A Review of **Freakonomics**

John DiNardo*

December 10, 2005

Contents

1 What Can I Say About Freakonomics? 2

2 Popular Popularizations 3

3 What Does one Expect from a Hot Fudge Sundae? 4
- 3.1 What to expect and what not to expect 4
- 3.2 The Drive For Narrative Thrust versus an Accurate Rendition of The Facts 6
  - 3.2.1 Prostitutes and Architects 7
  - 3.2.2 Abortion Prohibition in Romania 8
  - 3.2.3 Levitt the Scholar and the "Levitt" of **Freakonomics** 9

4 What is an Interesting Question? 10
- 4.1 What is a "Cause"? 11
- 4.2 A Simple Framework 13
- 4.3 Clear and Unclear Causal Questions 17

5 "Interesting" Questions in Freakonomics? 18
- 5.1 Can Regression Help Distinguish "Cause" from "Consequence"? 20
- 5.2 Possibly Well Posed But Confusing and/or Ambitious Questions 23
- 5.3 Why A Transparent Research Design Helps 24
- 5.4 Type I and Type II Error 30

6 The "Hidden Side of Everything" or the Leper's Squint? 31

*I had the good fortune to to be able to call on a large number of friends and colleagues for advice, several who provided detailed comments on a previous draft. I would like to thank Bob Axelrod, Martha Bailey, Thomas Buchmueller, Elena Delbanco, Peter DiCola, Concetta DiNardo, Jane Dokko, David S. Lee, Jim Levinsohn, Darren Lubotsky, Jordan Matsudaira, David Morse, Justin McCrary, Gary Solon, Robert Valletta, Jean Wohlever, Dean Yang and the students of Nathan Anderson's microeconomics class at the University of Illinois at Chicago for advice, encouragement, and discouragement. Several others have helped in ways they might not recognize. Though many disagreed with what I had written or how I had written it, they were all kind enough to provide helpful suggestions and/or provide admonishments to be clearer which have surely helped. Thanks also to Robert Willis and Becky Bahlibi for help in tracking down an unpublished paper by Lee Lillard.

1 What Can I Say About Freakonomics?

In a review of a book intended for scholars, one might expect a careful examination of the authors' claims and the evidence supporting their claims. The premise that supports that expectation is that a serious book merits a serious review. When a book is intended for non–scholars, however, the question arises: what is it reasonable to expect? Surely one does not expect the same degree of precision from a popularization of an important concept in physics as one expects from a book intended for scholars? Indeed, it is fair to expect anything at all (by way of fidelity to the evidence or the original source material) in a popularization?

Such questions can not be evaded in a discussion of **Freakonomics: A Rogue Economist Explores the Hidden Side of Everything** – written jointly by the University of Chicago economist Steven Levitt and New York Times journalist and author Stephen Dubner ("Confessions of a Hero-Worshiper" and "Turbulent Souls: A Catholic Son's Return to his Jewish Family"). Indeed, much of the surprise I experienced at reading the text might have been avoided if I had come to the book with the expectation that it was intended less as popularization and more as entertainment. This of course is not equivalent to the claim that a popularization can not be entertaining nor the claim that entertainment can not inform. Nor should one infer that all or most of the claims in the book have been created "out of whole cloth": it is not difficult to find many claims in **Freakonomics** that are well supported. Rather, the reader is given little guidance as to when to expect that a claim in the book is well–founded, mere speculation or in some cases, contrary to fact.

As **Freakonomics** is not what I might have expected, the first half of the essay first includes

1. a brief sketch of what *I* have come to expect from popularization; other readers will have different expectations.

2. Next, I discuss the failure of **Freakonomics** to meet those expectations, through a small number of examples in matters large and small. Someone expecting an accurate (albeit accessible) explanation of findings from actual research, will sometimes be surprised.

Consequently, instead of devoting the second half of the essay to a careful evaluation of the myriad claims made by Dubner and Levitt, I instead turn to address the key premises that underly the book:

3. "Economics is a science with excellent tools for gaining answers but a serious shortage of interesting questions."

   As the authors make much of the distinction between "correlation and causation" it seems worthwhile to spend sometime clarifying what we mean by that phrase. To do so, I first lay down a simple framework where it is relatively easy to distinguish between the two even if it isn't always clear what we have learned from the exercise. I use this framework to address whether the some of the questions in **Freakonomics** have answers or how we might recognize if they did.

4. As the book makes much of clever algorithms to detect cheating, I give a simple analysis I would have like to see discussed in **Freakonomics** – when will a pool of persons identified as "potential cheaters" by a clever algorithm be composed mostly of innocents? Even good cheating algorithms can have surprising negative consequences that should be contemplated before they are employed.

5. I conclude with a short discussion of another important premise of the book that "incentives are the cornerstone of modern life."

2 Popular Popularizations

I had the good fortune to be raised by parents with not a lot of formal schooling, but a great deal of intelligence, curiosity and "old fashioned" working class values. Having grown up in a rural village in Italy during a period of time when education was hard to come by, my parents greatly valued education.

While serving as reluctant conscript amidst the chaos we now call World War II, my father became friends with a man from Sicily who could neither read nor write. Unable to write himself, my father would transcribe this man's letters home to his mother. (she too was also unable to read or write; someone on her end would read the letters to her.) At one point my father volunteered to teach his friend how to read and write. The lessons never got as far as the subtleties of punctuation or capitalization, but they had their intended effect. One day my father found his friend reduced to tears for the mere fact of being able to read one of his mother's letters. The moral of the story was clear: teaching was just a good thing to do (even if you didn't get to all the details!)

Perhaps partly in response to having grown up in such a difficult environment, my parents kept the house full of of popularizations of mathematics and science, many written specifically for children. In many cases I remember much more of what I learned from these popularizations in childhood than anything I later learned in school. These experiences and later ones have left me quite fond of a largely abandoned (unfortunately) tradition of engagement by left intellectuals in the enrichment of working class culture, most notably in the form of popular expositions of science and mathematics. Perhaps one of the the best known books from this tradition is Lancelot Hogben's *Mathematics for the Millions: How to Master the Magic of Numbers* which treats its subject very seriously although aiming for a broad readership (Hogben 1968). As one physicist described it, "Hogben was an English socialist who believed that science and mathematics are grounded in practical affairs and dignify themselves in the service of democracy. The history of science, [Hogben] wrote, is the history of the constructive achievements of mankind and the democratization of knowledge" (Raymo 1996).

Another well known socialist in this tradition, Albert Einstein,[1] described his aims in writing a popular book on relativity:

> In the interest of clearness, it appeared to me inevitable that I should repeat myself frequently, without paying the slightest attention to the elegance of the presentation. I adhered scrupulously to the precept of that brilliant theoretical physicist L. Boltzmann, according to whom matters of elegance ought to be left to the tailor and to the cobbler. [However], I make no pretense of having withheld from the reader difficulties which are inherent to the subject. On the other hand, I have purposely treated the empirical physical foundations of the theory in a "step–motherly" fashion, so that readers unfamiliar with physics may not feel like the wanderer who was unable to see the forest for trees. (Einstein 1920)

Dubner and Levitt share neither the ideology of Hogben or Einstein nor their aims. Indeed, **Freakonomics** announces that is not intended to be taken seriously with its title: the authors' intent in describing Levitt as a "rogue economist" is to describe one who is playfully mischievous.The

[1] Einstein had a long engagement with popular movements. In the U.S., for example, Einstein served as the co-chair of an anti–lynching committee with leftist activist, singer, and actor, Paul Robeson Jr. (Simon 2005) An essay by Einstein called "Why Socialism?" made its appearance as the lead article in the very first issue of the *Monthly Review* an "Independent Socialist Magazine" in 1949. Perhaps most famously, in 1918, on the day Kaiser Wilhelm abdicated his position as Emperor of the German Empire and King of Prussia, Einstein posted a sign on his door announcing "Class Canceled: Revolution."

aims of Hogben and Einstein are very different. For both Hogben and Einstein popularization is about a special type of *engagement* with non–specialists. Popularization is not "a necessary (albeit low–status) educational activity of simplifying" which proceeds from the view that "genuine scientific knowledge belong[s] to a realm that can not be accessed by the public, but is the exclusive preserve of scientists."(Hilgartner 1990) For Einstein and Hogben, knowing that the reader may place some confidence in the rendition of the material, the writer has a duty to act in the best interest of the reader – to make it as simple as possible, although not withholding from the reader any unavoidable difficulties in the material.

3 What Does one Expect from a Hot Fudge Sundae?

3.1 What to expect and what not to expect

Freakonomics: A Rogue Economist Explores the Hidden Side of Everything is certainly popular. Indeed, my search for something comparable took me back more than 120 years.[2] Even with the uncertainty about what constitutes a best seller, it is clear that the book has reached a huge audience, especially for a book about "economics." Although not surpassing the excellent *Harry Potter and the Half Blood Prince* by J.K. Rowling in sales[3], it has spent considerable time in various top ten lists. Perhaps as testament to the book's large audience, one can even buy Freakonomics T-shirts from a website run by Levitt's sister.[4] Levitt and Dubner have also made an admirable effort in reaching a broad audience: Levitt has been on "The 700 Club" (a talk show by conservative businessman and religious broadcaster Pat Robertson), "The Daily Show with Jon Stewart" (a center–left parody of the news and news reporting) among other places; both authors write a column for the *New York Times Sunday Magazine* as well as participate in an active blog (just navigate from the book's web site to the URL `http://www.freakonomics.com`, where, among other things, they respond to a large number of readers inquiries.) I think public outreach is admirable; as economists we are not theorizing about Martians and a public role in the discussion is, for me, quite welcome.

Its popularity notwithstanding, **Freakonomics** is cut from quite a different cloth than *Mathematics for the Millions* or even more recent popularizations of "academic" research such as Stephen

[2]Andrè–Michel Guerry's (1883) *Essay on the Moral Statistics of France*, is all I could find although I am sure there are more recent comparisons. Nonetheless, the similarities, differences, and parallels of this book with **Freakonomics** are interesting. "Guerry's work appears to be the first to test 'armchair' assumptions about the relationship of certain variables to criminal behavior."(Reid 1985). Moreover, "amateur's loved Guerry's books."(Hacking 1990) and Guerry's maps "created a brief academic sensation."... Although "France during Guerry's day was obsessed by crime and suicide ... there was little evidence that the crime rate was actually rising dramatically; indeed crime rates dropped during the period 1818–1830 ... and there was a considerable fear of crime and the emergence of an impoverished underclass *les misérables* that many regarded as almost a race apart." As to the book itself, "Guerry avoided accepting any grand theoretical system for explaining the causes of crime and other social problems."(Whitt 2002) An importance contrast of Guerry's work with **Freakonomics** is the former detailed description of the data. Often this entailed clever and sophisticated graphical displays which are by and large not provided in **Freakonomics**.

[3]On virtually the day I wrote these words came the following entry from a web-site for the book in an entry on August 10, 2005 entitled "NAKED SELF-PROMOTION:" A nice surprise from our neighbors to the North: as of noon today, Freakonomics is the #1 seller on Amazon Canada. This, of course, marks the first time anywhere that Freakonomics is outselling Harry Potter. And if the success of Freakonomics in North America is surprising, prepare to be shocked by readers who have brought it to the top of the lists in the U.K., Brazil, and even Singapore. Who knew?"

[4]The URL is `http://www.yarnzilla.com`, which advertises the shirt:"This 100% cotton limited-edition t-shirt not only fits flatteringly, it tells the world that you know the difference between John Maynard Keynes and John Cougar Mellencamp. Put sumo wrestlers, crack dealers, and realtors on notice that you will not be *played*. Pop a few in your cart (shipping is $5 no matter how many you buy) and let your freak flag fly!"

Pinker's *The Language Instinct*. Nor is the book the serious attempt at addressing pressing social issues such as one finds in Drèze and Sen (1989) – the book is resolutely about being "fun." Although the book does make much of the distinction between "correlation and causation" someone looking for a "helping hand" through some knottier problems of non–experimental inference will be disappointed. Indeed, the book contains assertions about "regression analysis", "correlation", and "causality" that will provide at least a few "cringe worthy" moments for anyone who has attempted to write about these subjects. The authors clearly intend to spare the reader "difficulties which are inherent to the subject". Moreover, this is quite in line with their stated aims:

> Will the ability to think such thoughts improve your life materially? Probably not. Perhaps you'll put up a sturdy gate around your swimming pool or push your real–estate agent to work a little harder. But the net effect is likely to be more subtle than that. You might become more skeptical of the conventional wisdom; you may begin looking for hints as to how things aren't quite what they seem; perhaps you will seek out some trove of data and sift through it, balancing your intelligence and your intuition to arrive at a glimmering new idea.

If the central goal of the book is promote skepticism about the conventional wisdom and a call to to put claims on a solid evidentiary basis, it is fair to say that the book does so, although sometimes despite itself.

As a matter of both style and substance, another way **Freakonomics** differs from popularizations in the tradition of Hogben and Einstein is that it is not meant to be exclusively about its motivating subject – the "economics of every day life" (as if there was any other kind!). As one might correctly assume from the title, the book seems intended as part hagiography of Steve Levitt (with an occasional guest appearance by Levitt collaborator Roland Fryer) and a celebration of Levitt's work and his approach to economics (although it includes discussion of the work of others as well).[5]

Each chapter begins with a vignette about Levitt the person written in the voice of Dubner. He is portrayed as a loving father and husband (albeit with a penchant for nerdish proclivities), courageous and compassionate in the face of great tragedy, as well as a self–effacing but brilliant[6] academic, a truth seeker above the fray of ideology, a "noetic butterfly,"[7] a "demigod, one of the most creative people in economics and maybe in all social science."[8] Although a bit at odds with the book's aims of promoting skepticism about the "conventional wisdom"[9] most of this material is harmless at worst, although on occasion it may disquiet some.[10]

[5] By way of counterpoint, *The Language Instinct*, which has a large amount of discussion of Noam Chomsky's work in linguistics, has no detail about Chomsky "the person."

[6] For what it is worth, I have always thought "brilliance" a bit overvalued in the human sciences. Excepting the adjective "German", my views are closer to Friedrich Wilhelm III, King of Prussia, who in a letter to his minister of trade wrote "[In statistical work] the main requirement is order, completeness and reliability. To achieve these ends, German diligence, laboriousness and perseverance are more to the point than brilliant talent, so long as they do not actually destroy the latter. (As cited in Hacking (1990).)

[7] The American Heritage Dictionary defines noetic as "Of, relating to, originating in, or apprehended by the intellect."

[8] To be sure, Levitt does not take the hagiography seriously. In an interview with the *Financial Times*, Levitt explained: "He created a totally fictional account of me, one that was far more likable and interesting and smarter than I was, that people kind of fell in love with. Dubner had set this tone, this fake version of me, that we both could slip into and out of as we wanted."(Harford 2005)

[9] Also militating against the goal of promoting skepticism of the conventional wisdom or social science (except unintentionally perhaps) is the text's various declarations that economics is a "science." Perhaps ironically, "regression analysis", much of the evidence behind the "science" discussed in the book, is alone among the toolkit that is explicitly downgraded to the status of "art." (Page 163).

[10] Some of this material, almost certainly apocryphal, may also strike some as unintentionally disquieting. In

3.2 The Drive For Narrative Thrust versus an Accurate Rendition of The Facts

For me, the principle reason I view **Freakonomics** as entertainment and not as popularization in the spirit of Hogben or Einstein is its penchant for including assertions of various degrees of validity and treating them as if they were all equally valid. Some of this appears to be a consequence of how the book was written: Combined with Levitt's story, the chapters are sometimes awkwardly stitched together discussions of Levitt's work, large chunks of which have already appeared in articles in the *New York Times*, spiced with plenty of anecdotal information. Dubner and Levitt often begins with an "unusual" question – what do Sumo Wrestlers and School Teachers have in common? Frequently, a chapter begins with an invitation to the reader – well–intentioned but occasionally feckless (s/he cheats at golf) – to enter a world where "bad guys" (sumo wrestlers, Chicago public school teachers, real–estate agents, the Klu Klux Klan, the criminologist James Alan Fox) are caught in the act of cheating by the intervention of a "powerful set of tools", usually, but not always, applied by Levitt.

As a consequence of trying to satisfy so many different goals – telling Levitt's personal story whilst trying to weave together a large body of often disparate economic scholarship with a chatty narrative – **Freakonomics** is also less inclined to take its subject matter seriously. The authors clearly intended to provide more of a light–hearted romp through matters given much attention in academic Economics.

Most telling is Dubner and Levitt's decision to "withhold from the reader most of the difficulties" inherent to their chosen subjects: this has probably helped the book's popularity. Unfortunately, their decision has some other less fortunate consequences as well. One gets the feeling that the book was stitched together rapidly. (In the introductory material, for example, the Levitt character says that he doesn't have the time to write a book.) Consequently, they make silly errors that one would have thought might be easily detected by a well paid editor. On page 68, for example, we read that

> "...information asymmetries everywhere have in fact been mortally wounded by the Internet"

only to read on the very next page that:

> "The Internet, powerful as it is, has hardly slain the beast that is information asymmetry."

Such a transparent error is of little moment on its own. Clearly the reader can decide for herself whether the "beast of information asymmetry" is dead, mortally wounded, or alive and well (and taken up residence in the Harper Collins Editor's office.) Unfortunately, the book also has a penchant for taking assertions from sources of varying degrees of credibility and treating them as if they were all equally credible. The reader unfortunately is rarely given any clue about when to expect that an assertion is likely to be well–substantiated or merely a useful story–telling device. This is the case both when a point is made in passing, or when it is central to a longer discussion.

---

one vignette, for example, "Levitt" encounters an apparently indigent man. As described in the voice of Dubner, bereft of any other recognizable human feeling, the Levitt character's intense but solitary interest in the man is the provenance of the headphones he wears.

3.2.1 Prostitutes and Architects

An example of the former type, after listing "four meaningful factors that determine a wage"[11] Dubner and Levitt casually remark that "the delicate balance between these factors helps explain why, for instance, the typical prostitute earns more than the typical architect."[12] This struck me as an intriguing throw away line if it were true, but is it? Dubner and Levitt don't provide a reference, which is unfortunate, and after a bit of research my conclusion is that there are probably *no* careful estimates of the wage of a representative group (probability sample) of sex workers that would substantiate such a claim. Indeed it would be a real project to generate a serious estimate. It appears that some prostitutes receive high payments for some transactions (if information from Internet web sites such as `www.punternet.com` are to be believed (Moffatt and Peters 2004)) but this type of information is of limited use in estimating what a a "typical" prostitute earns.[13] A quick check of one source for an architect's mean yearly salary[14] reveals an estimate of $66,230 in May 2004. I was able to find only one estimate from an actual probability sample that describes the income of of what might be called a "typical prostitute."[15] Adjusting this 1989 estimate for inflation, an estimate for mean income for "Street Prostitutes" in Los Angeles is $36,325.[16] With a great deal of trepidation given the quality of the data and the likely existence of "compensating wage differentials" I would probably counsel the typical architect to keep his/her day job.

The most confusing twists and turns often involve discussions about specific research conducted by others. Compared to a work such as Stephen Pinker's *The Language Instinct* (whose clear but simplified expositions seem to bear a close relationship to the source work, even when that source

[11] "When there are a lot of people willing and able to do a job, that job doesn't generally pay well. ...the others are the specialized skills a job requires, the unpleasantness of a job, and the demand for services that the job fulfills."

[12] The four meaningful factors sound like boilerplate from a neo–classical model of wage determination. Labor economists working in the neo-classical tradition have generated many more than four meaningful factors. To take only one trivial example, in the U.S. the modal (most common) wage is often exactly or very near the minimum wage. See DiNardo, Fortin and Lemieux (1996), for example. In 1979, there is a huge spike in the density of wages near $2.90; in 1992, there is a spike at $4.25 instead. Perhaps this reflects the "delicate balance" of meaningful forces, or perhaps this reflected the fact that the value of the legal minimum wage in fact was $2.90 in 1979 and $4.25 in 1992.

[13] Edlund and Korn (2002) observe that even defining prostitution is quite difficult: "a prostitute cannot simply be a woman who sells her body since 'that is done by women who become wives in order to gain a home and a livelihood.'" See Edlund and Korn (2002) for references: they compile a number of anecdotal and more serious estimates of prostitute earnings, even though most of these estimates are from non–probability samples. Moffatt and Peters (2004) remark that "it is surprising that so little empirical work has been carried out on this 'oldest' profession" and cite no probability samples. Edlund and Korn (2002) cite one probability sample of U.S. prostitutes which is the unfinished Lillard (1998). This is the basis for the estimates I calculate.

[14] From the Occupational Employment Statistics, May 2004: Architects, Except Landscape and Naval. The annual earnings number is derived by assuming 2080 hours of work per year and multiplying the mean wage as calculated from the survey.

[15] For some idea of the difficulties involved see Kanouse, Berry, Duan, Lever, Carson, Perlman and Levitan (1999), and for evidence that convenience (non probability) samples may not be enough, Berry, Duan and Kanouse (1996).

[16] This estimate includes income from (a usually small amount of) non–sex related work. I have used the CPI-U of 124.0 for 1989 and 188.9 for 2004 and the estimate for mean of "Total Income, 1989" – 23,844.7 – from Lillard (1998) Table II.8. One limitation of this study for this exercise is that his probability sample of 901 prostitutes is designed to yield information on the "typical" sex worker working on the street in Los Angeles, this may or may not close to the "typical" prostitute. Moreover, even such issues as computing an average hourly wage are complicated: There is substantial heterogeneity in both payment and hours worked. Given the problem defining "prostitute," and the nature of the work and payment, calculating the wage of the typical prostitute would take a major research effort that to my knowledge has not been undertaken. Another surprising finding is that this wage is not much higher wages in the service sectors workers or for women more generally according to Lillard (1998). According to their data, prostitutes earn about 15 percent more than working women in Los Angeles generally, and about 28 percent more than service workers. In their sample, 69 percent of sex workers are white, and about 8 percent are hispanic. 33 percent have less than a high school degree, and 3.5 percent are college graduates.

work is rather difficult[17]), in **Freakonomics**, by contrast, it is sometimes difficult to recognize the cited research.

3.2.2 Abortion Prohibition in Romania

The inquisitive reader will find more surprises regards the long discussion in the chapter entitled "Where have all the criminals gone?" where the authors relate the story of Romanian dictator Nicolae Ceausescu's decision to declare abortion illegal in 1966 which frames much of the discussion. The narrative seems to suggest that this decision led to Ceausescu's eventual execution:

> It should not be overlooked that his demise was precipitated in large measure by the youth of Romania – a great number of whom, were it not for his abortion ban, would have never been born at all.

The text goes on to discuss the Romanian abortion ban referring to both popular articles as well as more scholarly publications. One surprising rendition of the originals includes a pair of papers by Cristian Pop-Eleches (Pop-Eleches 2005b, Pop-Eleches 2002), which is summarized in **Freakonomics** this way on page 118:

> Ceausescu's incentives produced the desired effect. Within one year of the abortion ban, the Romanian birth rate had doubled. These babies were born into a country where, unless you belonged to the Ceausescu clan or the Communist elite, life was miserable. But these children would turn out to have particularly miserable lives. Compared to Romanian children born just a year earlier, the cohort of children born after the abortion ban would do worse in every measurable way: they would test lower in school, they would have less success in the labor market, and they would also prove much more likely to become criminals.

The curious reader who tracked down the relevant papers by Pop–Eleches would be very surprised to learn that the description in **Freakonomics** is virtually the *opposite* of what is actually claimed.[18]

> On average, children born in 1967 just after abortions became illegal display better educational and labor market achievements than children born just prior to the change. This outcome can be explained by a change in the composition of women having children: urban, educated women were more likely to have abortions prior to the policy change, so a higher proportion of children were born into urban, educated households. (Pop–Eleches (2002), page 34)

While Pop–Eleches relates suggestive evidence that *conditional* on the usual list of demographic characteristics, a fetus born after than ban is more likely to engage in criminal behavior, Pop-Eleches' conclusion is that the effect is second order:

[17] I know nothing about linguistics, but having tried to plow through Miller and Chomsky (1963) I clearly recognized some of the themes of the original in Pinker's discussion of it on pages 93 forward in Pinker (1994). Reviews of the book by linguists and discussion with colleagues in linguistics confirm that impression.

[18] I am citing the most recent versions of these papers which, of course, could not have been used by Dubner and Levitt. The earlier versions did not vary appreciably except in details extraneous to this discussion. (Pop-Eleches 2005a)

> These results suggest that overall children born immediately after the ban of legal abortions have better educational outcomes than those born immediately prior the ban, implying that the positive effect due to changes in the composition of mothers having children more than outweighs all the other negative effects that such a restriction might have had. (Pop–Eleches (2002), page 20, 21)

The reader is given no hint that Dubner and Levitt's summary of Pop-Eleches' work so badly misrepresents its substance. It is unclear why/if they chose to do so. For me, this type of misrepresentation is especially unwelcome (and unnecessary) as Dubner and Levitt use the Romanian case as a "framework" on which to hang much of the book's discussion of Levitt's far more controversial claims about the impact of abortion legalization in the U.S.[19]

3.2.3 Levitt the Scholar and the "Levitt" of Freakonomics

Even Levitt's *own* research is discussed in a way that might surprise those who have read the originals. On page 126, Dubner and Levitt review Levitt (1997) which attempts to use political electoral cycles to identify a causal effect of police on crime. After a brief but accurate description of the research design they describe the results saying "it's possible to tease out the effect of the extra police [induced by electoral cycles] on crime."

Again, a surprise is in store for the reader of that passage.

Levitt (1997) estimates of the effect of police on specific crime categories using electoral cycles as an instrument. The original work makes at least two claims which relate to that passage from **Freakonomics**:

1. The estimates of the effect of police on crime using electoral cycles as instrumental variables in Levitt (1997) are "generally not statistically significant for individual crime categories."

2. These estimates although generally insignificant for individual crime categories "are significant for violent crime taken as a whole."

If that had been the end of the story, it might be fair to conclude from the research that "it's possible to tease out the effect of the extra police [induced by electoral cycles] on crime." However, Levitt (1997) began a story that Levitt (2002) concluded. The duly cited Levitt (2002), in fact, is a reply to the replication study McCrary (2002), neither mentioned or cited in **Freakonomics**. Unfortunately for the narrative, McCrary (2002) demonstrates that the second claim is based on on a programming error as Levitt (2002) concedes. As one reader described **Freakonomics** as an ice cream sundae, it might be said this uncomfortable fact about the actual research does not make its way into the chocolate sauce of the **Freakonomics** sundae.

Indeed, the summary by (McCrary 2002) is much more to the point: "While municipal police force size does appear to vary over state and local electoral cycles . . . elections do not induce enough variation in police hiring to generate informative estimates of the effect of police on crime." Levitt (2002) goes on to use a very different research design to investigate the question, but that is of no moment for the passage in **Freakonomics**.[20]

[19] N.B. I don't mean to suggest that Pop-Eleches' actual findings necessarily contradict any of the claims made elsewhere by Levitt about the U.S. case. Indeed, it not too difficult to tell either a story in which Pop-Eleches' actual findings are broadly consistent about Levitt's own findings on related issues or largely silent about the issues Levitt addresses. (The cited papers by Pop–Eleches, in fact, are not even primarily about a putative abortion-crime hypothesis in part because of scarcity of good crime data from Romania.)

[20] I do not mean to suggest that it is some sort of crime to commit a programming error. Mistakes are to be expected even from the most diligent researchers, and this is one reason scholarly journals make room for replication

My point here is not to debate the substantive questions. More police may or may not reduce crime. Electoral cycles may or may not allow a research to "tease out" an effect of police on crime (although the evidence suggests that they do not.) There may be other credible research designs that support this conclusion, although I am not aware of any.

Rather, the gap between the depiction of academic research in **Freakonomics** and the research it purports to describe is often much larger than I have come to expect from popularizations. Moreover, as this last example makes clear there is even a gap between "Levitt the scholar" and "Levitt" in **Freakonomics**: The latter made no error and was able to tease out an effect of police on crime. "Levitt the scholar" on the other hand, was conscientious in allowing another scholar to show that it was not possible to tease out an effect of police on crime with that research design. The general impression one receives is that Levitt did not carefully read much of **Freakonomics**: Levitt the scholar would not have made such an error.

I do not mean to suggest that it would have been easy to make the material accessible *and* more faithful to the source. Certainly, even the simplest mathematics is never an easy sell. For example, Varian (2002) spent a column in the *New York Times* discussing the putative Nash Equilibrium in a scene from Ron Howard's popular movie *A Beautiful Mind* about the life of the mathematician John Nash (based on the book by Sylvia Nasar)[21]

I *do* mean to suggest that **Freakonomics** is not a popularization in the tradition of Hogben and Einstein. Even where it was easy to provide guidance to the curious reader there are omissions that will surprise.[22]

If the many reviews of the book are any guide, most find the book "entertaining" even if "Levitt's only real message is to encourage confrontational questions"(Berg 2005) Indeed, one reviewer went so far as to suggest that "criticizing 'Freakonomics' would be like criticizing a hot fudge sundae"(Landsburg 2005). *De gustibus non est disputandum*: instead of providing a detailed critique of the assertions in **Freakonomics**, I will first use the book as largely as a springboard to discuss some of the books themes in a slightly broader context.

4 What is an Interesting Question?

Several different themes make an appearance in **Freakonomics**. One that seems fruitful to discuss is the assertion (page ix) that "economics is a science with excellent tools for gaining answers but a serious shortage of interesting question." I must confess, if I were to compose a tagline about economics it might be quite different: for me, there are an infinity of interesting questions; the

studies. Indeed, it was Levitt who graciously provided McCrary with the original programs and data that made it possible to demonstrate conclusively that a key claim of Levitt (1997) – that the estimates using this research design "are significant for violent crime taken as a whole" – was not in fact correct.

[21]In the movie, John Nash, one single guy among many is trying to pick up at least one girl at a local bar. After presumably working through the complex analytics of some unspecified game, he mysteriously concludes that the optimal strategy for each of the men is "don't go for the prettiest girl." The women's role in this game is left unspecified.

The problem of remaining minimally faithful to the original is often no easier when dealing with fiction. Ron Howard, who is also directing the fictional "Da Vinci Code", apparently sought advice on how to appeal to a broad audience who might be offended by the book's central premise – Jesus and Mary Magdalene had a child who was meant to be Jesus' true heir. According to a newspaper account, one piece of advice that he was given was to change the premise!(Waxman 2005)

[22]One example should suffice. Given the book's long discussion of the putative causal effect of abortion legalization on crime, it is bewildering why in the notes to pages 136-144 – which enumerate some recent work on the link between abortion and crime – the authors list Levitt and Donahue's "Further Evidence that Legalized Abortion Lowered Crime: A response to Joyce" (2004) without mentioning Joyce (2004a)!

problem is our tools are rather meager for making much headway with most of them, certainly as compared to the sciences such as physics.

In any case, there are different criteria one might pose for an interesting question. In social science research more generally, it seems to me that one relevant criterion is "answerability" – another might be "credibility." There are other criteria to be sure.

Although not all interesting questions are "causal", a lot of social science research purports to answer such questions. Questions can range from "ill–posed and unanswerable as stated" to "barely well–posed and difficult to learn about credibly" to "well–posed and straightforward to learn about credibly." (One curious phenomenon I have observed is that interest among social scientists is often highest in questions that strike me as ill–posed or impossible to answer, and lowest for questions which are arguably well–posed and answerable.) For me, the confusion is often the greatest in papers where there is no explicit discussion of an actual or even hypothetical policy. Often I find myself simply unable to *understand* either the question be asked or how I might evaluate the credibility of the answer given.

Given such a large selection from which to choose, it is interesting that **Freakonomics** often focuses on that part of Levitt's work where the questions are the least well–posed and the least amount of time on that part of Levitt's work which poses answerable questions with credible research designs. Levitt is not alone in posing such questions to be sure, and not all the questions the book takes up are ill posed. Nonetheless, a discussion seems warranted given the attention **Freakonomics** gives to causation: indeed, the book pokes fun at several persons who in their view fail to appreciate the "distinction between correlation and causation."[23]

In order to explain my premise that much of **Freakonomics** poses unanswerable questions at worst, or unclear questions at best, it will be necessary to lay down a simplistic framework in which what constitutes a clear question and a credible answer is relatively straightforward. I do so with far too much brevity:

1. First I explain what is meant by a "cause"

2. Second, I explain a *single case* where we *sometimes* have *some* hope for evaluating cause, the randomized controlled trial (RCT). My point is not to argue that this is the best or only way – surely our understanding of the world would be even more empty if it were based solely on this type of evidence. Rather, it is a framework in which it is easy to see what makes for a meaningful (albeit limited) question about causation and when we might have reason to believe that the results are valid.

3. I end with a simple example of a question that seems well posed but isn't.

In the subsequent section, I apply some of the lessons about *posing* meaningful questions from this framework to examples from Freakonomics.

4.1 What is a "Cause"?

It is not possible to provide a definition of "cause" in the social sciences that would perfectly discriminate cause from "correlation" in all contexts. At best, a social science (or even medical) "cause" that we will talk about is a faint echo of the notion of causality as is commonly used

[23] The contrast with Guerry (1883) is noteworthy for its modesty in this regard. "We have duly avoided any speculative consideration of causes and causal chains so as not to stray from the object of statistics, . . . does not directly show how they are linked. The study of causes is slow, difficult, and fraught with error."

in the hard sciences.[24] In part, this is because few concepts used by social scientists admit of much refinement in the way that, say, the notion of "mass" does in physics: however slippery the concept of "natural kind" is (Hacking 1991), it seems clear that myriad behaviors that fall under the rubric "crime" are *not* a natural kind in the same way that "mass" is to physicists. They don't obviously possess some common set essential properties such that it is obviously meaningful to study and describe all of them with a single term.[25] For example, war criminals, prostitutes and "johns" all commit "crime" but it is hard to believe that their might be laws that explain an essence common to these diverse activities (as well as other crimes such as strike–breaking, earnings misstatements, murder etc.)

Complicating matters further is the fact that the word cause has many (often contradictory) meanings. Aristotle, for example, wrote of four types of causes – material, formal, efficient, and final – none of which maps very well to the way the term is used by social scientists or in medicine. To take an example, what does it mean to say that Mrs. O'Leary's cow *caused* the Great Chicago Fire of 1871? Even if we were to agree (and perhaps we shouldn't (Bales 2002)) with this version of events:

> One dark night, when people were in bed,
> Mrs. O' Leary lit a lantern in her shed,
> The cow kicked it over, winked its eye, and said,
> There'll be a hot time in the old town tonight.

as to the "ultimate" cause of the fire, we might say the cause of the fire was Mrs. O'Leary's cow. We could also say that Mrs. O'Leary (and not her cow) was the cause of the fire since her placing of the lantern in the barn had the predictable consequence of igniting a blaze that would engulf much of Chicago. More policy relevant perhaps, we could cite lax fire regulations as the cause: perhaps Mrs. O'Leary would have been more cautious had the placing of a lantern in one's barn had been illegal. In today's language we might have talked about the failure to impose penalties that result in effective deterrence. More fancifully, we might even trace the cause back to U.S. agriculture subsidies. Without the government subsidies, maybe Mr. and Mrs. O'Leary would have not decided to take up dairy farming at all!

Thanks to Voltaire, perhaps the best known type of reasoning about ultimate "causes" is the famous Dr. Pangloss of *Candide*.[26] At one point Candide is reunited with his former teacher, Dr. Pangloss who has been reduced to a beggar, with his nose half-eaten off, covered in scabs. Surprised by this (and a lot of other) misfortune Candide "inquired into the cause and effect, as well as into the sufficing reason that had reduced Pangloss to so miserable a condition." As it turns out, Dr. Pangloss had "tasted the pleasures of Paradise" with Pacquette, a pretty servant girl, who had, as it turns out, been infected with a disease, the impressive genealogy of which Dr. Pangloss is able to trace back to a Countess, a Jesuit, a novitiate (among others) and ultimately

[24] Hacking (1995) observes that "causal generalizations lie between extremes. At the one end is the *strictly universal*: whenever there is an event or condition of kind K, then there results an event or condition of kind J. Old–fashioned physics preferred laws like that. At the other end are truly modest statements of *fairly necessary* conditions: Without events or conditions of kind K, events or conditions of kind J are unlikely to occur. In between we have probabilities and tendencies."

[25] See Nelson (1990) for a discussion of how even more basic economic concepts may not meet such a requirement.

[26] Voltaire describes Pangloss this way: "[He] was professor of metaphysico–theologo–comsolo–nigology. He could prove, to admiration, that there is no effect without a cause; and, that in this the best of all possible worlds, the baron's castle was the most magnificent of all castles, any lady the best of all possible baronesses. It is demonstrable, said he, that things cannot be otherwise than as they are: for all things having been created for some end, they must be necessarily be created for the best end. Observe, that the nose is formed for spectacles, and therefore we wear spectacles. The legs are visibly designed for stockings, and therefore we come to wear stockings." Chapter 1, (Voltaire 1796)

Christopher Columbus. Candide asks *why* did Dr. Pangloss suffer such a horrific fate? What *caused* his degradation? For Dr. Pangloss, causal questions were straightforward: things could not be otherwise than they are, all things are created for *some* end, and thus all things are created for the best. In this case, Dr. Pangloss concludes his suffering was "a thing unavoidable, a necessary ingredient in the best of worlds" for had this disease not come to pass "we should have had neither chocolate nor cochineal."[27]

Economists sometimes seem to flirt with a focus on "ultimate" causes: the quest to explain some phenomenon is considered complete when a behavior can be interpreted as the equilibrium outcome for some individualistic agents optimally maximizing utility. Indeed, a quick search of the web finds the phrase "provide an *economic* explanation for phenomenon 'X'" in many introductory economics exams. Sometimes this is useful. Sometimes this is not: it might be possible to describe a shy teenage boys presentation of some gushing romantic doggerel to the amour of his dreams as a method of making a credible commitment to a future of joint household production (she after all now has the means to embarrass him) but I am not sure such an explanation helps *me* understand *why* the boy acted as he did.

Dubner and Levitt seem to flirt with teleological reasoning that sometimes evokes Dr. Pangloss search for ultimate causes. For example, they liken Norma McCorvey's decision to pursue what became Roe v. Wade and its subsequent effect on crime to the "proverbial butterfly that flaps its wings on one continent and eventually causes a hurricane on another." Nonetheless, the search for ultimate causes *is not* what we generally have in mind when the word "cause" is used in social science. Instead, one usually has in mind an action, manipulation, or intervention that one is interested in and the possibility that their exists some "stable" relationship between the intervention and the consequences or outcome of the intervention. One useful expression of this notion is from Heckman (2005):

> Two ingredients are central to any definition [of causality]: (a) a set of possible outcomes (counterfactuals) generated by a function of a set of "factors" or "determinants" and (b) a manipulation where one (or more) of the "factors" or "determinants" is changed. An effect is realized as a change in the argument of a stable function that produces the same change in the outcome for a class of interventions that change the "factors" by the same amount. The outcomes are compared at different levels of the factors or generating variables. Holding all factors save one at a constant level, the change in the outcome associated with manipulation of the varied factor is called a causal effect of the manipulated factor.

What question is being answered and the credibility of the answer so obtained is clearest in the randomized controlled trial. In the next section, I highlight some of the obvious features of the RCT. The goal is not elucidation of an air–tight framework for inference or abduction or a claim that the RCT is an ideal, but rather a way to discuss the types of questions social scientists ask and what we might expect to learn.

4.2 A Simple Framework

One could easily fill a small library with all that has been written on causality. [28]

[27] See Chapter 4 of Voltaire (1796). The translator of this version of Voltaire's story attributes this style of reasoning to the "maxims of Leibniz" and as put into the mouth of Dr. Pangloss is a "most Capital and pointed stroke of Satire." Cochineal is apparently a red dye made from ground up insects.

[28] The extensive references in Shadish, Cook and Campbell (2002) would be a good start on such a library. The book itself contains a useful, albeit idiosyncratic discussion of some of these issues in causality by empirically

Instead, I will work with the simplest framework, and briefly discuss *a single case* where we *sometimes* have *some* hope of evaluating whether something "causes" another thing in the sense usually meant in social science. In the interests of brevity I have combined two related (but different) aspects of the problem: the "credibility of the research design" and the "well–posedness" of the question, even though they are somewhat distinct, and ignored several other important questions such as how one takes evidence from such simple setups and extrapolates to actual or contemplated policies.

The randomized controlled trial (RCT) is the simplest such framework in which to discuss causality. It is perhaps ironic that in medicine, the RCT has become known as "the gold standard", the technique owes much of its development to research on telepathy (mind reading) and is arguably best suited to situations "marked chiefly by situations of complete ignorance" (Hacking 1988).

In an RCT, a single potential cause is randomly "assigned" to a treatment group and a (inert) placebo is assigned to the control group.

Let y_i be an outcome which can be measured for all individuals, and let $T_i = 1$ signify that person i has been assigned to treatment and $T_i = 0$ otherwise. Suppose the following characterizes the true state of the world[29]:

$$y_i = \alpha + \beta T_i + f(X_i) + \epsilon_i \tag{1}$$

where α and β are constants, $f(\cdot)$ is some unknown function of all the observable characteristics that affect y_i before being assigned to the treatment or control, and ϵ_i is all the other unmeasurable influences.[30] A fundamental problem we face is that for some individual i we can only observe the person in one of the two states – treatment or control. Another related problem is that we don't observe everything that affects the outcome y. For any individual, then we can never be certain that some unobserved determinant of the outcome y is changing at the same time we are assigning the person to treatment or control.

The key to this design is that by coin toss or some other contrivance that generates "random numbers" persons are next assigned to either treatment or control in a way that is independent of their characteristics. If this assignment is conducted on a random sample of individuals from some population, then the mean outcome for individuals in the treatment group – $\overline{y}_{T=1}$ – is a good estimate of the average outcome of individuals from this population under the treatment – $\alpha + \beta + E[f(X_i)]$. By similar logic, $\overline{y}_{T=0}$, a good estimate of the average outcome for the control group – $\alpha + E[f(X_i)]$. The difference between these two means is likewise a good estimate of the average treatment effect for this group.[31]

---

minded social scientists, (although unfortunately for me unleavened with algebra or simple mathematics.) For a similar focus on treatment effects see Holland (1986). For more recent work see the very useful discussion in Heckman (2005). See also Freedman (2005) which includes an interesting discussion (among other things) of Yule's (1899) famous footnote; after pages and pages of correlations and discussions about them Yule disavows that the correlation he identified between pauperism and a specific type of provision providing food–relief was causal with a single footnote: "Strictly speaking, for 'due to' read 'associated with.'"

[29] Another way to proceed which is often helpful is to establish a notation for a counterfactuals. Let $Y_i(1)$ be the outcome when the person is assigned to the treatment and let $Y_i(0)$ be that same person's outcome when they are assigned to the control. The treatment effect for person i is then $\tau_i \equiv Y_i(1) - Y_i(0)$. It is generally impossible to observe τ_i since the individual is one state or the other. We could then talk about trying to define some average (for some population) of $E[Y_i(1) - Y_i(0)]$ as an object of interested. See Holland (1986) for an exposition along these lines. See Heckman (2005) for a critique of that approach and related points.

[30] We have already simplified the usual situation economists confront considerably, for example by treating β and α as constants. Quite reasonably, they might be expected to vary across individuals: in that case, the best one will generally be able to do is compute some sort of average effect.

[31] We have swept several issues under the rug that can even arise in a simple medical example. For instance, we are assuming that "general equilibrium" effects are unimportant so that one isn't concerned that the controls

The assertion that the estimate so formed is a "good" one is fortunately not one on that has to be taken solely on faith. While not "assumption free", our confidence in estimates generated this way *does not* rely on us having complete knowledge of the data generation process given by equation (1) although more knowledge helps! In a typical RCT, in fact, any of the variables in X_i are generally not used for any purpose but to test the design. Under random assignment, *any* X_i should be the same on average for the two groups. This is, of course, a consequence of random assignment that is routinely tested in every RCT. If the groups look very different on average, this is generally considered evidence against the design, and one reason to have less confidence in the results. It is the fact that the X are the same on average that gives us some reason to believe that the same is true for ϵ. Even in this simple case, we can never be sure that this is true. At best, the answers from identical experiments have the "tendency" to be correct.

Two attractive features of a well designed RCT which are usually too obvious to deserve mention become more important when one turns to the sorts of "approximations" we are often faced with in social science:

1. *Pre specified research design.* In an RCT, the researcher specifies *in advance* to the extent possible what conditions have to be satisfied, and what will be concluded (with the usual degree of tentativeness associated with any technique involving sampling) under every possible result of the experiment. If we are assessing the efficacy of a drug, for instance, it is pointless to decide in advance that the drug "works" and then massage the data, sample, specification, etc. until we "reach" that conclusion. Doing so would seem to vitiate using the RCT as a method for anything but confirming our previously held beliefs.[32] Indeed, historically and etymologically the notion of an "experiment" is intimately related to the effort to put one's views to the test (DiNardo 2006).

2. *"Transparent" research design.* In the classical RCT, for example, it is transparent what constitutes evidence against the design (for example, if the predetermined characteristics of the treatment and control and very different) and what comparison or regression coefficient constitutes evidence in favor or against the claim. In my own experience, when a research design fails to be transparent in this way, or the research fails to provide the relevant numbers that speak to the validity of the design or the conclusions, I generally conclude that the evidence does not support the researcher's claims. While such a harsh inference about the research may not always be correct, I have "sampled" enough research to suggest that it works remarkably well as an inference tool.

Another set of assumptions – again these are usually too obvious to be discussed in the case of the RCT – deal with whether a question or set of questions are "well posed" or "meaningful."

3. We can identify a "treatment" or "policy." At one level, since we are dealing with human beings one often has to carefully distinguish between "assignment to treatment" and the "treatment." You can assign someone to take a specific medicine but it isn't always reasonable

are affected by the treatment also. These and related concerns become even more important when we raise our ambitions to seek to extrapolate the results of the experiment to other possibly different contexts. There is a long tradition in Economics of seeking answers to these more difficult questions that dates back at least to the Cowles Commission (see Heckman (2000). For reasons of brevity, I focus on "simpler" less ambitious questions. (Heckman and Vytlacil 2005)

[32] For an illustration of evolving definitions of the "appropriate" specification *after* having seen the results, and the consequences of failing to adopt a pre–specified research design, see the discussion of Welch (1974), Siskind (1977), Welch (1976) and Welch (1977) in chapter 6 of Card and Krueger (1995). Although the extent of this research style is unknown, I suspect that the example is unusual only because it is documented.

to assume that the person has taken the medicine. Even if we can ignore such distinctions it may be difficult to identify what our treatment *is*. Even the most routine, small medical manipulation often comes bundled with other things. Many years ago it would have been a sound inference based on lots of unfortunate experience that the causal effect of a spinal tap (lumbar puncture) would be a serious headache afterward. Is this effect caused by the substance used to sterilize the needle? The type of needle? The size of the needle? Despite the fact that lumbar punctures have been performed for more than 100 years (Sakula 1991), these questions are a subject of a continuing debate despite *many* randomized controlled trials (Armon and Evans 2005).

4. The effect of a treatment is always *relative* to the control. The state of being assigned to the control is the "counterfactual" against which the treatment is evaluated. An effect is a comparison of outcomes in different possible states.

5. The treatment involves an "intervention" and/or is "manipulable". In the RCT, this is so basic it hardly deserves mention; it is, however, a subject of some debate among economists.[33] As I use the word "cause", it is not meaningful to ask what is the effect of "being black" on one's propensity for crime. Only in a fantasy world does it make sense to consider the fate of John DiNardo as a "black man." If a misguided social scientist had been able to secretly reach back into the womb to manipulate John DiNardo's DNA to make him "black" (something that would have no doubt come as a surprise to his Italian parents) would it even be meaningful to describe the person generated from that process as the "black John DiNardo?" to which the "white John DiNardo" could be compared? The issue is not "is such a manipulation possible" but "were such a manipulation conceivable, would it answer the question we are asking." If the answer to that question is "no", I would describe the question as ill–posed even if it is the answer to a different well-posed question. Some of debate the on this is perhaps merely a question of terminology. As I discuss later, I think it is possible to talk about the effect of changing a person's *perception* of the race of, say, a job applicant because it is perhaps meaningful to think about manipulating a person's perception of race.[34]

6. No matter how the treatment is assigned it always has the the same effect (β) on the

[33] See Granger (1986) for example.

[34] Moffitt (2005) for example, explains that

> "[The argument in Holland (1986) that race can not be a cause because it can not be manipulated results from] . . . a mistaken application of the experimental analogy, and the more basic counterfactual analogy is the superior and more general one. It does make conceptual sense to imagine that, at any point in the lifetime of (say) an African-American, having experienced everything she has experienced up to that time, her skin color were changed to white (this is sometimes called a gedanken, or thought, experiment). Although it is a well-defined question, it may nevertheless be unanswerable, and it may not even be the main question of interest. For example, would the individual in question move to a different neighborhood, live in a different family, and go to a different school? If not, the question is not very interesting."

While a distinction between comparisons one could make and those that are possible is important (I wish to think of manipulable quite broadly), I find such discussion confusing. If I were to wake up tomorrow and discover that I was "black" one possible reaction might be a visit to the Centers for Disease Control to learn if I had acquired an obscure disease!! Whether or not I moved to a different neighborhood, divorced my wife, etc., if that response were typical of other white folks who woke up one day to find themselves "black", I would nonetheless hesitate to say that the "causal effect of being black" (or white) is an increase in the probability that one makes a visit to the CDC, though it could be so described. Again, absent some discussion of a class of hypothetical manipulations *and* counterfactual states, for me it is hard to know what to make of such causes.

outcome.[35] For example, if the effect of aspirin on headache differs when it is given to a patient by a nurse than when it is given to a patient by a doctor, the most we can do is describe the causal effect of "nurse administered aspirin" or "doctor administered aspirin." In the limit, of course, if only the method of administration matters we might even wish to conclude that aspirin *qua* aspirin doesn't cause anything to do with headache.

7. I would add, although this is not properly thought of as a "requirement", that for me, the most interesting studies involve manipulations that correspond to real policies. In these cases, even if we learn little about the "structure" of a true model, we have perhaps learned a little about the consequences of one possible action we have taken.

In writing down this very abbreviated framework for inference, I do not mean to suggest by the foregoing that "best" evidence is always an RCT (see Heckman and Smith (1995) for one thoughtful discussion of the limitations of such evidence in social contexts) or that all meaningful questions satisfy the above desiderata, or that the only evidence that we should use to reflect about ourselves should come from RCTs[36] or approximations to RCTs.[37] Quite to the contrary, I don't even think that a singular focus on "well–posed" questions would be a good idea.[38]

I would even go further and suggest that in many areas under study by economists, the focus on "treatments" can be, perhaps unintentionally, narrow. As Thacher (2001) observes "Reducing crime is clearly one important goal for the police. But it must compete with other goals like equity, due process, just deserts, and parsimony." Rather, my argument is that if a putatively causal question can not be posed as some sort of "approximation" to a question satisfying the above desiderata, the burden of explaining what is meant in plain language should be borne by the author. Too frequently, however, it is not.[39]

4.3 Clear and Unclear Causal Questions

Unfortunately, it seems to me that there are many "commonsense" questions, often asked by economists and other social scientists, that don't satisfy the above desiderata and consequently are at least (potentially) confusing or undecipherable. Consider the "commonsense" question: "what is the effect of divorce on children?" It sounds simple enough, but is actually quite hard to understand.

[35] In the interests of brevity, I have omitted discussion of "random coefficient" models, for example, which allow for the possibility that the effect of a treatment is not the same for all persons. This possibility is not to be confused with the condition I have described here. To illustrate, we do not require that the returns to an extra year of school in terms of wages may to be uniform across all types of persons for the question about the returns to schooling to be well posed. One may, for example, measure higher returns to an additional year of schooling for children from low socio-economic status backgrounds than those from rich backgrounds. However, the requirement I have stated in this context is that the same types children receive the returns to schooling regardless of how they were "encouraged." For a nice exposition of how a little bit of formal modeling can make sense of heterogeneous treatment effects, see Card (1999).

[36] Indeed, it is unfortunately not uncommon to see even the RCT badly executed. In work I have been doing with Jane Dokko and Justin McCrary evaluating RCTs for treatments of chronic pain, it is trivial to find hundreds of examples that are so poorly executed up as to make confident inference about anything impossible.

[37] Even the need for randomization is not obvious on all or even most contexts. See Harville (1975) for one such exposition and Heckman (2005) for a broader and more recent discussion.

[38] In this regard, the philosopher Ian Hacking has done a great deal to show that useful work can be done in areas that vary quite widely in how well posed the questions are. For a study of statistical questions, see Hacking (1965), the role of experimentation in natural science (Hacking 1983), multiple personality disorder (Hacking 1995) and the "social construction of reality" (Hacking 2000) for example.

[39] This point is not in anyway unique to me. For different, but not unrelated views of these issues with relevance to social science see Holland (1986), Freedman (1999), Pearl (1997), Heckman (2005) and Shadish et al. (2002), to name just a few.

Clearly (?) having wonderful parents (whatever that is) is good. If one begins with the premise that wonderful parents (whatever those are) are less likely to divorce than un-wonderful parents, (say because it is more difficult to be wonderful with one's child where one is involved with an ongoing battle with one's spouse), it's pretty hard to think about what the "effect of divorce" would be.

Consider a pool of "divorce prone" parents and imagine randomly assigning a "marriage encouragement." Half of these parents might be assigned to the treatment, the other half to the control and the average outcomes compared. Assuming the encouragement works, we would expect the treated group to be much more likely to remain married than the controls. If divorce was a bad thing we would expect average outcomes to be better for the kids in the treatment group.

Whatever other problems this set up has or doesn't have, it hard to imagine that the "effect of divorce" could be separated from the type of "encouragement." Consider encouragement method one: for people in the treatment group, we promise to shoot the parents if they try to get divorce. I think it is safe to say that the rate of divorce would be lower among those the "encouraged" to remain married relative to the control. Now consider encouragement method two: for people in the treatment group, we provide free marital counseling, health care for the children, money if their income is low, a new job if they hate their boss, etc. In this case too, I suspect, we would find the divorce rate to be lower in the treatment group. So far, so good.

However, one surely expects that the outcomes of the children who received the "marriage treatment" not be the same in both cases, though the putative treatment is the same. I have deliberately chosen these fanciful encouragements (assignment mechanisms) to make my argument clear: unless there is a set of widely different encouragements that "manipulate" marriage and can be reasonably supposed to have the same "effect", working "only" through the probability of remaining married, then we can't really talk about the effect of "divorce" *per se*.[40]

5 "Interesting" Questions in Freakonomics?

Many of the "interesting" questions in economics (and Freakonomics as well) strike me as poorly posed at best. Though some of these questions might admit of a meaningful causal (or other) interpretation, one often hopes for more explanation than is provided in several of the examples **Freakonomics**. Indeed, the divorce example above is arguably a bit more clear than the example they pursue in two chapters – "how much do parents really matter?"

Let me begin with stating that there is much I agree with in the chapter:

1. The advice of "parenting experts" should be met with deep skepticism at best.

2. The research in Cullen, Jacob and Levitt (2003) justifies a long discussion (in my view, even more than the book provides. It is qualitatively several notches above most of the research done on school choice, and the paper itself is a marvel of clarity and honest reporting of results.) This isn't the case solely because randomization was involved (even though that

[40]This is related to the usual requirement of instrumental variables estimators such as 2SLS that there exist a variable (the encouragement) which acts to generate experimental variation in the "endogenous" variable (in this case marriage) that be excludable in the equation determining the outcome. In this interpretation, the "problem" is is the inability to "control" for the independent effects of death threats or large amounts of assistance on child outcomes. My point here, however, is similar to the point made above. If we are unable to agree upon a set of manipulations of marriage which (when manipulated) produce the same "effect" then we are reduced to being able to describe the causal effects of "death threat induced" marriage or "financial assistance induced" marriage, etc. In the limit, if each manipulation of the marriage probability yields a different "effect" (and not merely because the set of people induced to remain married is different) then it does not seem fruitful to discuss an effect of marriage *per se*.

helps the credibility substantially) but because it provides a useful lesson, *inter alia*, about how much hard work is involved to get a credible answer even in "easy contexts."

3. Even though I can't come up with a simple "experiment" to test the hypothesis that "honesty may be more important to good parenting than spanking is to bad parenting" (Page 171). I think honesty is a good strategy (even if it didn't have a causal effect on a child's test scores; the salient issues (for me) have to do with ethical behavior.)

Unfortunately, much of the chapter is a discussion of Fryer and Levitt (2004b) (pages 163 – 176) and is, for me, is at best a long hike in a forest of confusion. Surprisingly, they use it for a short tutorial about regression analysis ("knowing what you now know about regression analysis, conventional wisdom, and the art of parenting") and they spend a great deal of time discussing what is essentially a pair of "kitchen sink regressions" (regressions with enormous numbers of covariates) from Appendix A-2 of Fryer and Levitt (2004b) using data from the Early Childhood Longitudinal Study of test scores. In their presentation, they invite the reader to consider several things that are positively correlated with a child's test scores (presumably after conditioning on a huge laundry list of (unmentioned) variables):

> the child has highly educated parents, the child's parents have high socio–economic status, the child's birth mother was thirty or older at the time of her first child's birth, the child had low birthweight, the child's parents speak English in the house, the child is adopted, the child's parents are involved in the PTA, the child has many books in his home.

as well as things that aren't correlated (by which they mean, I believe, so imprecisely estimated that a null hypothesis of no correlation can not be rejected using standard procedures):

> the child's family is intact, the child's parents recently moved into a better neighborhood, the child's mother didn't work between birth and kindergarten, the child attended Head Start, the child's parents regularly take him to museums, the child is regularly spanked, the child frequently watches television, the child's parents regularly read to him every day.

At some points, they seem to suggest that the results of this analysis speak to nothing causal: "the ECLS data don't say that books in the house [or any of the variables in their analysis] *cause* high test scores; it says only that the two are correlated." Elsewhere they seems to suggest the opposite:

> Now a researcher is able to tease some insights from this very complicated set of data. He can line up all the children who share many characteristics – all the circuit boards that have their switches flipped in the same direction – and then pinpoint the single characteristic they *don't* share. This is how he isolates the true impact of that single switch – and, eventually, of every switch – becomes manifest. (Page 162)

This last description seems more apt about how one learns to program a computer without a manual than anything to do with learning about "causes" in social science. In my experience, I have never seen any case where anything is learned by selective interpretation of scores of coefficients in kitchen sink regressions.

For example, whatever one thinks of Head Start, making anything of the observation that "according to the [kitchen sink regression using] ECLS data, Head Start does nothing for a child's future test scores" seems unwise at best. The research design can not credibly support that inference.

To make this clear, consider other inferences (albeit undiscussed in **Freakonomics**) from the same regressions. Why not, for example, observe that participation in WIC (Women, Infants, and Children) significantly lowers test scores?[41]. Perhaps such assistance actively harms children: I would argue that the good reason for avoiding *that* inference works just as well as a rationale for avoiding the inference they *do* make about Head Start: there is no reason to believe that (conditional on the other non randomly assigned regressors) that a coefficient in a kitchen sink regressions reliably informs us about causation.

Again, even kitchen sink regressions have their place: one can sometimes make a case for inclusion of scores of covariates in some very selected contexts. Despite the commonness of such analyses in economics, however, an algorithm which allows the research to decide which coefficients represent "causal" effects and which ones are regression artifacts *after* one has seen the regression output is unlikely to result in much progress in understanding.

5.1 Can Regression Help Distinguish "Cause" from "Consequence"?

Chapter 6, "Perfect Parenting, Part II; or: Would a Roshanda by Any Other Name Smell as Sweet?" begins this way:

> Levitt thinks he is onto something with a new paper about black names. He wanted to know if someone with a distinctly black name suffers an economic penalty. His answer – contrary to other recent research – is no. But now he has a bigger question: Is black culture a cause of racial inequality or is it a consequence? For an economist, even for Levitt, this is new turf – "quantifying culture" he calls it. As a task, he finds it thorny, messy, perhaps impossible, and deeply tantalizing.

As with eugenics, the history of social science research suggests that scholarly research into race that makes extensive use of correlations should be taken with a large grain of salt. The most well–known example perhaps is the controversy over the 1840 census which involved the putative correlation between the number of "insane and idiotic colored persons" living in a state and the proportion that were slaves. The data, which were faked (but still available today from the ICPSR) show that incidence of insanity is far, far lower in the South and the implication for the debate on slavery was clear (Grob 1978). (A far different version of "acting white" is mentioned several times in **Freakonomics**.) When talking about race, it is my view that being *clear* about what is meant is of even more important.

As someone who is frequently called upon as an econometric "script doctor" to "fix the econometrics" of some existing paper which is putatively about "causation", I have found it useful to begin with two seemingly simple questions:

1. What is y, the outcome, you wish to explain?
2. What are your key x variables and what potential causes are you interested in?

As a practical matter, the inability to provide a simple reply to the question is a good predictor (cause ?) of my inability to understand the empirical work. The above quote from **Freakonomics**

[41] From Appendix A-2, when the dependent variable is Math Score the coefficient on WIC is -0.120 with a standard error(0.020). When the dependent variable is reading scores, the coefficient on WIC is -0.104 with a standard error (0.021)

is in a chapter which, *inter alia*, discusses research from Fryer and Levitt (2004a) and (far more briefly) Bertrand and Mullainathan (2004).

Let x_i be defined as the "black culture" of individual i as defined by Fryer and Levitt – their "Black Name Index" (BNI).[42]

In Fryer and Levitt (2004a) much of the evidence on whether "black names" are cause or consequence comes from two types of regressions. Superficially, it would appear that they run the regressions "both ways:" in some, x_i is an independent variable, in a second set, it plays the role of a dependent variable. As is well–appreciated, this is a problem even when it occurs in different literatures (Kennan 1989).

Further inspection suggests that this is not strictly the case: in the first set of regressions (see Table II "Determinants of name choices among blacks" of Fryer and Levitt (2004a)) the dependent variable is the BNI of a given child and the explanatory variables are a number of things many of which are presumably correlated with outcomes (mother's age at time of birth, father's age at time of birth, months of prenatal care, percentage of Black babies in zip code, per capita income in the birth place, parental education, etc.). In another set (Table III "The Relationship Between Names and Life Outcomes"), BNI becomes an explanatory variable, and the dependent variables are outcomes such as "percent Black in residential zip code as an adult", years of education (the woman herself), the woman's age at first birth, etc.

Fryer and Levitt (2004a), are forthright in admitting that their evidence is consistent with a number of very plausible (but very different) alternatives that are consistent with their regressions but not necessarily with their conclusion: "With respect to this particular aspect of distinctive Black culture, we conclude that carrying a black name is primarily a consequence rather than a cause of poverty and segregation."

I have no wish to dispute their conclusion; rather I wish to suggest that there is no configuration of the data of which I am aware which would credibly support the view held by Fryer and Levitt *and not support very different alternatives.* In short, this is because it is very difficult to know what is being asked and what would constitute an answer. Put differently, it seems to me that there is at least one ill posed question floating about. Is it possible to talk meaningfully about "manipulating" culture? (and if one could, would one want to?)[43] Might reasonable people agree on some variable, policy, etc. that served exclusively to manipulate black culture and affected economic outcomes only through its effect on "culture." It is not even clear that "culture" and "economic outcomes" or "racial inequalities" are distinct entities. Indeed, as the word is often understood culture often includes the distribution of "economic outcomes." For instance, one might remark: "the fact that Bill Gates earns several times more in a year than the sum earned by all Chicago Public School teachers is a distressing fact about U.S. culture."

Further muddling the issue is the way Levitt and Dubner discuss studies such as Bertrand and Mullainathan (2004):

> So how does it matter if you have a very white name or a very black name? . . . In a typical audit study, a researcher would send two identical (and fake) résumés, one with

[42] I am stipulating, of course, that Levitt and Fryer's measure of "distinctively black name" – crudely put a function of the relative frequency with which a specific name is chosen for black children and the relative frequency with which the same name is chosen for white children – provides a measure of whatever "culture" is. They refer to this as the "Black Name Index" (BNI). A lot of non–obvious measurement issues arise. A few moments reflection, for instance, makes clear that the level of "black culture" is, by definition, a function of "white" culture. Second, a white man named Maurice Ravel might be measured as have more black culture than a black man named Paul Robeson, Jr. regardless of their actual "culture" if Maurice was relatively more popular among blacks than Paul.

[43] The paper seems to suggest that they have the usual "manipulationist" version of cause in mind. For example, there is a brief mention of the fact that there are no obvious instrumental variables which would be of no moment unless they conceived of a potential manipulation.

> a traditionally minority–sounding name, to potential employers. The "white" résumés have always gleaned more job interviews. ... The implication is that black–sounding names carry an economic penalty. Such studies are tantalizing but severely limited, for they can't explain *why* [someone with a black sounding name like] DeShawn didn't get the call.

First, as Bertrand and Mullainathan (2004) are clear to explain, they are *not* interested in the lifetime "economic cost" of a black sounding name, which is not obviously an interesting and/or well–posed question. Rather they are interested in "experimentally manipulat[ing] [employer's] perception of race." Unlike "culture" or an individual's "black name" Bertrand and Mullainathan's question seems well-posed: it much easier to conceive of a salient experiment manipulating "perceptions" than a salient experiment manipulating the naming decisions of parents. One can argue that the causal effect of manipulating perceptions of race is "uninteresting" on a number of grounds, not the least of which is that the manipulation itself doesn't suggest an intervention we might wish to undertake as a society.[44] Nonetheless, the question seems well–posed and may be answerable with regression, even if one wants to argue that it is uninteresting on other grounds.[45]

Second, although Dubner and Levitt are correct to argue that studies involving résumé randomization are unlikely to tell us "*why* DeShawn gets fewer callbacks" – as I explained in section (4.1) it is not clear what a satisfactory explanation of "why" would look like. It is even harder to understand how the type of of regressions performed in Fryer and Levitt (2004a) would, in principle, help be relevant to this discussion. (Again, they might be, but the link is not obvious to me.) Perhaps like Dr. Pangloss, we could trace Jamal's bad luck with employers to necessity: it is necessary for this to be the case, for us to be able to live in this the best of all possible worlds.

Why questions, or more specifically information on mechanisms, require a lot more than a set of OLS estimates. It is certainly the case that even in an RCT on a treatment for head pain, for example, we get meager information at best on the mechanism by which the treatment has its effect.[46]

More generally, reasoning backward from an effect (not calling back Jamal) to a "cause" (why employers don't call Jamal) in social science is generally fraught with peril – people are complicated

[44] When I teach applied econometrics I discuss Fryer and Levitt (2004a) and Bertrand and Mullainathan (2004) as a pair. Most of my graduate students conclude that Fryer and Levitt (2004a) do not pose a "meaningful causal question." Among the same graduate students the most frequent objection to the conclusions in Bertrand and Mullainathan (2004) is that the experiment doesn't manipulate racial perceptions as much as it manipulates perceptions of "uncommon-ness". The argument is that employers believe that "Moon Unit" and "Dweezil", for example, are less productive than "Jean" and "John." For what it is worth, this seems besides the point. By construction, "black sounding names" are more uncommon than white sounding names in the U.S.: there are fewer blacks than whites. If "Dweezil" or "Beauregard" don't get call backs that would be interesting but not of obvious importance to Bertrand and Mullainthan's thesis. By way of analogy, what would I make of the fact –supposing it were true– that in South Africa, where I assume white names are more uncommon, I learned that that in a broad sample of employers, Johannes, Hedrik, Balthazar and Pieter (the names of the last South African Apartheid Prime Ministers) get *more* callbacks than the presumably more common Black African names of Jayaseelan, Mbhazima, and Zwelinzima (the first names of the most recent General Secretaries of the Congress of South African Trade Unions).

[45] The fact that employers call back "Jamal's" much less frequently than "John" may not be based solely on self–conscious racial hatred, but might reflect "only" "statistical discrimination" (i.e. employers are merely acting as sophisticated econometricians, extracting all the useful information not provided by a résumé about the likely productivity of workers based on their first names, and then choosing based exclusively on "merit",) or some other mechanism (although this may be of little comfort to Jamal or John.) See Thacher (2002) for a thoughtful discussion of the issues involved in "profiling."

[46] The mechanism by which sumatriptan reduces the frequency of migraines is a subject of constantly evolving debate although there is a mountain of RCT evidence that has (at least limited) success in some types of migraineurs.

enough that there is rarely a single answer to the question "why" – often there are many interacting "reasons." Absent some fairly articulated model of how the world works, it seems difficult to even know what would constitute a good answer. To me, it often seems that putative explanations of "why" some complex human interaction occurs are frequently used as a device to *end* a debate just at the point when the issue begins to get interesting. If X is *the* reason Y occurs, why look further? Many readers might be familiar with this aspect of some answers to "why" questions: one thinks of a parent who tries to end a long conversation with a child whose replies to a parent's increasingly complicated responses is "Why?" Again it is not that a satisfactory answer to such question is not desirable: it just seems like way too much to hope from a small set of OLS regressions.

Finally, in asking a regression to distinguish "black culture" as a *cause*, from black culture as a *consequence* of economic conditions, we are very far from the types of questions I discussed in section 4.2, but there is no clear discussion in **Freakonomics** of what question is being ask and the "ground rules" that we might use to determine when the question is answered satisfactorily. It is possible that the question is well posed, but at a minimum, it is not very obvious. After reading **Freakonomics** and the original source material, I haven't gained any understanding of issues involved or even how to think about what are the answerable questions.

5.2 Possibly Well Posed But Confusing and/or Ambitious Questions

For me the most confusing section of **Freakonomics** is the discussion of "Why do drug dealers live with their moms" and "Where have all the criminals gone?." Between them, the chapters contain references to scores of articles of varying degrees of scholarship. Much of the former chapter discusses Levitt's work with sociologist Sudhir Alladi Venkatesh who collected a large amount of detailed data on one Chicago gang. For those surprised as to why gang members don't frequently live in the nicest homes in town, it will be a useful corrective. (For an earlier discussion that covers similar ground see Reuter, MacCoun and Murphy (1990).) The discussion also includes the conclusions of some very careful work by Almond, Chay and Greenstone (2003) that document the key role that hospital integration in Mississippi played in improving the appalling infant mortality rate of black children – before integration, these infants were often left to die of very preventable causes such as diarrhea and pneumonia.

Sometimes causal questions are reasonably well posed but difficult to answer. Consider Dubner and Levitt's argument that "it is clear that one of the major factors pushing [the upward trend in violent crime during the 1960s] was a more lenient justice system"[47]

This is a very difficult claim to establish at best and **Freakonomics** cites no research that speaks directly to that question. How might one try to assess effect of the "likelihood" of punishment on crime rates or how "lenient" the justice system? Part of the problem is that an "ideal" experiment to evaluate the importance of long prison sentences would involve randomizing punishment regimes and comparing crime outcomes between those exposed to high punishment regimes and low punishment regimes.[48] As Kessler and Levitt (1999) observe, it is important to distinguish between long sentences *incapacitating* or *warehousing* criminals and *deterring* persons from committing crime. Moreover, for periods during this "idyllic" 1960s, I'm not aware of any research with credible designs that reach the conclusion that changes in deterrence (within the ranges we typically see in U.S. data) matter very much. Moreover, to judge from Katz, Levitt

[47] In an early chapter Dubner and Levitt remark that "The 1960s and 1970s were, in retrospect, a great time to be a street criminal in most American cities. The likelihood of punishment was so low – this was the heyday of a liberal justice system and the criminals' rights movement – that it simply didn't cost very much to commit a crime. (Page 111)

[48] For one recent attempt see Lee and McCrary (2005).

and Shustorovich (2003) prison conditions were significantly *less* idyllic in the 1960s. One of their admittedly crude proxies is state level prison death rates (not from executions, but illness, etc.). Over the period 1950 to 1990 this averaged 3.10 death per thousand prisoners. From Figure 1 in of that paper it appears that death rates were at least twice as high during the 1960s as the 1980s.[49] Perhaps "a more lenient justice system" was a "major factor." As a reader, it was not at all clear why. At a minimum, it would have been nice to have some discussion of the distinction between "deterrence" and "incapacitation" and some documentation to point the curious reader to the basis for the claims.[50] and some discussion of the basis for the claims.

5.3 Why A Transparent Research Design Helps

Much of the chapter on "where have all the criminals gone?" deals with Romania's abortion ban, which I discussed earlier. This chapter also includes the controversial material on whether "abortion lowers crime rates."

As a purely personal matter, given the long, deep, and ugly relationship between statistical analysis and eugenics, what might emerge from this debate seems too meager to justify the effort on this subject.[51] Merely participating in the discussion one runs the risk of coarsening of the debate on how we treat the poor – the usual the target of eugenic policies.[52]

[49] **Freakonomics** refers readers curious as to whether politicians had gone "soft on crime" to three articles by Nobel Laureate Gary S. Becker: Becker (1994), Becker (1985) and Becker (1993) originally published in *Business Week*. The most salient of these three is perhaps Becker (1993) which *inter alia* cites Wilson and Herrnstein (1985) as summarizing the evidence on whether "appropriate punishments – especially raising the certainty of punishment via more police, quicker trials, and higher conviction rates – are effective in reducing the number of criminals who rob, steal, or rape. My judgment on the evidence is a bit more skeptical. Though the cited book often has useful discussion, it a bit of a curiosity in many respects from today's vantage point. Wilson and Herrnstein (1985), for example, include an explicitly Skinnerian theory of crime, which to its credit, is quite clear and laid out. More embarrassing for me – when I was reading this book in the library – was the section of the book that included several photographs of naked men to illustrate "body types" alleged to be often correlated with crime. I myself apparently have the criminal body type!

[50] See Kessler and Levitt (1999) and Lee and McCrary (2005). Long sentences might lower crime either by merely "warehousing" criminals so they can't commit crimes (except in prison) and "deterring" them – causing them to revise the calculations that lead to the criminal behavior.

[51] Eugenics, often popular among "progressive" members of the elite, was a leading motive for the development of regression. Sir Francis Galton, who gave us the word "regression," was an ardent eugenicist. For example, what is now the "Galton Laboratory, Department of Human Genetics and Biometry" at University College London, was originally named the "Galton Laboratory of National Eugenics."

[52] Indeed, the debate has grown coarser. William Bennett, a former government official, after appearing to dismiss the "abortion – crime" hypothesis in **Freakonomics**, remarked in in a talk show that: "I do know that it's true that if you wanted to reduce crime, you could – if that were your sole purpose, you could abort every black baby in this country, and your crime rate would go down. That would be an impossible, ridiculous, and morally reprehensible thing to do, but your crime rate would go down." I of course agree that "it would be a morally reprehensible thing to do." On the other hand, the premise that "you could abort every black baby in this country and the crime rate would go down" is unsupportable at best, racist at worst.

Levitt's thoughts on the subject (as well as a transcript of the relevant portion of Bennett's remarks) are available at the website `http://www.freakonomics.com/2005/09/bill-bennett-and-freakonomics.html`.

For what it's worth, Levitt's remarks are admixture of what strike me as reasonable assertions and others that are confusing at best, wrong at worst. For example, on his blog `www.freakonomics.com` Levitt argues:

...

6 "If we lived in a world in which the government chose who gets to reproduce, then Bennett would be correct in saying that "you could abort every black baby in this country, and your crime rate would go down." Of course, it would also be true that if we aborted every white, Asian, male, Republican, and Democratic baby in that world, crime would also fall. Immediately after he made the statement about blacks, he followed it up by saying, "That would be an impossible, ridiculous, and morally reprehensible thing to do, but your crime rate would go down." He made a factual statement (if you prohibit any group from reproducing, then the crime rate will go down), and then he noted that just because a statement is true, it doesn't mean that

Caveats aside, here goes.

In their original article, Donohue and Levitt (2001) cite two possible "theories" about the consequences of abortion legalization. Neither of them fit well into the framework described in Section 4.2. Note that one could conceive of cases where abortion might be thought of (for better or worse) as a treatment: that is generally true when the subject of interest was child–bearing women (not their fetuses). The question of what happened to the welfare of women who are given the choice of having abortion relative to those that have been denied such choice, is well posed. One merely would seek to compare a group of women given the opportunity to have a an abortion to those who did not. Of course, this is much easier said than done (and indeed is the subject of much of the pre–Donohue and Levitt (2001) work by economists on the consequences of abortion legalization.)

The "effect" of abortion legalization on *crime*, of course, is a whole different matter. Donohue and Levitt (2001) discuss two possible mechanisms at length.

Donohue and Levitt (2001) first argue that "The simplest way in which legalized abortion reduces crime is through smaller cohort sizes."

While possibly "simple", it is amazingly difficult to articulate clearly in a regression framework where the unit of observation is the individual. At its core this hypothesis appears to include the implicit assertion that among other things, my mother's decision not to abort the fetal John DiNardo caused some other children's propensity to commit crime to increase. (Although it should be said, it clearly raised mine!) Such effects are difficult to identify, even in the easiest cases (Manski 1993).

A far more subtle mechanism is distinct from the first, although it could certainly interact with it. "Far more interesting from our perspective is the possibility that abortion has a disproportionate effect on the births of those who are most at risk of engaging in criminal behavior."Donohue and Levitt (2001)

To anyone who has given the problem of "missing data" some thought, it is difficult to be sanguine about the possibility of inferring much about the criminal propensities of those who are never born. Even in the context of a medical RCT, the analogous problem of attrition is often distressingly difficult to cope with. Moreover, the problem is so difficult that in the RCT one

it is desirable or moral. That is, of course, an incredibly important distinction and one that we make over and over in Freakonomics.

7 "There is one thing I would take Bennett to task for: first saying that he doesn't believe our abortion-crime hypothesis but then revealing that he does believe it with his comments about black babies. You can't have it both ways."

As far as I can tell the statement about lowering the *level* of crime by aborting Native American, Republican, . . . fetuses is a non-sequitor at best. Bennett is clearly talking about the *rate* of crime. I can only make sense of the statement by construing it to mean that ridding the planet of human life would eliminate crime (at least that caused by humans.)

As to the rest of the explanation:

- One does not make a "factual statement" by claiming that "if you prohibit any group from reproducing, then the crime *rate* [my emphasis] will go down." I know of no "successful" eugenic program that has "lowered the crime rate."
- Neither is there any reason to believe that "if we lived in a world in which the government chose who gets to reproduce, then Bennett would be correct in saying that 'you could abort every black baby in this country, and your crime rate would go down.'"
- Contrary to Levitt's claim, I do not think it necessary to believe that the termination of black fetuses would lower the crime rate *even if* the causal effect of abortion legalization in the U.S. had been a reduction in crime. As I explain below, even if one stipulates that crime reduction was a causal effect of abortion legalization in the U.S. this would tell us nothing about the causal consequences of aborting black (or any) fetuses.

often abandons hope of modeling non-response or sample selection and seeks merely to bound the difference between the treated and control groups (Horowitz and Manski 1998). Indeed, one rarely confronts a situation where attrition from the study is the "goal" of the treatment – with good reason.

Moreover, as Donohue and Levitt (2001) observe, there are many mechanisms besides abortion to either stop the "criminogenic" fetus from being born or prevent the child from becoming a "criminal" once born.

> Equivalent reductions in crime could in principle be obtained through alternatives for abortion, such as more effective birth control, or providing better environments for those children at greatest risk for future crime. Donohue and Levitt (2001)

A description from **Freakonomics** provides one possible suggestion:

> How, then, can we tell if the abortion-crime link is a case of causality rather than simply correlation?
>
> One way to test the effect of abortion on crime would be to measure crime data in the five states where abortion was made legal before the Supreme Court extended abortion rights to the rest of the country.... And indeed, those early-legalizing states saw crime begin to fall earlier than the other forty-five states and the District of Columbia. Between 1988 and 1994, violent crime in the early-legalizing states fell 13 percent compared to the other states; between 1994 and 1997, their murder rates fell 23 percent more than those of the other states. (page 140)

Of the identification strategies employed in this literature, this is the most transparent. To understand what is going on, assume that pre-Roe legalization provided a Brandiesian natural experiment of sorts. Instead of the individual being the unit of observation, think of each state as sort of identical petri dish to which a drop of abortion legalization is being added. Fifteen to twenty five years later, the petri dishes will be checked again to seem how much per capita crime is occurring. If legalization had been an actual experiment (perhaps run by a dictator), we might have expected half the states to be legalizers and the other half to never legalize (assume that items in the petri dishes can't jump into other petri dishes.) That of course did not happen. In this case, the experimenter added a drop of legalization to 5 states in 1970, and then added a drop to the remaining states a scant three years later. Of course, it wouldn't be clear that even in this experiment you could detect an "effect" on crime unless the effect were large relative to the variation across the petri dishes we would expect in the absence of any experiment.[53] (Note of course, that such an experiment could provide us essentially no information on the "mechanisms" – it could be a complicated interaction of many things having little to do with selective abortion or cohort size *per se*. Merely the *option* of having an abortion might change outcomes for many reasons.)

Though one would not know from reading **Freakonomics**, Donohue and Levitt (2001) argue that this research design is inadequate.[54]

[53] Indeed, this or similar identification strategy is employed in such work as Charles and Stephens (2006), Gruber, Levine and Staiger (1999), Bitler and Zavodny (2002), as well as Joyce (2004b). Gruber et al. (1999) detect a rather small (and brief) effect on the total number of children born from this identification strategy.

[54] They argue *against* the identification strategy both on *a priori* grounds and on *ex post* grounds (the implausibility of the results so obtained.) In Donohue and Levitt (2001), for example, when they deploy that identification strategy, they report that "the cumulative decrease in crime between 1982-1997 for early-legalizing states compared with the rest of the nation is 16.2 percent greater for murder, 30.4 percent greater for violent crime, and 35.3 percent greater for property crime. Realistically, these crime decreases are too large to be attributed to the three-year

Consequently, much of this is beside the point. Donohue and Levitt (2001) argue that evidence from such a research design is only "suggestive."

The bulk of their argument centers on their attempts to "more systematically" analyze the relationship with an analysis of state level crime data on lagged "abortion rates."

Consider equation (1) from Donohue and Levitt (2001)

$$A_t \equiv \text{Effective Abortion}_t = \sum_a \text{Abortion}^{\star}_{t-a} \frac{\text{Arrests}_a}{\text{Arrests}_{\text{total}}}$$

which they label the "effective abortion rate" (the asterisk seems to be an acknowledgment of the fact that they don't have reliable data on the abortions before it became legal.) They then divide this by the number of live births to get an "effective abortion ratio"

$$\mathfrak{A}_{st} = \frac{A_{st}}{\text{LB}_{st}}$$

Much of the more systematic evidence on the link between abortion legalization and crime is a result of regressions of the form:

$$\log \text{Crime Per Capita}_{st} = \beta_1 \mathfrak{A}_{st} + X_{st}\Theta + \gamma_s + \lambda_t + \epsilon_{st} \tag{2}$$

where s and t refer to states and years and each observation is the relevant state/year average or value. X_{st} are a set of covariates, γ_s are a set of state dummy variables and λ_t are a set of year fixed effects. ϵ is a random disturbance that is presumably uncorrelated with any of the regressors. In words, up to a constant that differs by states, absent variation in X or the (modified) abortion ratio, it is assumed that trends across state in crime would be the same.

Stipulating that all of the data used to generate this specification are fine[55] I find it impossible to interpret the coefficients at all. In common econometric parlance, the abortion ratio is "endogenous". Indeed, some work has looked a the effect of economic and other conditions on abortion (Blank, George and London 1996): that is, something akin to $\mathfrak{A}$ is the *dependent* variable in the regression. Donohue and Levitt (2001), however, spend surprisingly little time discussing the issue.[56]

Moreover, I don't know what the "ground rules" that a skeptical, but persuadable person should use for evaluating this regression. Other than the "the coefficients look reasonable" – what would speak to the credibility of the research design or what should lead me to reject it?

The notion that we should be reassured about the existence of an "abortion –crime" link because the OLS coefficient on $\mathfrak{A}$ in a regression like equation (3) is robust to the inclusion of some covariates is not obvious. One "intuition" that motivates investigating whether a result is "robust" to the inclusion of a large number of explanatory variables comes from the RCT. On average, if we repeat the experiment, the answer we get from including covariates and from excluding covariates should be the same.

head start in the early-legalizing states." The reservations in Donohue and Levitt (2001) about the estimates generated with this identification strategy do not appear in **Freakonomics** which selectively discusses some comparison between early and late legalizing states.

[55] This is perhaps more than we should stipulate to: our knowledge of the number of illegal abortions today or abortions that preceded abortion legalization in the 1970s is meager at best. Moreover, Donohue and Levitt (2001) and other researchers do not have data on the amount of crime committed by individuals of a given age. At best one has very crude proxies. See Charles and Stephens (2006) or Joyce (2004b) for discussion.

[56] In the published version of the paper, the word "endogeneity" appears only regarding a discussion of two right hand side variables – number of police and prisons – which are "lagged to minimize endogeneity." The word "exogeneity" appears in confusing discussion about the difference between high and low abortion states (page 401.)

On the other hand, clearly it does not make sense to think of $\mathfrak{A}$ as "randomly assigned." Indeed, if abortion legalization is all about "selection" – i.e. the difference in the crime propensities of those born and those not born – pure random assignment of abortion (a thought too grotesque to even contemplate) would not merely leave the statistical problem unsolved, it would answer a different (even more uninteresting) question. For example, in one version of the Donahue–Levitt story, abortion matters for crime because it is the consequence of *choice* made by women to *selectively* abort some fetuses and not others. "Random abortion" would, on the other hand, would produce no "selection effect" – studying such "random" variation in abortion ratios would be silent about the putative effects of legalizing abortion.[57]

If thinking about the regression as an approximation to some sort of randomized controlled trial doesn't help, how is one to even *assess* or interpret the specification? What is missing from this research is *either* a similarity to the simple type of question I described in Section 4.2 *or* an explicit model of the link between abortion legalization and cohort size. (See Gruber et al. (1999) for one simple example of a model.) Absent that, it is hard to understand why this (or similar evidence) should persuade anyone (one way or the other.)

Consequently, I've only been able to guess at what valid interpretation of the coefficient on the abortion ratio would be. One guess is that this regression is makes sense under some explicit model of fertility, abortion, crime, etc. but no such model has been provided.[58] The absence of such an articulated model in Donohue and Levitt (2001) is surprising since selection is not merely a nuisance, but is the object of interest.

Recent revelations regarding a programming error in Donohue and Levitt (2001) are instructive in this regard. The regression Donohue and Levitt (2001) claims to have run looks something like:

$$\log \text{Arrests}_{stb} = \beta_1 \mathfrak{A}_{sb} + \gamma_s + \lambda_{tb} + \theta_{st} + \epsilon_{stb} \tag{3}$$

where b denotes year or birth, so for example, $\mathfrak{A}_{sb}$ denotes the modified abortion ratio for the cohort born in state s in birth year b. As Foote and Goetz (2005) demonstrates, due to a programming error, the set of state$\otimes$year dummy variables (the terms represented by θ_{st}) were supposed to have been included were not. Foote and Goetz (2005) go on to argue that the "correct" specification should include θ_{st} and that the dependent variable should be log Arrests per capita not the total arrests in the state. On his authors blog, Levitt (2005) has a variety of responses:

1. Foote and Goetz (2005) correctly identified that there was a programming error. "Once you made those changes [included the state$\otimes$year effects and used per capita arrests] the results in originally Table 7 disappear[ed]."

2. The regressions were discussed in a section of the paper that "was the most speculative of analysis of all that we did and frankly we were surprised it worked at all given the great demands it put on the data."

3. The data used in Levitt (1997) and Foote and Goetz (2005) are noisy and that using a new measure of the abortion ratio, the coefficient on the abortion ratio is "significantly different

[57] The fact that abortion ratios are surely endogenous makes it impossible for me to understand other research designs in the paper such as a comparison of states with high versus low abortion rates.

[58] One might, for example, write down some version of the basic "selection bias" model (Gronau 1974, Lewis 1974, Heckman 1979).

What is usually required for such a system to be identified is parametric knowledge of the above equations and/or a variable which affects the probability of being born, but is uncorrelated with the determinants of crime. See for example, Heckman and Robb Jr. (1986), Ahn and Powell (1993), Das, Newey and Vella (2003) for a discussion. Given the fact that the data is collected at the level of the state and not the level of the individual, one idea is to write down an explicit model as in Gronau (1974) and Lewis (1974). See also the useful appendix in Card and Rothstein (2005).

from zero" except when they perform an instrumental variables analysis using one measure of the abortion ratio as an instrument for the other.

The criticisms in Foote and Goetz (2005) are thoughtful and carefully executed. The fact that Levitt and Donahue made it relatively easy to identify the error is a testament to their scholarship which is greatly valued.

Where both Levitt (2005) and Foote and Goetz (2005) go wrong, however, is appearing to stipulate that (apart from measurement error, etc) it is meaningful to interpret the estimate of β_1 as the "effect of abortion legalization." It is not obvious why any reasonable person would think so. I also acknowledge that there are few "perfect" regressions. Unlike Levitt (1997), however, what is required for the coefficient are neither spelled out nor obvious. Since the regression can not be treated as an "approximation" to a randomized trial – a case where what constitutes evidence against the design is clear – it is not obvious under what conditions (what data generation process) would an OLS regression of this sort produce a reliable answer to the question addressed. No description is provided anywhere in the literature. Consequently, as to the claim that the regression is "speculative" analysis, I concur. But what regressions in the paper go beyond speculation and support the hypothesis and the research design? Elsewhere Levitt has described the regression as part of a "collage of evidence."The Economist (2005) Given the utter non–transparency of the research design, a "Rorschach ink blot" seems more appropriate than "collage." To me, it seems that it could only convince the already–convinced.

A bit more intuitively, there are a long list of reasons the abortion ratio might vary in ways having nothing to do with abortion *per se.* Women's access to contraception and other types of fertility control were undergoing some fairly significant changes. For instance, 1970 witnessed the passing of the Public Health Services Act which greatly increased some women's access to birth control – especially poor women, much of this around the time of Roe v. Wade. During this time, changes in the economic condition of women changed fairly dramatically. Surely this matters for the abortion ratio. What effect did access to better birth control and changes in norms do to likelihood that a child might become a criminal. Are such changes plausibly "controlled for?" Should they be? How would we know if they had? The "experiment" involving early and late legalizing states is already a rather mongrel experiment; having eschewed this experiment, what is left?

Did legalizing abortion lower crime? The reader who has suffered through this discussion and remains curious is encouraged to read the criticisms by Joyce as well as the original work by Donahue and Levitt as well as contemplate what an "ideal" experiment or a fully articulated structural model would look like. How well do the research designs approximate one or the other?[59] The regression I have discussed can not be interpreted as an "approximation" to randomized controlled trial. No explicit structural model is given. Is it possible that under some state of the world a regression coefficient from something like equation (3) interpretable? I suppose so, but neither the original research (nor the critiques that followed) provide any help on this front.

Perhaps if the "experiment" involving early versus late legalizing states been "big enough", the effects "large enough" and if everything else had remained "quiescent enough" a debate such as we have witnessed on the evidence would have never ensued. The answer would have been far more obvious and far less demanding of the need for just the "right" specification and the right data. As this discussion may have made clear, it wasn't.[60]

[59] For what little it is worth, my judgment is that the data do not support any claim about the "effect" of abortion legalization in the U.S. (or Canada) on crime, in essence, if not in all the particulars, endorsing the conclusions in Joyce (2004b).

[60] N.B. this is not an endorsement of the view that only uncontroversial results are to be trusted. To the contrary, the level of criticism of a finding is seems to be better predicted by how unpopular (or "un-Theoretically Correct")

5.4 Type I and Type II Error

Elsewhere, the "focus on the hidden side" seems to ignore some potentially important issues. To illustrate, let me choose one such case where a little statistics might have gone a long way. In the chapter entitled "What Do School Teacher's and Sumo Wrestler's have in Common", the authors discuss some work by Levitt on detecting "teacher cheating." In the telling, the cast of heroes includes the CEO of the Chicago Public School system and the villains include the school teachers and their labor union ("When [Duncan] took over the public schools, his allegiance lay more with the schoolchildren and their families than with teachers and their unions.") The basic method is to analyze the *pattern* of test answers. Answers that depart from the posited (ad hoc) data generation process are flagged as "cheating." For obvious reasons, at no point in the process described is actual data on *observed* teacher cheating used. As a consequence, the algorithm described has no way of discriminating between the case where a teacher selectively "corrects" a subset of answers for a class, from those cases where the students (unknown to the teacher) have obtained copies of a subset of the answers, to name one (perhaps unlikely) situation. At a most basic level, of course, there is no *perfect* way to "detect teacher cheating" with statistical analysis[61] and I don't mean to suggest that Levitt and Dubner suggest this.

Indeed, the chapter indicates that the "teacher cheating" algorithm was not the sole method used to assess guilt (one hopes so) but remarks with little further curiosity that "the evidence was strong enough only to get rid of a dozen of them." Given the rest of the discussion, this might come as quite a surprise. Why would such a clever algorithm work so poorly in a situation when there was much cheating?

Anything but a perfect "test" for the existence or "non–existence" of something (virus, cheating, etc.) commits two types of error – in unhelpful terminology, Type I and Type II. I find the legal metaphor the easiest way to remember the distinction. The legal system in the U.S. (at least

it is, rather than any supposed weaknesses of the research design. Deaton (1996) provides a couple of illustrations of this tendency in the context of the debate on Card and Krueger (1995). One doesn't have to endorse *any* of the conclusions in Card and Krueger to recognize that this problem is real: "June ONeill, [then] Director of the Congressional Budget Office, the agency charged with credibly assessing the effects of government policies, reminded [her] audience at an American Enterprise Institute meeting [about the effect of the minimum wage] that *theory is also evidence*." [my emphasis]

A more ironic illustration from Deaton (1996):

> That evidence may have to be discarded in favor of "science" could hardly be better argued than in Nobel Laureate James Buchanans words in The Wall Street Journal: "no self-respecting economist would claim that increases in the minimum wage increase employment. Such a claim, if seriously advanced, becomes equivalent to a denial that there is even minimum scientific content in economics, and that, in consequence, economists can do nothing but write as advocates for ideological interests. Fortunately, only a handful of economists are willing to throw over the teaching of two centuries; we have not yet become a bevy of camp-following whores."

[61] To make this clear, consider an analysis made by officials responsible for New York's Powerball lottery. In the March 30, 2005 drawing a startling number of persons (110) got five out of six numbers correct. According to a news report (Lee 2005), past experience with the lottery had lead them to believe that in the 29 states where the game is played, the average number of winners would be more like four or five. Cheating? Fraud? As the report explains, graud was definitely one suspect, but not the only one. "Earlier that month, an ABC television show, "Lost," included a sequence of winning lottery numbers. The combination didn't match the Powerball numbers, though hundreds of people had played it: 4, 8, 15, 16, 23 and 42. Numbers on a Powerball ticket in a recent episode of a soap opera, "The Young and the Restless," didn't match, either. Nor did the winning numbers form a pattern on the lottery grid, like a cross or a diagonal. Then the winners started arriving at lottery offices." (Lee 2005)

The first winner came in, and failed to admit cheating. The second winner came in and did the same. So did the third. Indeed, this was not a case of fraud or cheating. All three reported that they had chosen their number on the basis of a fortune cookie. Lottery investigators finally even managed to locate the fortune cookie maker who verified that his factory had produced the fortune cookie.

nominally) attempts to minimize Type I error – sending an innocent person to jail. Type II error is the opposite mistake – exonerating the guilty. In practice, there is a traded between the two types. One way to avoid Type II error is to declare everyone guilty; declare everyone innocent and one avoids Type I error at the expense of Type II error.

If the fact that only a "handful" were caught was a surprise to the reader, it wouldn't be a surprise to those familiar with Tversky and Kahneman (1974) who argued that people are frequently inattentive to "base rates" (although that interpretation is subject to a lively debate.) The canonical problem can be illustrated by making a few assumptions about the algorithm discussed in **Freakonomics**. Suppose that the probability of being detected cheating, given that you cheat is 0.90 – the probability of Type I error is .1. Also assume that the algorithm incorrectly identifies you as a cheater when you are not is .06 – Type II error. Further suppose that 4 percent of teachers cheat – this is the crucial "base rate." Slightly more formally:

$$
\begin{aligned}
\Pr(D|C) \equiv \Pr(\text{Detected Cheating by Algorithm}|\text{Engaged in Cheating}) &= .90 \\
\Pr(D|\sim C) \equiv \Pr(\text{Detected Cheating by Algorithm}|\text{Not Engaged in Cheating}) &= .06 \\
\Pr(C) \equiv \Pr(\text{Engaged in Cheating}) &= .04
\end{aligned}
$$

I wasn't able to locate the actual numbers in **Freakonomics** and the ones I have chosen seem a bit optimistic for the algorithm they describe (albeit a bit pessimistic about the fraction of cheating teachers). If they were correct, however, it would explain why only a handful of those identified by the algorithm were finally identified as cheaters – despite the large pool of potential cheaters. Many statistically naive readers might conclude that virtually all of those identified as guilty were indeed guilty. The test looks pretty accurate. Few detected cheaters are innocent, and cheaters have a good chance of being caught. However, even in this example, of the roughly 9 percent of teachers classified as cheating on the basis of the algorithm, the majority (about 62 percent) would actually be innocent. This strikes me as a frighteningly high percentage, but perhaps others will disagree.[62] A more thoughtful analysis would go even further: does it treat different but morally homogeneous groups differently? It would almost certainly give one a moment's pause if an algorithm was only (or mostly) able to detect cheating among the the lowest paid teachers with the most difficult students, but that did a poor job of detecting cheating among the most affluent. **Freakonomics** unfortunately discusses none of these issues.

6 The "Hidden Side of Everything" or the Leper's Squint?

> Standing before the altar gazing down the length of the nave to the great west door of [St. Mary's Church in Youghal, County Cork, Ireland] one can detect, high above and slightly to the right, a small opening. From this vantage point many centuries ago the town's lepers, reaching the opening by a special entry, could peek out at the devotions of the notables and merchants mustered below. Hence the ancient name for such an opening "the leper's squint." (Cockburn 1993)

[62] The calculation is:

$$
\begin{aligned}
1 - \Pr(C|D) &= 1 - \left\{ \frac{\Pr(D|C)\cdot\Pr(C)}{\Pr(D|C)\cdot\Pr(C) + \Pr(D|\sim C)\cdot(1-\Pr(C))} \right\} \\
&= 1 - \frac{.9(.04)}{.9(.04) + .06(.96)} \\
&= 1 - 0.385 \\
&= 0.615
\end{aligned}
$$

Although I am not a "linguistic determinist" of any stripe, one depressing feature (for me) of **Freakonomics**– a feature endemic to much social science writing – is the tendency to replace commonsense descriptions of behavior with essentially scientistic explanations. Abstract and technical language can be useful, but (especially in a popularization) I think it should be avoided as much as possible.

Some of this is probably inevitable: at some level this tendency is merely one manifestation of the fairly universal concept of "shop talk" – the shorthand people use to communicate concepts quickly. If my own experience is any guide, sometimes this language is required by reviewers and editors. For example, I have always been a bit hesitant to even write about immigrants and immigration for fear of some awful construction such as "the effect of Hispanic status". For me at least, sometimes such language can get in the way so much that it can be quite difficult to think clearly. In my own case, I've embarked on a self–help program to drop the use of the word "incentive" entirely [a quest in which I have not been entirely successful.]

Tied up with the use of language, is the world view advocated by Levitt and Dubner. Many reviewers have found the perspective of Levitt and Dubner "refreshing", "broad", etc. and if Levitt and Dubner have broadened the perspective economists can bring to their subject, this is surely welcome. "Broad minded economist" is not quite an oxymoron, but it often seems that way. There is no reason why economists shouldn't study political institutions, nor should political scientists be required to ignore "economics." Even the designation of fields of study as "economics" as distinct from "political science", "sociology", or "psychology" seems so intrinsically muddled that it is surely unwise to require research to fit into any particular "box." A strict distinction between fields is likely to generate a lot of foolishness. If **Freakonomics** encourages a crossing of disciplinary boundaries, that would, in itself be welcome.

Despite being a book that eschews a "unifying theme", **Freakonomics** has at least one central argument: "incentives matter" – on the other hand, it is not clear what an incentive "is." The helpful index to the book lists the following: *incentives, bright line versus murky, as a cornerstone of modern life, criminal, definitions of, discovery and understanding, economic, of experts, invention and enactment of, moral, negative versus positive, power of, of real estate agents, schemes based on, of schoolteachers, social, study, tinkering with, trade-offs inherent in.*

The authors discuss several types of incentives: economic, social, and moral which they define as "simply a *means* of urging people to do more of a good thing and less of a bad thing. [*my emphasis*]" As the authors are aware (they've discussed the issue in their blog and elsewhere) the term *incentive* is a very elastic one. My qualm is that it is so elastic as to be a hindrance to clear thinking.

In Dubner and Levitt's hands, the assertion that incentives are the "cornerstone of modern life" often comes off as a two part tautology. The first part of the tautology is: "when incentives matter, they matter." The second part of the tautology is that when incentives don't matter, it is because of "moral incentives"

Despite it's widespread usage, I'd like to take this opportunity to lobby (unsuccessfully for certain!) for the (at least temporary) banishment of the term "moral incentive". The way the term incentive is typically used by economists evokes, for me at least, a kind of Skinnerian behaviorism which in popular writing was most cogently demolished by Chomsky (1971), (although still alive among some social scientists. See footnote 49.) It is easy to get confused about whether negative and positive incentives, for example, are merely synonyms for the Skinnerian notions of negative and positive reinforcement.[63]

[63]Part of the problem, of course, is that the terms positive and negative reinforcement are notoriously hard to define in a non–circular way. Should you doubt that confusion between the Skinnerian notion of reinforcement and incentive is possible, consider the following definition of "incentive." This definition that follows *began* as a definition

Like much else in **Freakonomics**, Dubner and Levitt do not take the framework that seriously. Skinner's very explicit and detailed discussion, by contrast, is so clear that it has always struck me as a argument of the *reductio ad absurdum* sort (Skinner 1957).

Nonetheless, the term moral incentives seems to elide an important distinction between an action I (or a government, or a business) might take to affect a person's behavior – a manipulation if you will – and an aspect of a person's internal state, in what in earlier times a social philosopher might have described as a person's "soul" or "beliefs" or "convictions."

Consider one of their illustration of moral incentives (page 21) – "when the government asserts that terrorists raise money by selling black–market cigarettes that acts as a jarring moral incentive." If the persons in the government are making a well–informed, truthful, and salient claim why not merely call it information? If the claim is otherwise, why not refer to it as propaganda? (or marketing in more polite language.) Calling it "moral incentives" seems to me to conflate two very different things: deliberate manipulations outside the person, with inner states (unless one is a Skinnerian in which the inner states are infinitely flexible.)

Even the designation of moral incentives as "negative" or "positive" seems to conflate things "external" to a person with a person's inner states. For example, the practice of the Roman Catholic church in drawing up an index of "prohibited" books could be described as an "negative moral incentive" to not engage in reading such books, although to take a personal example, the *Index Librorum Prohibitorum* (despite its official demise) provided me with a wonderful reading list when I was in high school (over the objections of the nuns who taught me.) Isn't it just plainer to say that often "incentives" don't matter or that the attempts of others to control what we think or believe sometimes (thankfully) don't work? More optimistically, economists or those in a position to do so have only the crudest tools and knowledge to manipulate us?

Perhaps I read more into the use of the word incentives than is there. However consider Dubner and Levitt's description of the "typical economist's view" of incentives:

> Economists love incentives. They love to dream them up and enact them, study them, and tinker with them. The typical economists believes the world has not yet invented a problem that he can not fix if given a free hand to design the proper incentive scheme. His solution may not always be pretty – it may involve coercion or exorbitant penalties or the violation of civil liberties – but the original problem, rest assured, will be fixed. An incentive is a bullet, a lever, a key: an often tiny object with astonishing power to change a situation.

In this respect, I am apparently closer to the typical linguist than I am to the typical economist. Consider this critique of Skinner's discussion of the implications of operant condition for human behavior:

> Humans are not merely dull mechanisms formed by a history of reinforcement and behaving predictably with no intrinsic needs apart from the need for physiological satiation. Then humans are not fit subjects for manipulation, and we will seek to design a social order accordingly. (Chomsky 1971)

of the word "reinforcement". To turn it into a definition of "incentive" I merely changed the words "surroundings", "reinforcement" and "animal" in a Wikipedia entry on operant conditioning (Wikipedia 2005): "[an] incentive is any change in an person's environment that (a) occurs after the person behaves in a given way, (b) seems to make that behavior re-occur more often in the future and (c) that re-occurrence of behavior must be the result of the change." An almost identical exercise can be performed on the salient passages of my undergraduate Psychology textbook (Mussen, Rosenzweig, Aronson, Elkind, Feshbach, Giewitz, Glickman, Murdock Jr., Wertheimer and Jr. 1977).

I do not mean to suggest that Dubner and Levitt believe that humans are "dull mechanisms" formed only by a history of "incentives." I mean to suggest only that an apt metaphor to talking about humans as "behaving according to their incentives" is the Leper's Squint at the beginning of this section. It is not a viewpoint that is always entirely without merit. Just a narrow one.

References

Ahn, Hyungtaik and James Powell, "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 1993, *58*, 3–29.

Almond, Douglas V., Kenneth Y. Chay, and Michael Greenstone, "Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality in Mississippi," Unpublished Manuscript, Department of Economics, University of California – Berkeley November 2003.

Armon, Carmel and Randolph W. Evans, "Addendum to assessment: Prevention of post-lumbar puncture headaches: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology," *Neurology*, 2005, *65* (4), 510–512.

Bales, Richard F., *The Great Chicago Fire and the Myth of Mrs. O'Leary's Cow*, Jefferson, NC: McFarland & Company, Inc., October 2002.

Becker, Gary S., "Tailoring Punishment to White–Collar Crime," *Business Week*, October 28 1985, p. 20.

——, "How to Tackle Crime? Take a Tough, Head On Stance," *Business Week*, November 29 1993, p. 26.

——, "Stiffer Jail Terms Will Make Gunmen More Gun Shy," *Business Week*, February 28 1994, p. 18.

Berg, Chris, "Why do drug dealers live with their mums?," *IPA (Institute of Public Affairs) Review*, June 2005, *57* (2), 46.

Berry, Sandra H., Naihua Duan, and David E. Kanouse, "Use of Probability Versus Convenience Samples of Street Prostitutes for Research on Sexually Transmitted Diseases and HIV Risk Behaviors: How much does it matter?," in Richard B. Warnecke, ed., *Health Survey Research Methods Conference Proceedings*, Hyattsville, MD: Department of Health and Human Services, April 1996, pp. 93–97.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, September 2004, *94* (4), 991–1013.

Bitler, Marianne and Madeline Zavodny, "Did Abortion Legalization Reduce the Number of Unwanted Children: Evidence from Adoptions," *Perspectives on Sexual and Reproductive Health*, January/February 2002, *34* (1).

Blank, Rebecca, Christine George, and Rebecca London, "State Abortion Rates: The Impact of Policies, Providers, Politics, Demographics, and Economic Environment," *Journal of Health Economics*, October 1996, *15* (5), 513–553.

Card, David, "The Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam: North Holland, 1999, pp. 1801–1863.

___ **and Jesse Rothstein**, "Racial Segregation and the Black–White Test Score Gap," Working Paper 109, CEPS, Princeton May 2005.

Card, David Edward and Alan B. Krueger, *Myth and measurement : the new economics of the minimum wage*, Princeton, N.J.: Princeton University Press, 1995.

Charles, Kerwin Kofi and Melvin Jr. Stephens, "Abortion Legalization and Adolescent Substance Use," *The Journal of Law and Economics*, Forthcoming, October 2006.

Chomsky, Noam, "The Case Against B.F. Skinner," *New York Review of Books*, December 30 1971.

Cockburn, Alexander, "Beat The Devil," *The Nation*, September 6 1993, *257* (7), 234.

Cullen, Julie Berry, Brian A. Jacob, and Steven D. Levitt, "The Effect of School Choice on Student Outcomes: Evidence From Randomized Lotteries," Working Paper 10113, National Bureau of Economic Research 2003.

Das, Mitali, Whitney K. Newey, and Francis Vella, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, January 2003, *70*, 33–58.

Deaton, Angus S., "Letter from America: The Minimum Wage," *Newsletter of the Royal Economic Society*, October 1996, (95), 13.

DiNardo, John, "Natural Experiments," in Steven N. Durlauf and Lawrence E. Blume, eds., *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, Forthcoming 2006. Preliminary version at: `"http://http://www-personal.umich.edu/~jdinardo/ne6.pdf"`.

___ **, Nicole Fortin, and Thomas Lemieux**, "Labor Market Institutions and The Distribution of Wages, 1973-1993: A Semi-Parametric Approach," *Econometrica*, September 1996, *64* (5), 1001–1045.

Donohue, John J. III and Steven D. Levitt, "The Impact of Legalized Abortion on Crime," *Quarterly Journal of Economics*, 2001, *116* (2), 379–420.

___ **and** ___ , "Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce," *Journal of Human Resources*, 2004, *39.*

Drèze, Jean and Amartya Sen, *Hunger and Public Action*, Oxford: Clarendon Press, 1989.

Edlund, Lena and Eveyln Korn, "A Theory of Prostitution," *Journal of Political Economy*, 2002, *110* (1), 181–214.

Einstein, Albert, *Relativity: The Special and General Theory*, New York: Henry Holt and Company, 1920. Translated by Robert W. Lawson.

Foote, Christopher L. and Christopher F. Goetz, "Testing Economic Hypotheses with State–Level Data: A Comment on Donohue and Levitt," November 22, 2005, (05–15). `http://www.bos.frb.org/economic/wp/wp2005/wp0515.pdf`.

Freedman, David A., "From Association to Causation: Some Remarks on the History of Statistics," *Statistical Science*, 1999, *14*, 243–258.

——, "Statistical Models for Causation: A Critical Review," in Brian Everitt and David Howell, eds., *Wiley Encyclopedia of Statistics in the Behavioral Sciences*, Chichester: John Wiley and Sons, 2005.

Fryer, Roland G. Jr. and Steven D. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, August 2004, *119* (3), 767–805.

—— **and** ——, "Understanding the Black–White Test Score Gap in the First Two Years of School," *Review of Economics and Statistics*, 2004, *86* (2), 447–464.

Granger, Clive, "Statistics and Causal Inference: Comment," *Journal of the American Statistical Association*, December 1986, *81* (396), 967–968.

Grob, Gerald N., *Edward Jarvis and the Medical World of Nineteenth–Century America*, Knoxville: University of Tennessee Press, 1978.

Gronau, Reuben, "Wage Comparisons – A Selectivity Bias," *Journal of Political Economy*, 1974, *82* (6), 1119–1143.

Gruber, Jonathan, Phillip Levine, and Douglas Staiger, "Abortion Legalization and Child Living Circumstances: Who is the 'Marginal Child'," *Quarterly Journal of Economics*, 1999, *114* (1), 263–291.

Guerry, André-Michel, *Essai sur la statistique moral de la France. A translation of Andre–Michel Guerry's Essay on the Moral Statistics of France: a sociological report to the French Academy of Science edited and translated by Hugh P. Whitt and Victor W. Reinking, 2002* 1883.

Hacking, Ian, *The Logic of Statistical Inference*, Cambridge: Cambridge University Press, 1965.

——, *Representing and intervening: Introductory topics in the philosophy of natural science.*, Cambridge, England: Cambridge University Press, 1983.

——, "Telepathy: Origins of Randomization in Experimental Design," *Isis*, September 1988, *79* (3), 427–451.

——, *The Taming of Chance* number 124. In 'Ideas in Context.', Cambridge, England: Cambridge University Press, August 1990.

——, "A Tradition of Natural Kinds," *Philosophical Studies*, February 1991, *61* (1/2), 109–126.

——, *Rewriting the Soul: Multiple Personality and the Sciences of Memory*, Princeton, NJ: Princeton University Press, 1995.

——, *The Social Construction of What?*, Cambridge, MA: Harvard University Press, 2000.

Harford, Tim, "Odd numbers The man said to be America's most brilliant young economist is left cold by fiscal drag or monetary policy. He's more interested in aborted criminals and cheating sumo wrestlers," *Financial Times*, April 23 2005, *FT Weekend Magazine – Feature*, 25.

Harville, D. A., "Experimental Randomization: Who Needs It?," *American Statistician*, 1975, *29*, 27–31.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

____ , "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *Quarterly Journal of Economics*, February 2000, *115* (1), 45–97.

____ , "The Scientific Model of Causality," Unpublished Paper, University of Chicago, University College London, and the American Bar Foundation April 28 2005.

____ **and Edward Vytlacil**, "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, May 2005, *73* (3), 669–738.

____ **and Jeffrey A. Smith**, "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 1995, *9* (2), 85–110.

____ **and Richard Robb Jr.**, "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in H. Wainer, ed., *Drawing Inferences from Self–Selected Samples*, New York: Springer-Verlag, 1986.

Hilgartner, Stephen, "The Dominant View of Popularization: Conceptual Problems, Political Uses," *Social Studies of Science*, 1990, *20*, 519–539.

Hogben, Lancelot Thomas, *Mathematics for the Millions*, New York: W. W. Norton, 1968.

Holland, Paul W., "Statistics and Causal Inference," *Journal of the American Statistical Association*, December 1986, *81* (396), 945–960.

Horowitz, Joel L. and Charles F. Manski, "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, May 1998, *84* (1), 37–58.

Joyce, Theodore, "Did Legalized Abortion Lower Crime," *Journal of Human Resources*, 2004, *39* (1), 1–28.

____ , "Further Tests of Abortion and Crime," NBER Working Paper 10564, National Bureau of Economic Research, Cambridge, MA June 2004.

Kanouse, David E., Sandra H. Berry, Naihua Duan, Janet Lever, Sally Carson, Judith F. Perlman, and Barbara Levitan, "Drawing a Probability Sample of Female Street Prostitutes in Los Angeles County," *Journal of Sex Research*, February 1999, *36* (1), 45–51.

Katz, Lawrence, Steven D. Levitt, and Ellen Shustorovich, "Prison Conditions, Capital Punishment and Deterrence," *American Law and Economics Review*, 2003, *5* (2), 318–343.

Kennan, John, "Simultaneous Equations Bias in Disaggregated Econometric Models," *Review of Economic Studies*, January 1989, *56* (1), 151–156.

Kessler, Daniel and Steven D. Levitt, "Using Sentence Enhancements to Distinguish Between Deterrence and Incapacitation," *Journal of Law and Economics*, April 1999, *42* ((1, Part 2),), 343–363.

Landsburg, Steven E., "When Numbers Solve a Mystery," *Wall Street Journal*, April 13 2005.

Lee, David S. and Justin McCrary, "Crime, Punishment, and Myopia," NBER Working Paper 11491, National Bureau of Economic Research, Cambridge, MA June 2005.

Lee, Jennifer 8 (Eight), "Who Needs Giacomo? Bet on the Fortune Cookie," *The New York Times*, May 11 2005.

Levitt, Steven D., "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, June 1997, *87* (3), 270–290.

——, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Reply," *American Economic Review*, September 2002, *92* (4), 1244–1250.

——, "Back to the drawing board for our latest critics ... and also the Wall Street Journal and (Oops!) the Economist," http://www.freakonomics.com/blog/2005/12/05/back-to-the-drawing-board-for-our-latest-criticsand-also-the-wall-street-journal-and-oops-the-economist/ December 5 2005.

Lewis, H. Gregg, "Comments on Selectivity Biases in Wage Comparisons," *Journal of Political Economy*, 1974, *82* (6), 1145–1155.

Lillard, Lee A., "The Market for Sex: Street Prostitution in Los Angeles," Unpublished Manuscript, RAND September 1998.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, July 1993, *60* (3), 531–542.

McCrary, Justin, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment," *American Economic Review*, September 2002, *92* (4), 1236–1243.

Miller, George A. and Noam Chomsky, "Finitary Models of Language Users," in R.D. Luce, Robert R. Bush, and Eugene Galanter, eds., *Handbook of Mathematical Psychology*, Vol. 2, New York and London: Wiley and Sons, 1963, pp. 419–491.

Moffatt, Peter G. and Simon A. Peters, "Pricing Personal Services: An Empirical Study of Earnings in the UK Prostitution Industry," *Scottish Journal of Political Economy*, November 2004, *51* (5), 675–690.

Moffitt, Robert A., "Remarks on the Analysis of Causal Relationships in Population Research," *Demography*, 2005, p. Forthcoming.

Mussen, Paul, Mark R. Rosenzweig, Elliot Aronson, David Elkind, Seymour Feshbach, James Giewitz, Stephen E. Glickman, Bennet B. Murdock Jr., Michael Wertheimer, and Lewis O. Harvey Jr., *Psychology: An Introduction*, second ed., Lexington, MA: D. C. Heath & Co., 1977.

Nelson, Alan, "Are Economic Kinds Natural?," in C. Wade Savage, ed., *Scientific Theories*, Vol. 14 of *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, 1990, pp. 102–135.

Pearl, Jude, "The New Challenge: From a Century of Statistics to the Age of Causation," *Computing Science and Statistics*, 1997, *29*, 415–423.

Pickett, Joseph P. et. al, editors, "Rogue," in "The American Heritage Dictionary of the English Language," fourth ed., Boston: Houghton Mifflin, 2000.

Pinker, Steven, *The Language Instinct: How the Mind Creates Language*, New York: William Morrow and Company, 1994.

Pop-Eleches, Cristian, "The Impact of an Abortion Ban on Socio–Economic Outcomes of Children: Evidence from Romania," Unpublished Manuscript, Columbia University, New York November 2002.

——, Personal Communication September 2005.

——, "The Supply of Birth Control Methods, Education and Fertility: Evidence from Romania," Unpublished Manuscript, Columbia University, New York March 2005.

Raymo, Chet, "Growing Up with the BOMC," *Boston Globe*, July 15 1996, p. C2.

Reid, Sue Titus, *Crime and Criminology*, fourth ed., New York: Holt, Rinehart and Winston, 1985.

Reuter, Peter, Robert MacCoun, and Patrick Murphy, "Money from Crime: A Study of the Economics of Drug Dealing in Washington, D.C.," RAND Report R3894-RF, RAND, Santa Monica, CA June 1990.

Rowling, J. K., *Harry Potter and the Half-blood Prince*, New York: Arthur A. Levine Books, 2005.

Sakula, A., "A hundred years of lumbar puncture: 1891-1991," *Journal of the Royal College of Physicians of London*, April 1991, *25* (2), 171–175.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi–Experimental Designs for Generalized Causal Inference*, Boston: Houghton Mifflin Company, 2002.

Simon, John J., "Albert Einstein, Radical: A Political Profile," *Monthly Review*, May 2005, *57* (1).

Siskind, Frederic, "Minimum Wage Legislation in the United States: Comment," *Economic Inquiry*, January 1977, *15* (1), 135–138.

Skinner, B. F., *Verbal Behavior*, Englewood Cliffs, NJ: Prentice Hall, 1957.

Thacher, David, "Policing is Not a Treatment: Alternatives To The Medical Model of Police Research," *Journal of Research in Crime and Delinquency*, 2001, *38* (4), 387–415.

——, "From Racial Profiling to Racial Equality: Rethinking Equity in Police Stops and Searches," Working Paper 02–006, Gerald R. Ford School of Public Policy, Ann Arbor August 2002.

The Economist, "Oops–onomics; Economic Focus," *The Economist*, December 3 2005.

Tversky, Amos and Daniel Kahneman, "Judgement under uncertainty: Heuristics and biases," *Science*, September 27 1974, *185* (4157), 1124–1131.

Varian, Hal R., "Economic Scene," *New York Times*, April 11 2002, pp. 2, Section C.

Voltaire, *The History of Candid; or All for the Best*, Cooke's ed., London: C. Cooke, 1796. Translated from the French of M. Voltaire. Embellished with superb engravings.

Waxman, Sharon, "Sprinking Holy Water on the 'Da Vinci Code'," *New York Times*, August 7 2005.

Welch, Finis, "Minimum Wage Legislation in the United States," *Economic Inquiry*, September 1974, *12* (3), 285–318.

——, "Minimum Wage Legislation in the United States," in Orley Ashenfelter and James Blum, eds., *Evaluating the Labor Market Effects of Social Programs*, Princeton, NJ: Princeton University Press, 1976.

——, "Minimum Wage Legislation in the United States: Reply," *Economic Inquiry*, January 1977, *15* (1), 139–142.

Whitt, Hugh P., "Inventing Sociology: André–Michel Guerry and the *Essai sur la statistique morale de la France*," in Hugh P. Whitt and Victor W. Reinking, eds., *Essai sur la statistique moral de la France. A translation of Andre–Michel Guerry's Essay on the Moral Statistics of France: a sociological report to the French Academy of Science edited and translated by Hugh P. Whitt and Victor W. Reinking*, Studies in French Civilization, Lewiston, New York; Queenston, Ontario; Lampeter, Ceredigion, Wales: The Edwin Mellen Press, 2002.

Wikipedia, "Reinforcement— Wikipedia, the free encyclopedia," [Online: Accessed August 20, 2005] 2005.

Wilson, James Q. and Richard J. Herrnstein, *Crime and Human Nature*, New York: Simon and Schuster, 1985.

Yule, G. Undy, "An Investigation into the Causes of Changes in Pauperism in England, Chiefly During the Last Two Intercensal Decades (Part I.)," *Journal of the Royal Statistical Society*, June 1899, *62* (2), 249–295.