Founding moral reasoning on evolutionary psychology: A critique and an alternative

Saras D. Sarasvathy

4547 Van Munching Hall
R.H. Smith School of Business, University of Maryland
College Park, MD 20742
Phone: (301) 405-9673.
Email: saras@rhsmith.umd.edu

I would like to thank Ed Freeman for giving me the opportunity to write this critique; Anil Menon for incisively reviewing several earlier drafts and being an invaluable comrade-in-arms; and, Ed Hartman for providing thought-provoking critical comments that have helped re-shape the final version in important ways.

Cosmides and Tooby (2000) seek to clarify and illuminate our understanding of how we reason – particularly how we reason about social interaction laced with moral intonations. The method they have chosen involves connecting laboratory experiments with some evidence from the larger body of evolutionary theories including paleo-anthropology and behavioral ecology. My critique of their position has to do with the semantic bankruptcy of second-hand syntax borrowed from the physical and biological sciences and with the fact that such a debt seduces us to completely ignore history. Economics as social physics has for a long time done disservice both to society and physics, by disregarding the variety and reality of empirical evidence in the social sphere on the one hand, and on the other, by failing to acknowledge that physical concepts such as "entropy" "invariance" and "symmetry principles" lose their precision and become meaningless when carried over into what William James would call the "blooming buzzing confusion" of human affairs. Similarly the conceptualization of "Economics as evolutionary biology" or business ethics as consequences of evolutionary adaptation endangers our roots in historical reality while at the same time falling prey to an oversimplified, tautological and superficial understanding of evolutionary adaptation. Moreover, such oversimplification understates the pluralistic views of eminent evolution theorists, including those of Darwin himself, particularly with regard to the richness and variety of non-adaptive and non-selective explanations.

In the interest of limiting the scope of this article, I am going to undertake a critique of just one example of the "ethics as adaptation" story – i.e., the cheater detection mechanism (Cosmides and Tooby, 2000). But, with minor modifications, all my arguments can be applied to the other mechanisms and subroutines in the studies. I have organized my critique in two stages. In the first stage, I will advance arguments as to why I find the particular adaptation story

that the authors advance for their experimental results unpersuasive even when I fully accept the value of their experimental results. In the second stage, I will grant them their adaptation story and critique the implications of such stories for business ethics and for future research.

I concur with, and congratulate the authors on their experimental results. While the Wason tests used in their experiments are novel and interesting, their results are entirely in line with earlier experimental work in cognition and decision making *without invoking an adaptation story*. Cognitive scientists and others have known since the fifties — if not earlier — that problem isomorphs or changes in the domain of logically similar problems do create differential outcomes. This can be seen even in toy problems such as the Tower of Hanoi and the fisherman problems. In other words, problem representations matter (Simon, 1975). Their results are also in line with those of Robin Dunbar who showed that we are better at dealing with social interaction than with scientific argumentation (1996).

To summarize the results of over forty years of related experimental work – we suck (if I may use the technical term!) at solving problems based on formal calculi such as the ones involving constrained optimization or propositional logic or any experientially meaningless problem representation. So what can we conclude from this in terms of the adaptationist story? To begin to answer that, let us examine closely just one of the numerous threads in the tangle of experiments over the last forty years. In the 60's and 70's Bar-Hillel (1980), Tversky and Kahnemann (1982) and others conclusively established the general incompetence of humans in solving statistical choice problems involving base rates. In short, when problems are cast in terms of choosing balls from bins, or deciding which doors to open and other stylized scenarios, human beings consistently ignored base rates. Yet as Gigerenzer, Hell, and Blank (1988) showed, when the same base-rate fallacy problem was cast as a football (soccer) wager, the

proportion of people who do take into account base rates increased dramatically. The moral is not that we evolved to wager on football games. The sober lesson is that meaningful framing based on informative descriptions rooted in previous experiences does a lot more to help solve problems than precise axiomatics or stylistic representations of logical rigor. As Gigerenzer argues rather simply and elegantly, we are capable of "intuitive statistics" whether such an intuition "evolved" in the adaptationist sense or not. Arguments such as Gigerenzer's, so long as they rest on the *sufficiency* of evolutionary explanation, may be useful irrespective of the "truth" of their adaptationist claims. The complications of adaptive *necessity* however, whether embodied in a collection of subroutines or neural circuitry or genes or whatever else, are merely excess material, cluttering our explanations and getting in the way of our understanding.

So without further belaboring the point about our incompetence in using content-free formal logic to any desirable extent, let us examine a different aspect of the experimental results in this study. May be the key result here is not that we are bad at content-free logics, but that we are surprisingly good at "cheater detection" – which is after all the real issue at hand in this study. Let us for the moment accept the result – i.e. the existence of a particular "subroutine" in our brains. Does it follow that this mechanism -- or more accurately, ability -- adaptively evolved over thousands of millenia to solve particular social interaction problems such as the rather implausible ones used in the study?

I submit that the authors make a case for this through the following three points:

1. Throughout our evolutionary history, humans have had to solve similar social contract problems;
2. Cheaters can benefit by not keeping to their end of the bargains (get something for nothing); and,
3. People who can detect cheaters have an advantage over those who cannot.

Ergo, we, the survivors come equipped with a mechanism to detect cheaters.

This may make for an adaptation *story* – and I will argue later that even the story is implausible at best. But it certainly does not make for a good adaptation *theory*. In the simplest case, as I understand it, a true adaptation theory (not story) begins with a heterogeneous population. The population may be grouped into two sets. At any point in time, we can associate an average numerical fitness with each set; the fitness is typically computed based on a set of common measurements (energy consumption, gestation period etc). If one set has a higher fitness than the other, then other things being equal, it is possible to show that the members of that set may be expected to have greater reproductive success. If, for the time under consideration, we can explain the difference in proportions of the two sets as a result of the fitness values measured or estimated for that period, then we may say we have an adaptation theory. Furthermore, it is not enough for an adaptation theory to show something exists – it has to be able to trace its adaptive history as a changing distribution of the population over time.

Typically, the adaptive explanation is couched in terms of regression analysis (this can provide an explanation based on observables rather than on nebulous concepts like "fitness"), but there are weaker alternatives like game theoretic modeling, stochastic modeling, and/or computer simulations. In using these though, there are some important caveats: What is being explained cannot be used as an explanatory variable. Mere existence is not evidence of superior fitness. Superior fitness does not guarantee survival. Also, today's success may be someone's dinner tomorrow! In sum, evolution is about sufficient conditions for survival not necessary ones.

The adaptive explanation for the results of the study under consideration here does not meet the criteria for an adaptation theory. For example, I've seen no proof for the minimal requirement that the ability to solve Wason problems is inheritable. Or for the requirement that this ability leads to improved reproductive success. But perhaps I am unfairly pushing the

argument beyond its scope. As the authors have explained elsewhere, inheritability is not an issue for their studies since they merely use the adaptationist explanation to generate hypotheses. If so, then what is the basis of their claim, "I will present experimental findings suggesting that our minds contain evolved mechanisms specialized for reasoning about problems …" Furthermore, using a theory to generate hypotheses that are not falsified in the results is the same as using the theory for explaining the results, especially since no alternative explanations based on history are even considered or explored. And even more importantly, no alternative hypotheses involving explicit *design* of social contracts to detect and punish cheaters based on collective historical experience were generated and eliminated. Hence I will now try and develop an alternate hypothesis based on history and social action rather than biology.

To do that, I will release the authors from the rigor required of an adaptation *theory* and examine their arguments as a *story* of adaptation instead. Here's how such a story might go: Imagine a group of pre-historic hominoids, some of whose brains are "programmed" with the cheater detection subroutine, or with punitive sentiments against non-participants, and others whose brains are not so "programmed." Then according to the adaptationist argument, the ones with the subroutine have an evolutionary advantage and over millenia increase in numbers in the population until the others who do not have the subroutine become virtually extinct. Fast forwarding through to the present, I am simply overwhelmed by the historical evidence that the exact opposite to the predicted outcome seems to have occurred. According to Thomas McCraw, (1997) as late as the eighteenth century, less than 4% of homo sapiens appear to have managed to convince the other 96% of us that we should hand over the products of our work and even the very essentials of our freedom to their will and disposition. The few cheaters in our midst appear to have successfully sold the majority of us a variety of bad goods and broken promises. Be it the

divine rights of kings, the absolute rule of the husband in the household, well-deserved incarnations of the higher castes in India, or the sub-human status of African slaves in the New World, most of us have been unable to detect the "cheaters" in our midst. For practically all of recorded history, a few political charlatans in cahoots with a handful of religious leaders have sold us "requirements unfulfilled" and have paid for our lives and liberties with "benefits undelivered." It is not all that better today. Whether it is divorce rates or dot com bubbles, we are still for the most part unable to detect "cheaters" in any non-trivial sense. So based on the historical evidence, I could argue that we do not have any evolved cheater detection mechanism at all and that is why we need to develop them through conscious action, such as collective bargaining, legislative initiatives, contractual due diligence and other explicit mechanisms of fair trade *designed* for defining and detecting cheaters, not evolved in any sense in our brains.

What about the counter-argument that the authors are not claiming that the presumed cheater detection mechanisms are omniscient, or capable of resolving complex social contract problems? But then, of what use is a cheater-detection mechanism that can solve toy Wason problems, but is incapable of detecting cheaters who, by spinning unverifiable social contracts, can obtain significant advantages including *reproductive* advantages for themselves (for example, a simple examination of caste rules in India shows very clearly how higher castes have claimed such advantages). All in all, the historical evidence is overwhelmingly against any adaptationist arguments to explain the laboratory results in this study.

But, of course, the results stand on their own and need explanation. What would an alternative explanation of the results be? Let us look at the particular experiment involving the transportation to Boston as opposed to the stolen watch. The reason I fail to detect the logical flaw while solving the transportation to Boston problem is that it is content-irrelevant and even

meaningless in my own previous experience: it is reminiscent of calculus problems that urge you to solve problems involving cooling rates of cups of coffee, or sliding ladders or the crazy fly oscillating between trains heading for a collision! But when someone takes my watch and gyps me for the price, I can easily detect it because both my own previous experience and my memory of our collective experience in history points to this being a cheater problem. Since trade and contractual exchange constitute one of the oldest phenomenon in all cultures -- even tribes at war with each other in ancient times would stop fighting at the end of each day so they could trade with each other! – trade matters to us. Taking the cab or train does not matter all that much – in fact when going to Boston all I care about is not having to drive.

And so most human beings would "get" any problem couched in simple terms of trade or contracts or promises to be fulfilled, with one major caveat. Fair trade assumes that value can be measured and clearly verified upon delivery. When the exact same problem of trade involves quantities such as martyrdom and loyalty and marital fidelity, we will find that the so-called cheater detection subroutine fails us miserably. This can easily be verified by conducting the same experiment as the current authors did except replacing the propositional calculus based descriptive rules with contracts involving non-measurable and unverifiable quantities and subsequently using measurable and verifiable quantities such as dollars. That is exactly the task of market transactions, to design mechanisms to make services rendered and benefits derived easier to measure and verify and trade in.

Therefore, the falsifiable alternative hypothesis I am putting forward here is: *When quantities to a contract are difficult to measure and verify, people will fail to spot cheaters; when the same quantities are made verifiable and measurable, they will quickly develop new cheater detection mechanisms – right there in the lab – in social time, not evolutionary time.*

Furthermore, the more tightly tied the quantities and qualities to be measured are to our actual *lived* experience – i.e., our own past history and the collective cultural history we are raised in, the quicker we will develop cheater detection mechanisms. To give the most recent and raw of all such examples, if the young suicide bombers in the Middle East could somehow *verify* whether the fundamentalist cleric really delivers heaven and 63 virgins in exchange for their so-called martyrdom, we would find that they will very quickly develop a "cheater" subroutine that keeps them from making the trade in the first place. In sum, the development of this "mechanism" or ability need not be a consequence of millions of years of evolution, but a simple consequence of Lamarckian learning originating in our immersion in particular social groups.

Besides the inherent pitfalls of transferring arguments from biological adaptation and the massive historical evidence against it, there is another set of arguments why the adaptationist explanation for the results observed in the study is not very compelling. Recall the observation that Borges (1962) made in his essay "The fearful sphere of Pascal." He said: "It may be that universal history is the history of the intonations given a handful of metaphors." Ever since we have posited the existence of the brain, scholars have used metaphors to describe its workings and how it came to be. And with the supreme genius of the obvious, have (perhaps understandably) used the hottest and hippest of the tools of their time as their primary metaphors. The brain is a clock in the age of the clock; genes are messages and the brain the telegraph in the age of Darwin and Bell; and of course, the mind is but the hardware implementation of evolved subroutines in the age of the Turing machine and the double helix. In other words, it is one thing to consider an information processing system as a useful way to theorize about the mind; and quite another to posit actual subroutines that "explain" its functioning. To paraphrase the words

of Herbert Simon, the author of the metaphor of mind as information processing system: "The mind is an ink blot. It is whatever you see in it."

Furthermore, even if any such subroutines do exist, it still does not follow (as I argued above) that they were evolved specifically to solve any particular problem. Also, even if they *did* evolve to solve particular problems, there may still be a variety of other alternatives that impact their existence and function. In the most rigorous of adaptation theories, proof for the existence of particular mechanisms or subroutines may at best illustrate their evolutionary sufficiency – proving evolutionary necessity is another matter altogether. In fact, better scholars than I, including eminent geneticists and evolutionary scientists such as Gould and Lewontin have powerfully argued against the Cosmides and Tooby (1994: 328) claim that "… there is only one class of problems that evolution produces mechanisms for solving: *adaptive* problems." For example the mechanisms may be able to identify and solve new problems that had nothing to do with their evolution; or they may have no function whatsoever since the problem they evolved to solve is no longer relevant.

Evolutionary theorists (For example, Gould and Lewontin, 1979; Gould and Vrba, 1982) have identified several non-adaptive mechanisms including exaptations, serendipities, and redundancies. In the spirit of reasoning through metaphor, let us look at the history of technology as an analogue to the history of evolved "mechanisms" in the brain. The history of technological evolution too is filled with exaptations, serendipities, and redundancies. To give you just one example for each of these non-adaptive mechanisms: Viagra was developed while in search of anti-coagulants (exaptation); penicillin because someone neglected cleaning out a lab dish (serendipity); and the VCR embodies a collection of innumerable features that no one ever uses (redundancy). Similarly I could argue that any evolutionary history of neurological

subroutines is also most likely strewn with exaptations, serendipities, and redundancies. So where is the compelling case to believe that the cheater detection mechanism or any other subroutine has adaptively evolved to solve any particular problem in our evolutionary history? All that the results tell us is that for reasons we can only speculate about, we are very good at solving social interaction problems as opposed to other logically equivalent problems that do not make sense in terms of our life experiences and that we do not in some way care about.

At this point, I would like to move to the second phase of my critique where I examine the consequences of *accepting* the adaptationist argument for business ethics and our future research. If we accept the idea that evidence for the existence of any particular "mechanism" or "subroutine" or a "grammar of social interchange" is also evidence for its adaptive usefulness in evolutionary selection, we fall into the implicit Panglossian ethic of the "just-so-story." Such an ethic, and I insist that it is indeed an ethic, goes as follows: *If something has adaptively evolved into existence, it deserves to exist – it has in some way "earned" its existence in an evolutionary sense. Otherwise, by default, it would not have come to be.* Translated into modern economics, this ethic would imply that since Enron has made it to the top ten of the Fortune 500 companies, it deserves to exist and may even be emulated, instead of being summarily eliminated.

While this might appear as an unfair flippancy on my part in interpreting the arguments in the paper in favor of an adaptationist explanation, I would urge that the implication I derive is not that far-fetched. If our rational powers or subroutines for reasoning about particular domains are inevitably and exclusively created through adaptive evolution, and not through learned historical experience and ongoing meaningful exchanges of ideas and goods, what would be the basis for our ethics in general and business ethics in particular?

In a powerful, but non-technical thesis titled "Biology as ideology" R. C. Lewontin (1991) argues my case far more eloquently and in greater detail. And the concluding paragraph of his thesis lays out for us the inescapability of social action as the primary basis for the creation both of our individual identities as well as the spatial and temporal contexts within which those identities get forged and become meaningful: *History far transcends any narrow limitations that are claimed for either the power of genes or the power of the environment to circumscribe us. Like the House of Lords that destroyed its own power to limit the political development of Britain in the successive Reform Acts to which it assented, so the genes, in making possible the development of human consciousness, have surrendered their power both to determine the individual and its environment. They have been replaced by an entirely new level of causation, that of social interaction with its own laws and its own nature that can be understood and explored only through that unique form of experience, social action.*

To look to biology or other so-called "hard" sciences to tell us what to do in the social sphere is (to paraphrase McCloskey, 2000) to busy ourselves with games in the sand-box while time runs out on the human condition. As ethicists and economists, we can only so long shirk the hard chores that need to be tackled in our own domain by distracting ourselves with the cool toys fashioned by other disciplines. The physicist's metaphors ask us to equilibrate between forces outside our control; the biologist's to adapt to an environment not of our making; and the theologian's to answer to a divinity we cannot question. I suggest it is time we took ownership of the problems in our disciplines and allowed the problems to drive our search for solutions instead of fabricating pseudo-problems that we can then pound into alignment with the tools others have developed for other purposes. I posit social action – organized through conversation, community, and commerce – as a powerful alternative tool to the physicist's equilibrium, the

biologist's fitness, and the theologian's salvation for the conduct and future course of human affairs. And in the particular realm of business ethics, I hope we can embrace economics predominantly as a historical science that wields linguistic tools through which we can negotiate better societies into existence – both in the academy and the world outside it.

In the hope of such a possibility for the uses of our reason, whatever its origins may be, it might be worth re-reading the poets.

Possibilities

A week ago on longer clocks than ours
a supernova in Orion lit
the sky like a full moon. The dinosaurs
might have looked up and made a note of it
but didn't, and the next night it blinked out.
The next day from a metaphoric tree
my father's father's beetle brow and snout
poked through the leaves. Just yesterday at three
he spoke his first word. And an hour ago
invented God. And, in the last hour, Doubt.

I, because my only clock's too slow
for less than hope, hope he will not fall out
of time and space at least for one more week
of the long clock. Think, given time enough,
what language he might yet learn to speak
when the last hairs have withered from his scruff,
when his dark brows unknit and he looks out,
when the last ape has grunted from his throat.

-- John Ciardi

Bibliography

Bar-Hillel, M. "The base-rate fallacy in probability judgments." *Acta Psychologica*, 44, (1980): 211-233.

Borges, Jorge Luis. *Labyrinths: Selected Stories & Other Writings*, New Directions, (1962).

Cosmides, Leda, and Tooby, John. "The Cognitive Neuroscience of Social Reasoning." In *The New Cognitive Neurosciences* edited by Michael S. Gazzaniga, MIT Press (2000).

Cosmides, Leda, and Tooby, John. "Better than Rational: Evolutionary Psychology and the Invisible Hand." *The American Economic Review* 84, 2 (May 1994): 327-332.

Dunbar, Robin. *Grooming, Gossip and the Evolution of Language*. Faber Faber and Harvard University Press (1996).

Gigerenzer, Gerd, Hell, Wolfgang, and Blank, Hartmut. "Presentation and content: The use of base rates as a conyinuous variable." *Journal of Experimental Psychology: Human Perception and Performance*, 14, (1988): 513-525.

Gould, and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm: A critique of the adaptationist programme." *Proceedings of the Royal Society of London 205* (1979): 281-288

Gould, Stephen J. and Vrba, Elizabeth S. "Exaptation: A Missing Term in the Science of Form." *Paleobiology* 8, 1 (1982.): 4-15.

Lewontin, Richard C. *Biology as Ideology: The Doctrine of DNA*. New York: Harper Collins, (1991).

McCloskey Deirdre. How *to Be Human (Though an Economist)*. Ann Arbor: University of Michigan Press, (2000).

McCraw, Thomas K. *Creating Modern Capitalism*. Cambridge, MA: Harvard University Press (1997).

Simon, Herbert A. "The functional equivalence of problem solving skills." *Cognitive Psychology* Vol. 7, (1975): 268-272.

Tversky, Amos and Kahneman, Daniel. "Judgment and uncertainty: Heuristics and biases." In *Judgment and Uncertainty* edited by D. Kahneman, P. Slovic, and A Tversky, New York: Cambridge University Press. (1982): 3-20.