The Origin of Bounded Rationality and Intelligence[1]

ANDREW W. LO

Charles E. and Susan T. Harris Professor, MIT Sloan School of Management
Principal Investigator, MIT Computer Science and Artificial Intelligence Laboratory
Director, MIT Laboratory for Financial Engineering
Chief Investment Strategist, AlphaSimplex Group, LLC

ABSTRACT. *Rational economic behavior in which individuals maximize their own self-interest is only one of many possible types of behavior that arise from natural selection. Given an initial population of individuals, each assigned a purely arbitrary behavior with respect to a binary choice problem, and assuming that offspring behave identically to their parents, only those behaviors linked to reproductive success will survive, and less successful behaviors will disappear exponentially fast. This framework yields a single evolutionary explanation for the origin of several behaviors that have been observed in organisms ranging from bacteria to humans, including risk-sensitive foraging, risk aversion, loss aversion, probability matching, randomization, and diversification. The key to understanding which types of behavior are more likely to survive is how behavior affects reproductive success in a given population's environment. From this perspective, intelligence is naturally defined as behavior that increases the likelihood of reproductive success, and bounds on rationality are determined by physiological and environmental constraints.*

I WOULD LIKE TO thank Dr. Harriet Zuckerman for inviting me to speak before the American Philosophical Society. I have to say that it is a particular pleasure and privilege for me to be delivering this talk on bounded rationality and intelligence to the society founded by one of the earliest

and most astute students of human behavior.

I am referring, of course, to Benjamin Franklin, and to one of his many insights into the foibles of *Homo sapiens*, the so-called "Benjamin Franklin effect." Franklin first stumbled upon this anomaly when he asked to borrow a book from a member of Congress who was particularly antagonistic toward him. To his amazement, this individual agreed to lend him the book and afterward became much more friendly, which Franklin (1987, 1404) summarized in the maxim "He that has once done you a kindness will be more ready to do you another than he whom you yourself have obliged." Most of you are probably more familiar with the modern translation of this insight: "No good deed goes unpunished."

The study of human behavior is common to many scholars and disciplines. Anthropology, archaeology, biology, computer science, economics, engineering, history, psychology and the cognitive neurosciences, and sociology all involve human behavior to some degree. Of course, the depth engagement varies across these disciplines—computer science touches upon human behavior only within the narrow subfield of artificial intelligence, whereas all of psychology is dedicated to it. However, despite the difference in emphasis, the various depictions of human behavior across disciplines ought to be mutually consistent. For example, anthropological studies of mating behavior in hunter-gatherer societies should be consistent with the biology of human reproduction, e.g., the physiology of sexual reproduction, gestation lags, and the genetic basis of heredity. It would be a strange state of affairs if anthropologists had a model of human reproduction different from that of biologists, even though anthropology and biology have vastly different objectives, methods, and models. After all, with respect to human behavior, both fields are focusing on a common subject: *Homo sapiens*.

This basic scientific principle of cross-disciplinary consistency—known more recently as "consilience" (Wilson 1998)—can be used to leverage insights from one field to produce discoveries in several others. In re-introducing this term into the popular lexicon, Wilson 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." For example, the physics of radioactive decay of carbon-14 atoms provided a remarkably accurate method for determining the age of fossils, revolutionizing the field of archaeology (Libby 1946). But inconsistencies across disciplines can also yield new insights through the identification and resolution of such conflicts.

A case in point, and the subject of my talk today, is the idea of *Homo economicus*, rational economic man. Virtually all economic models share the

common assumption that individuals behave rationally at all times and under all circumstances, seeking solely to further their own self-interest subject to resource constraints. And when individually rational parties are allowed to interact freely, Adam Smith's "invisible hand" is the result, guiding scarce resources to their most valuable uses while equalizing supply and demand across all consumers, producers, and markets. The influence of this paradigm goes far beyond academia—it underlies current macroeconomic and monetary policies, and has also become an integral part of the rules and regulations that govern financial markets today (see, e.g., Kocherlakota 2010 and Hu 2012).

However, an accumulation of evidence from psychology, behavioral economics and finance, the cognitive neurosciences, and biology have highlighted significant and persistent contradictions between *Homo economicus* and *Homo sapiens* (Kahneman et al. 1982; Thaler 1993). The financial crisis of 2007–09 has only created more skepticism about the neoclassical economic paradigm of individual and aggregate rationality.

In this talk, I propose to reconcile these contradictions by showing that evolutionary principles applied to a simple binary-choice model developed by Thomas J. Brennan and me can fully explain several common behavioral anomalies (Brennan and Lo 2011; 2012). A key feature of this model is the way the environment affects an individual's reproductive success—in one environment, natural selection favors individually rational decisions, but in another environment, natural selection punishes rationality in favor of random behavior. Instead of focusing our attention on the decision maker, we should be studying the environment in which individuals are making their decisions.

By understanding the role of the environment in shaping behavior, we not only reconcile rationality with various anomalies, but also obtain a natural definition of intelligence and a framework for determining its limits: intelligence is any behavior that increases the likelihood of reproductive success, and it is ultimately bounded by the biological cost of engaging in such behavior.

This is a rather bold claim. Rather than attempt to defend it at the outset, let me begin with a simple example.

1. An Example: Probability Matching

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

A practical example of this phenomenon comes from the pilots and crew members in the U.S. Army Air Force who flew bombing missions over Germany during World War II.[2] Before each mission, bomber crews had to make a choice as they boarded their aircraft: should they wear a parachute or a flak jacket? Parachutes were much bulkier in those days, and flak jackets were even bulkier because steel plates were sewn into their linings (this was well before Kevlar was invented). As a result, crew members couldn't wear both; they had to choose one or the other. Parachutes were needed in case the bomber was shot down by German ground-based anti-aircraft artillery, and flak jackets protected the soldiers' bodies from the shrapnel of the artillery's high-explosive shells, which often pierced the plane's fuselage.

Bomber crews knew that the chances of getting shot down were considerably less than those of getting hit with anti-aircraft fire. They also knew that each bombing mission was plausibly unrelated to the previous one, so there was no carryover in terms of the odds. The rational choice would have been to wear a flak jacket all the time. But that's not what the bomber crews chose. It turns out that they switched between parachutes and flak jackets, roughly in proportion to their chances of getting shot down versus getting hit with anti-aircraft fire—in other words, they probability matched! This behavior frustrated military officials, who felt responsible for the young men who risked their lives on these deadly missions, but they were unsuccessful in changing the bomber crews' choices.

2. An Evolutionary Basis for Probability Matching

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, as we show in Brennan and Lo (2011). To see why, consider the hypothetical case of an 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 three 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

---

[2]I am grateful to Steve Ross for providing me with this fascinating illustration of probability matching. Ross credits Amos Tversky with this example, but Tversky passed away before we could confirm the details.

nests, but nests on the plateau survive, yielding three 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. Now 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.

In fact, the behavior that maximizes the growth of the population is for individuals to randomize their nesting choice by choosing the valley with 75% probability and the plateau with 25% probability (Brennan and Lo 2011). In other words, probability matching behavior, also known as "Herrnstein's Law" (Herrnstein 1961; Cooper and Kaplan 1982; Brennan and Lo 2011), is evolutionarily dominant in this special case. However, by dominant, we mean that probability matching 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.

Table 1 provides a numerical simulation of this simple ecology for several possible behaviors including deterministic choice and randomized choice with choice probability f=20%, 50%, 75%, and 90%. The values in the right-most column of table 1 show that always choosing the valley does, indeed, maximize the growth rate as long as it doesn't rain, but once it does, extinction occurs. The behavior that maximizes the expected growth rate across all conditions is the probability-matching behavior f=75%.

In Brennan and Lo (2011), we develop a general mathematical framework for analyzing the evolutionary properties of the binary choice model and show that, although probability matching arises only in certain special cases, there are many other types of environments in which the evolutionarily dominant behavior is not deterministic. In particular, when faced with environmental randomness that affects the entire population in the same manner—specifically, when fecundity is the same random variable across all individuals—and where the type of randomness yields extreme outcomes for different behaviors, deterministic behavior cannot survive because at some point an extreme outcome will occur, wiping out that subpopulation. The only way to survive is to randomize.

However, when fecundity is not the same random variable across all individuals, randomization may no longer be the dominant behavior. In fact,

Generation	f = .20	f = .50	f* = .75	f = .90	f = 1
1	21	6	12	24	30
2	12	6	6	57	90
3	6	12	12	144	270
4	18	9	24	387	810
5	45	18	48	1,020	2,430
6	96	21	108	2,766	7,290
7	60	42	240	834	21,870
8	45	54	528	2,292	65,610
9	18	87	1,233	690	196,830
10	9	138	2,712	204	590,490
11	12	204	6,123	555	1,771,470
12	36	294	13,824	159	5,314,410
13	87	462	31,149	435	15,943,230
14	42	768	69,954	1,155	0
15	27	1,161	157,122	3,114	0
16	15	1,668	353,712	8,448	0
17	3	2,451	795,171	22,860	0
18	3	3,648	1,787,613	61,734	0
19	9	5,469	4,020,045	166,878	0
20	21	8,022	9,047,583	450,672	0
21	6	12,213	6,786,657	1,215,723	0
22	0	18,306	15,272,328	366,051	0
23	0	27,429	34,366,023	987,813	0
24	0	41,019	77,323,623	2,667,984	0
25	0	61,131	173,996,290	7,203,495	0

TABLE 1. Simulated population sizes for binary-choice model with five sub-populations in which individuals choose *a* with probability *f* and *b* with probability 1–*f*, where *f* = 0.20, 0.5, 0.75, 0.9, 1, and the initial population is 10 for each *f*. Reproductive uncertainty is systematic and also binary, with Prob(μ_a=3 , μ_b=0) = 0.75 and Prob(μ_a=0 , μ_b=3)=0.25. In this setting, probability matching *f**=0.75 is the growth-optimal behavior.

suppose the number of offspring produced by each individual is a random variable that is unique to that individual, i.e., the number of offspring produced by each individual is independently and identically distributed across the entire population. This is the case of "idiosyncratic" reproductive risk because each individual's realized number of offspring is an independent draw, albeit from the same distribution, and each realization is statistically unrelated

to any other individual's realization.

In this case, we find that the evolutionarily dominant behavior is deterministic: always choose the behavior that yields the higher expected number of offspring. The intuition for this remarkably general result is straightforward: because fecundity is independently and identically distributed across the population, even if all individuals choose action a, the chance that all realizations of this choice lead to extinction is exceedingly remote. However, in the case of systematic reproductive risk, all individuals choosing a obtain the identical number of offspring; if this realization is 0, the entire population is wiped out.

These two extreme cases—systematic and idiosyncratic reproductive risk—can be summarized by the observation that when reproductive risk is systematic, natural selection favors randomizing behavior to avoid extinction, and when reproductive risk is idiosyncratic, natural selection favors deterministic behavior because it yields a higher growth rate without risking extinction. In other words, "Nature abhors an undiversified bet."

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—namely, environments with systematic reproductive risk—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. Even in our simple binary choice model, we find that several commonly observed behaviors such as risk aversion, loss aversion, and randomization are adaptive traits that can emerge organically through evolution.

3. Bounded Rationality

To illustrate this process of adaptive behavior, consider the problem everyone confronts each day: deciding what to wear. For concreteness, my wardrobe consists of 5 jackets, 10 pairs of pants, 20 ties, 10 shirts, 10 pairs of socks, 4 pairs of shoes, and 5 belts. Now this may seem like a rather limited collection, but the various combinations of these items amount to two million unique outfits. Granted, not all of these outfits are equally compelling from a fashion perspective, hence the decision problem: I want to select the outfit that will maximize the impact of my appearance. Suppose it takes me one second to evaluate the fashion content of each outfit—how long will it take me to get dressed? The answer is 23.1 days!

So how am I able to get dressed in less than five minutes? First, I do not optimize; specifically, I do not maximize any utility function subject to a wardrobe constraint. Instead, I use a heuristic that has been honed through years

of trial-and-error and experience. For example, when I was six years old, the comic-book and television character Superman was the superhero of choice, and some clever marketing executive realized that placing Superman emblems on children's clothing could boost sales dramatically. I nagged my mother for weeks to buy me a jacket with a Superman emblem until she finally gave in. On the following Monday morning, I spent so much time in front of the mirror modeling my new jacket that I was late for school. So late, in fact, that I needed a note from my mother to explain why I was late before I could enter the classroom.

I was completely mortified to walk into my classroom so late, which is confirmed by the fact that, forty-three years later, I still remember that event as though it had happened this morning. From that day on, it never took me more than five minutes to get dressed. This is a concrete example of emotion working together with logical deliberation to create a heuristic through the machinations of a six-year-old child's brain, which permanently altered my behavior.

Heuristics provide useful sequences of actions, i.e., behaviors, that are adapted to solve specific environmental challenges. The economist/psychologist/computer scientist Herbert Simon (1955) created a new term for this kind of behavior—"satisficing"—to distinguish it from the kind of optimizing behavior that economists routinely assume. Satisficing behavior is not necessarily optimal, but is, instead, satisfactory. Almost immediately, this notion of "bounded rationality" was attacked by Simon's critics, who asked how one knows what "satisfactory" means; what determines the limits to rationality? In Brennan and Lo (2011; 2012), we provide an answer: the limits are determined by natural selection operating on heuristics in various environments. In one environment, a particular heuristic may yield near-optimal results for an individual, in which case the heuristic will persist. However, in a very different environment, that same heuristic may be disastrous, in which case it will eventually be eliminated through the forces of natural selection. The graceful undulations of a great white shark pursuing its prey in deep water are a fearsome spectacle to behold, but those very same undulations of a beached shark look clumsy and comical.

In my own case, getting dressed in under five minutes worked well until I showed up at a friend's wedding rehearsal dinner in sneakers and jeans; embarrassed yet again, I developed a more nuanced heuristic for choosing the proper attire for a given occasion. Bounded rationality has more to do with natural selection of heuristics under specific environmental constraints than with whether individuals are rational or irrational.

4. Intelligence

These examples bring us to the main focus of my talk today—what is intelligence and where does it come from? Many educators and learned scholars

of psychometric testing would argue that it is impossible to come up with a single and simple definition of intelligence, particularly given contemporary social mores on the importance of teamwork, collective intelligence, and "emotional quotient" or E.Q. However, from an evolutionary perspective, we can formulate a definition of intelligent behavior that is at once simple, general, and operationally meaningful: any behavior that confers increased reproductive success to the individual.

To fully appreciate the implications of this proposed definition, a few examples and counter-examples may be helpful. When we deem a professional colleague of ours to be "intelligent," what do we mean? This term is used sparingly in the academy, and refers neither to the mental gymnastics of an idiot savant nor to the encyclopedic knowledge of someone with eidetic memory. Instead, we reserve this most coveted accolade for those capable of developing new and important insights from the same set of facts available to the rest of us. But what do we mean by "new and important" insights? These are new ideas that, by consensus and through empirical verification using the scientific method, allow us to better understand, and, therefore, better predict and control our destiny. Newton's law of gravity, Darwin's theory of evolution, and Einstein's theory of relativity are examples of extraordinary feats of intelligence and they have significantly increased the reproductive success of our species.

These philosophical musings have a rather concrete instantiation from an evolutionary perspective on behavior: intelligence is any behavior that is positively correlated with reproductive success. In the binary choice model of Brennan and Lo (2011; 2012), this definition can be made mathematically precise. Let I_{it} denote the binary choice variable of individual i in generation t, where $I_{it}=1$ represents a choice of action a, leading to a random number x_a of offspring, and $I_{it}=0$ represents a choice of action b leading to a different random number x_b of offspring. Then our evolutionary definition of intelligence is captured by the condition that I_{it} is positively correlated with x_a-x_b, i.e., when i chooses x_a, x_a-x_b tends to be larger and when i chooses x_b, x_a-x_b tends to be smaller.

Now this simple condition may be a far cry from Einstein's theory of relativity, but in Brennan and Lo (2012) we argue that the difference may be more a matter of degree than mechanism. In a binary choice model where all forms of behavior are equally represented in an initial population, and all correlations—positive, negative, and zero—are also equally represented, natural selection will eventually yield a population dominated by individuals with behavior that is maximally correlated with their reproductive success. An ecologist observing this population after many generations will see one group of individuals emerge with the following characteristic: more often than not, their choice of a or b will tend to coincide with the higher-yielding

of the two random variables x_a or x_b. The group with the highest correlation between I_{it} and x_a-x_b will eventually dominate the population, and will appear to "know" the right thing to do, especially in comparison with other individuals with lower correlations. This is the beginning of intelligence.

Skeptics might balk at the simplicity of the binary choice framework, arguing that this can hardly be considered true intelligence because correlation between actions and reproductive success can be purely coincidental, having nothing to do with intention, planning, and deliberate execution. But this is precisely the point of our model—in its most primitive form, intelligence need not be intentional; it only needs to offer reproductive advantage. In this respect, intelligent binary choice is no different from any other heritable trait such as webbed feet, sharp teeth, or an exoskeleton that gives an individual a leg up on its competitors. The difference is perhaps more semantic than substantive—we use "intelligence" to describe adaptive *behavior,* whereas we use "adaptation" to describe adaptive *traits*.

Of course, the binary choice framework is silent on the precise mechanism through which correlation between behavior and reproduction arises. This level of generality is both intentional and significant because it demonstrates that certain behaviors are universal across both individuals and species, persisting irrespective of the specific biological structures involved. For example, positive correlation between behavior and reproduction might emerge through pattern recognition of protected nesting sites, from superior foraging techniques, or from sheer physical prowess in defending offspring. Despite the different types of adaptive behavior in each of these contexts, the impact on fecundity may be the same, in which case these behaviors are equally likely to be represented in the population through natural selection. Moreover, achieving positive correlation between behavior and reproductive outcomes is not costless, but almost always imposes physiological burdens that reduce fitness. The trade-off between these costs and the benefits of positive correlation yields an explicit bound on intelligence. The mathematical foundations of this trade-off are developed fully in Brennan and Lo (2012).

5. Conclusion

Although economics occupies an enviable position among the social sciences because of its rigorous axiomatic foundations and the uniformity of beliefs among its practitioners, *Homo economicus* is a fiction that can no longer be maintained in light of mounting evidence to the contrary from many other quarters. These inconsistencies illustrate the value of "consilience," both in its absence and in the struggle to achieve it.

In comparing the rate of progress in the medical versus the social scienc-

es, Wilson (1998, 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. . . .
>
> 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 tough medicine for economists to take, but it provides a clear directive for improving the status quo.

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 into 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, and the binary choice model may provide one path to get there.

References

Brennan, T. J., and A. W. Lo. 2011. The origin of behavior. *Quarterly Journal of Finance* 1:55–108.

———.2012. An evolutionary model of bounded rationality and intelligence. *PLoS ONE 7*: e34569. doi:10.1371/journal.pone.0050310.

Cooper, W. S., and R. H. Kaplan. 1982. Adaptive coin-flipping: A decision theoretic examination of natural selection for random individual variation. *Journal of Theoretical Biology* 94:135–51.

Franklin, B. 1987. *The autobiography of Benjamin Franklin.* In *Benjamin Franklin: Writings,* ed. J. L. Lemay. New York: Library of America.

Herrnstein, R. J. 1961. Relative and absolute strength of responses as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behaviour* 4:267–72.

Hu, H. 2012. Efficient markets and the law: A predictable past and an uncertain future. *Annual Review of Financial Economics* 4.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kocherlakota, N. R. 2010. Modern macroeconomic models as tools for economic policy. *The Region* (Federal Reserve Bank of Minneapolis) (May), 5–21.

Libby, W. F. 1946. Atmospheric helium three and radiocarbon from cosmic radiation. *Physical Review* 69:671–72.

Simon, H. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics 69:* 99–118.

Thaler, R. H., ed. 1993. *Advances in behavioral finance*. New York: Russell Sage Foundation.

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