivors, it will reflect such advantageous behavior disproportionately to the extent that behavior is heritable. While probability matching is, indeed, irrational from the perspective of maximizing an individual's expected wealth, its evolutionary advantage is clear.

This broader perspective suggests that the economic notion of rationality is not wrong, but simply incomplete—humans usually do maximize their expected wealth but, under certain circumstances, they may engage in other types of "hard-wired" behavior that are far more primitive. Probability matching is likely to be a vestigial evolutionary adaptation that may not increase the chances of survival in the current environment, but nevertheless is still part of our behavioral repertoire. Using a simple binary choice model, Brennan and Lo (2009) show that

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

several commonly observed behaviors such as risk aversion, loss aversion, and randomization are adaptive traits that can emerge organically through evolution.

The natural follow-on question—one that lies at the heart of the grand challenge question posed above—is why do we choose one particular behavior from our repertoire for a given occasion and not another, and how does that repertoire change over time and across circumstances? The answer to this question has obvious consequences for virtually all economic models, yet the tools by which we will solve this challenge may come from other disciplines such as psychology, neuroscience, ecology, and evolutionary biology. Other examples of important questions about economic behavior that fall outside standard economics are:

- How do emotions affect the stability of preferences over time and circumstances?
- What role does memory play in economic decisionmaking?
- What do "theory of mind" experiments imply for strategic behavior?
- Can robust optimal control explain the regulatory challenges of fast-paced innovation?
- Does network analysis provide new insights for systemic risk in the financial system?

By reconciling the inconsistencies and contradictions between disciplines, we can develop a broader and deeper understanding of *Homo sapiens*.

These examples illustrate the value of "consilience", a term re-introduced into the popular lexicon by E. O. Wilson (1998), who attributes its first use to William Whewell's 1840 treatise *The Philosophy of the Inductive Sciences*, in which Whewell wrote "The Consilience of Inductions takes place when an Induction, obtained from one class of facts, coincides with an Induction, obtained from another different class. This Consilience is a test of the truth of the Theory in which it occurs." In comparing the rate of progress in the medical vs. the social sciences, Wilson (1998, p. 182) makes a thought-provoking observation:

> There is also progress in the social sciences, but it is much slower, and not at all animated by the same information flow and optimistic spirit…
>
> The crucial difference between the two domains is consilience: The medical sciences have it and the social sciences do not. Medical scientists build upon a coherent foundation of molecular and cell biology. They pursue elements of health and illness all the way down to the level of biophysical chemistry…

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

> Social scientists by and large spurn the idea of the hierarchical ordering of knowledge that unites and drives the natural sciences. Split into independent cadres, they stress precision in words within their specialty but seldom speak the same technical language from one specialty to the next.

This is a bitter pill for economists to swallow, but it provides a clear directive for improving the status quo.

Although economics occupies an enviable position among the social sciences because of its axiomatic consistency and uniformity, *Homo economicus* is a fiction that can no longer be maintained in light of mounting evidence to the contrary from allied fields in SBE. For disciplines in which controlled experimentation is possible, consilience may be less critical to progress because inconsistencies can be generated and resolved within the discipline through clever experimental design. But for disciplines such as economics in which controlled experimentation is more challenging, consilience is an essential means for moving the field forward. And even in fields where experiments are routine, consilience can speed up progress dramatically. The revolution in psychology that transformed the field from a loosely organized collection of interesting and suggestive experiments and hypotheses to a bona fide science occurred only within the last three decades, thanks to synergistic advances in neuroscience, medicine, computer science, and even evolutionary biology. This could be the future of economics.

The NSF's SBE Directorate has a unique opportunity to foster consilience in the Social, Behavioral, and Economic sciences by taking up the grand challenge question proposed at the start of this essay. Developing a complete theory of human behavior that is truly predictive in all contexts will require contributions from and collaborations between many disciplines: economics, engineering sociology, anthropology, psychology, neuroscience, ecology, evolutionary biology, and computer science. However, unlike the usual inter-disciplinary grants—which are often as effective as arranged marriages—RFPs centered on particular aspects of human behavior rather than specific disciplines will naturally draw the relevant fields together in productive ways.

Beyond issuing new RFPs, the NSF can encourage consilience through other means. Holding annual conferences at NSF in which principal investigators from difference disciplines

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

are invited to come together to share their latest research, as well as their frustrations and open challenges, would be a natural extension of the NSF's activities. Providing "summer camps" for NSF graduate fellowship recipients at the start of their graduate careers, where they are exposed to a broad array of NSF PIs, who would be asked to deliver overview lectures about the biggest challenges in their respective disciplines, is another way to "seed" the next generation of scholars. Finally, changing the very review process of NSF grants to be more cross-disciplinary may create greater diversity in the type of research conducted, increasing the likelihood of consilience in the SBE Directorate and across the entire NSF research portfolio.

References

Brennan, T. and A. Lo, 2009, "The Origin of Behavior", to appear in *Quarterly Journal of Finance*.

Lo, A. and M. Mueller, 2010, "WARNING: Physics Envy May Be Hazardous To Your Wealth", *Journal of Investment Management* 8, 13–63.

Wilson, E., 1998, *Consilience*. New York: Alfred A. Knopf.