

The Problem of Big Data

AN A PRIORI EPISTEMOLOGICAL APPROACH TO TECHNOLOGICAL ADVANCEMENT

DUSTY WUNDERLICH

PRESENTED TO THE LIBERTARIAN SCHOLARS CONFERENCE
AND THE MISES INSTITUTE

Contents

Abstract	2
Epistemology Matters	3
Empirical Fantasies.....	7
Austrian Credit Cycle Theory.....	9
Subprime Mortgage Crisis – Mispriced Risk	11
Student Loans Bubble – Education Arbitrage	14
The Big Data Narrative.....	16
Moving the Goal Posts – Big Data Gymnastics	16
The Big Data War on Savers and Pensioners.....	17
The Future of Big Data.....	19
Big Data Discrimination – The Social Score.....	19
Dataism – The New Collectivism	22
Conclusion	25
References.....	26

Abstract

The third and fourth industrial revolutions created rapid advancement in technology while simultaneously creating a shift in epistemology. The advent of the personal computer and internet open the gateway for new opportunities in communication and disseminating information. For the first time in history computing power and data storage is allowing for the combination and analysis of large data sets. The promise of big data and artificial intelligence is one that will allegedly change humanity. However, is new technology and data creating new knowledge or are we wrapping ancient philosophical inquiries in a shiny technological wrapper.

For decades society has been promised smooth economies and prosperous environment using sophisticated modeling and econometrics. In practice it has driven more extreme business cycles the more advanced technology has become. Technology has not changed the underlying foundation of epistemology and the same problems we are facing now were addressed by many great minds during the enlightenment era and throughout Austrian economic thought.

As society shifts the epistemology pendulum further to empiricism history echoes the words of warning from Kant, Descartes, Mises, and Menger. The great rationalists understood the limitations of induction and scientific inquiry. If they were here today they would likely state that the problem of big data is the problem of induction and its natural conclusion will end in circular reasoning.

The stakes are higher now as we turn to technology and empiricism for societal solutions, technology stands to potentially compound societal issues. Humanity now has more information and justification of a partial picture to make claims which makes it inherently more dangerous. The war for power will be waged through big data and artificial intelligence, those wanting to assert centralized authority and dominance will use the illusion of technology to accomplish this task.

This paper explores the importance of epistemology as it relates to new technological advancement and the dangers of trusting technology to deliver new truth. The problem we face is not the technology, it's the epistemology and worldview behind the technology. Mises warned us of this problem in his great book the Epistemological Problems of Economics.

“Only a perfect being, whose omniscience and omnipresence would enable him to survey all the data and every causal relationship, could know how each erring human being would have to act at every moment if he wanted to possess the divine attribute of omniscience. If we were to attempt to distinguish rational action from irrational action, we should not only be setting ourselves up as a judge over the scales of value of our fellow men, but we should also be declaring our own knowledge to be the only correct, objective standard of knowledge.”

The issue at hand is the illusion of omniscience and omnipresence, there are no perfect all-knowing beings in society nor are there perfect all-knowing technologies. However, big data and artificial intelligence will be used as a proxy to be that perfect being and system. Only sound epistemology can defend against this intrusion on our liberty and human action.

We only need to look to the field of economics and the capital markets to see this problem in action. The 2008 mortgage crisis is a perfect example of the systemic failure of financial innovation and misguided epistemology.

Central bankers and hedge fund managers around the world took an a posteriori methodology to the markets, believing their models could measure and properly assess risk. More simply stated, they could control the markets.

Two core fundamental issues illustrate the limitations of big data and econometrics in this example. First, the use of econometrics to manipulate interest rates, Zero Interest Rate Policy (ZIRP), disconnects the market from time preference ordinary interest rates leading to a misallocation of resources. This results in an overemphasis of behavioral data for pricing risk, specifically, the consumer credit score. Trillions of dollars in capital are reliant on a posteriori knowledge causing exponential risk in the market due to malinvestment.

When supply of savings and time preferences do not drive interest rates in a marketplace there is no way to properly price risk on capital. Without the foundation of market interest rates, we turn to induction, an over emphasis on a posteriori data to price risk. Technology tries to mask the logical fallacies and explain away historical truths to maintain the status quo.

ZIRP and human behavior data like credit scores have illustrated the inefficiency of technology and misguided epistemology. Without sound theory and reasoned a priori thinking, society is bound to continue to chase the technological utopia.

“If the past, by bringing surprises, did not resemble the previous past to it (what I call the past’s past), then why should our future resemble our current past?” Nassim Taleb

Epistemology Matters

The promise of big data and artificial intelligence to change society is an exciting concept to behold. The Economist highlighted data as the most valuable commodity in the world on the cover of their magazine, replacing oil for the top spot. Bigger and bigger sets of data matched with greater computing power is bound to bring new discoveries and change society. Big data will bring society closer to the utopia we all want. Or will it?

Big players in the field of Economics and Econometrics are starting to position big data as the next generation of economic advancement. Hal Varian, Google Chief Economist, discusses new tools to manipulate and analyze large data sets in his recent publication. Stanford Economists, Liran Einav and Jonathan Levin, give a road map for big data and economic policy in their publication, The Data Revolution and Economic Analysis.

The popular leading economic minds are pointing to the darlings of Silicon Valley and their ability to gather and store larger and larger data sets. Furthermore, their technological abilities to take that data and analyze and manipulate for more precise econometrics. There is no doubt that the technology and data they are discussing is exciting but what is missing from all these giants of Economic thinking is a simple foundation of epistemology. There is a significant leap that is being made from

big data and new technology to enhanced knowledge. Unfortunately, in making this leap we are bypassing hundreds of years of epistemological thought.

Before we start envisioning technological utopias created from artificial intelligence and big data an important question must be addressed. Are we masking age old philosophical problems with a shiny new technological wrapper? Should society be looking to technologists and scientists for what the future unfolds or the great epistemology philosophers of the enlightenment era? Have we distanced ourselves so far from philosophy that we are now forming society devoid of a knowledge foundation?

The popular leading minds and modern literature are guides to understand the philosophy of our times and where it is likely going. In turning to literature about big data and economics we begin to unravel where institutions want to take big data. The literature written on big data and economics states that bigger data sets and technology will allow for greater modeling and prediction of economic activity. Varian points to the data collection power of Google as an indicator of importance.

“Due to the rise of computer-mediated transactions, many companies have found it necessary to develop systems to process billions of transactions per day. For example, according to Sullivan, Google has seen 30 trillion URLs, crawls over 20 billion of those a day, and answers 100 billion search queries a month” (Varian, 2014)

The above example and many others in modern literature addressing big data continue to point to the size of data as a basis for knowledge advancement. The assumption of bigger data being better must be addressed head on. Regardless of how many impressive examples economists and technologists give about big data there is no set of data nor system that is all knowing, omniscience and omnipresent. Therefore, this is just another impartial picture of a complex system that is ever changing and evolving. Ludwig Von Mises addressed this very issue in *Epistemological Problems of Economics*.

Statesmen, field marshals, and stock-market speculators act differently at present from the way in which they would act if they knew exactly all the data needed for an accurate judgement of conditions. Only a perfect being, whose omniscience and omnipresence would enable him to survey all the data and every causal relationship, could know how each erring human being would have to act at every moment if he wanted to possess the divine attribute of omniscience. If we were to attempt to distinguish rational action from irrational action, we should not only be setting ourselves up as a judge over the scales of value of our fellow men, but we should also be declaring our own knowledge to be the only correct, objective standard of knowledge. (Mises, *Epistemological Problems of Economics*, 1933)

Mises understood the complexity of individual valuation scales that are constantly changing. He knew that to truly understand human behavior and economics there had to be an epistemological foundation. The question then arises why leading economists and technologists are leaving out the

foundational importance of epistemology behind their theories of big data. Einav and Levin's publication give us great insight into possible motivations behind their big data philosophy.

One area of government activity where we could imagine such products is consumer protection. The key challenge in consumer protection is to keep individuals from making decisions they will (predictably) come to regret without proscribing individual choice. Behavioral economics has emphasized that one way to strike this balance is through the framing of decisions (e.g., well-chosen defaults), and another way is through the careful presentation of information. (Liran Einav, 2013)

The above statement illustrates the epistemological foundation being laid for the future of big data economists and econometrics. The use of empirical data to "help" guide consumer decisions is the basis of a posteriori knowledge, it is the use of data to attempt to predict future subjective valuations of individuals. Mises saw this much differently, and the process of valuation is best left to deductive methodology of the individual.

Since satisfaction and dissatisfaction depend only on the subjective view of the individual, there is no room for argument on this question in a science that does not presume to establish a scale of values or to make judgements of value. Its conception of an end, in the strict sense, is more deductive than empirical: ends are determined by the wishes and desires of the individual. Whenever reference is made to the greater or lesser appropriateness of means, this can only be from the point of view of the acting individual. (Mises, *Epistemological Problems of Economics*, 1933)

Mises was correct in recognizing that individual valuations are driven by wanting to achieve a desired end. Individual valuation scales are constantly changing and evolving. Therefore, the use of empirical data and induction will never be sufficient for accurate modeling of behavior.

David Hume understood this problem well when he documented the problem of induction in the 1700's. The issue becomes the presupposition that future events will occur like the past. The above example regarding consumer protection presupposes that empirical data regarding consumer regret will hold true with future consumer decisions. This is what Hume defined as uniformity of nature, similar effects come from similar causes. Hume rightly understood that this logical reasoning is circular and cannot hold true. Not to mention it is particularly impossible to pin causality and uniformity of nature with highly complex systems like human behavior.

Big data and artificial intelligence are nothing more than an extension of a posteriori and inductive reasoning. At its core the problem of big data is the problem of induction. For centuries we have used empirical evidence and scientific theory to advance society and fool ourselves into thinking that this is the full representation and path to truth. There is no denying induction has its place in society with the hard sciences or within highly controlled environments. However, applying that same methodology to complex systems and soft sciences is extremely flawed.

Many great minds have wrestled with the problem of induction over the centuries. There is no greater example of this old philosophical problem than the black swan theory. This ancient parable has roots back to second century poets yet still expanded on in modern times. Second century Roman Poet Juvenal stated the following before there was presumed a black swan in existence, “a rare bird in the land and very much like a black swan.” Centuries later British philosopher John Stuart Mill added to the black swan theory, “No amount of observations of white swans can allow the inference that all swans are white, but the observation of a single black swan is sufficient to refute that conclusion.” The proverbial black swan still exists today as it did in the 2nd century and it is now taking form in big data.

More recently the black swan theory has been expanded upon by derivatives trader and scholar, Nassim Taleb. Taleb dedicated a book to the black swan theory with an emphasis on exploring the impact of unpredictable and outlier events, specifically humanities blindness to their possibilities. He dedicates multiple chapters in his books to the flaws of induction, “the inability to predict outliers implies the inability to predict the course of history” (Taleb, *The Black Swan*, 2010).

Taleb’s humility comes from a strong epistemological foundation and because of his work in epistemology he understands the limitations of induction. Specifically, the complexity of human behavior and our inability to have all the available information. Taleb frames the complexity of human behavior brilliantly in this statement, “remember that you are a Black Swan” (Taleb, *The Black Swan*, 2010).

He goes further in stating the inherent issues in induction by addressing the fact we do not have all the available data, “the problem with experts is that they do not know what they do not know” (Taleb, *The Black Swan*, 2010). Taleb’s positions are a far cry from what we are seeing within academic institutions and leading technology companies. In our society we worship the expert and pay homage to technological advancements. Unfortunately, what’s missing in all this is sound epistemology, we are trying to build a plane in midair.

Another important addition to the problem of induction in recent years comes from quantum theorists. The work being done in this arena recognizes the problem of observation and its ability to effect the outcome. The observer’s effect is rarely discussed within the soft sciences but is a reality that quantum theorists are well aware of in their work within controlled environments.

Researches from Weizmann Institute of Science issued a paper stating their research supports the observer effect, “One of the most bizarre premises of quantum theory, which has long fascinated philosophers and physicists alike, states that by the very act of watching, the observer affects the observed reality” (Science, 1998). Although this research is specific to physical systems it illustrates the potential for error even in highly controlled environments.

The observer effect in the soft sciences is better known as the Hawthorne effect. It is well documented that subjects of data collection have a propensity to change behavior, this was originally discovered in the Hawthorne study in the 1920’s. In addressing the problem of big data the

observer's effect and bias must also be addressed, by the very nature of observation we are likely changing the observed reality which means that this could skew the collection and validity of data.

Intellectual honesty is the core shift being witnessed in today's intellectuals versus those during the enlightenment period. The explosion of induction and empiricism during the enlightenment era did not prohibit many of the intellects from being honest about its limitations. For example, empiricist philosophers and economists Davide Hume and John Stuart Mill were intellectually honest in knowing that empirical discipline would not work with economic science.

David Hume set the basis for this with his work on the Problem of Induction, "the supposition that the future resembles the past, is not founded on arguments of any kind, but is derived entirely from habit." (Hume, 2017). Mill later added onto Hume's work and understood specifically the limitations with induction and economics.

Mill argued in his Principles of Political Economy that economics was an "inexact and separate science," whose general principles were essentially known a priori and which held only subject to ceteris paribus clauses Mill's apriorism proved to be hugely influential in later economics. Lionel Robbins and the Austrian economists, such as Mises. (Hoover, 2006)

A deep history of thought from strong intellectual minds from many different camps of epistemology illustrate the limitations of induction and empirical data. It is not a leap to apply this same foundation and philosophical problem to modern technologies like big data and artificial intelligence. The premises are the same and that leads me to the core issues that need to be addressed around the Problem of Big Data.

1. The future does not always resemble the past;
2. All data is not available for observation;
3. Statistical and logical fallacies create different interpretations of data; and,
4. Observance can create disruption in collecting data.

Without being intellectually honest about the above problems of big data then we are bound to fall prey to its consequences. The next section will explore the results of epistemology in economics built on induction and a posteriori reasoning.

Empirical Fantasies

Big data is used daily throughout our lives and the impacts of induction can go largely unnoticed. As the use of big data increases in our society the opportunity for its impact in our lives will become more prevalent. This makes epistemology inherently valuable when assessing how big data will impact our lives for better or worse.

One area where big data and scoring is already highly prevalent is the capital markets. Specifically, the credit markets rely heavily on empirical data and scoring systems. The consumer credit score is a great example of the type of empirical data used to allocate capital. The credit score and how it is used is something worth exploring further.

There is currently \$3.9 Trillion of outstanding consumer debt and \$15 Trillion in outstanding mortgages in the United States. That is nearly \$20 Trillion of debt relying on credit scoring methodology, an empirical data set based on historical human behavior. I am not denying the predictive nature of the credit scoring system, however, its ability to properly assess and price risk is only as good as the epistemology behind its use.

Credit score is often used inductively to understand which consumers are the best credit risk. The empirical data from credit scoring is typically used to make observations and then identify patterns. The observations and patterns will eventually lead to a hypothesis. An example of a hypothesis would be consumers with 750 or higher credit scores are a good credit risk. This hypothesis would eventually lead to a theory. For example, consumers with a 750 credit score or higher have a high probability of paying their obligations fully and on time.

The immediate flaw in this credit score theory based on empirical evidence is that the theory does not take into effect all information for all time. The theory also errors in assuming past behavior predicts future behavior. As Hume stated, inductive theories would assume the uniformity of nature. The premise in this logical argument is based on the uniformity of nature and since we are dealing with unpredictable human behavior this is a flawed argument.

The reality is a consumer with a 750-credit score will not always be a 750-credit score for all of time and space, thus, the past does not always represent the future. No amount of big data can predict what a 750-credit score today will be in 12, 24, or 36 months from now. This creates an inherent logical flaw within the theory of consumer credit scoring. Credit scoring can be predictive at the moment it is applied but becomes null the moment time is introduced.

This subtle but important logical error can have exponential impact on millions of individual choices. The 550-credit score is likely to not get credit and without credit an individual can not improve credit creating a vicious cycle. However, the 550-credit scores behavior could easily be in line with a 750-credit score in the future and the same is true of the 750-credit score behavior being in line with a 550-credit score in the future. The limitations of this inductive system are creating economic social classes based purely on past behavioral data.

To properly assess a big data sample like the credit score it must be deduced from an a priori foundation. The Austrian Business Cycle Theory (ABCT) stands as a foundational a priori truth within Austrian Economics. Its eloquent and concise description of business cycle serves to guide epistemology on subsets of data like the credit score. Below is an explanation of the theory from the Mises Institute.

“The business cycle describes regularly occurring booms and busts observed in the economy and the Austrian business cycle theory is an explanation of this phenomenon from the Austrian School. Originally developed by Ludwig von Mises in the 1912 Theory of Money and Credit it was elaborated on by Hayek and others.

In one classical rendition: Banks expand credit well beyond their own assets and by the funds of their clients, often supported or encouraged by the setting of low interest rates by a central bank. This additional credit flow into the economy from increased borrowing for capital projects stimulates economic activity.

Projects which would not have been started before, seem now profitable, creating malinvestment. They increase demand for production materials and for labor and their prices rise, which, in turn, leads to an increase in prices of consumption goods. If the banks would stop the extension of credit, the boom would be rapidly over. To prevent the sudden halt of this boom (and the resulting collapse of prices), the banks must create more and more credit, and the prices will rise even more.” (Institute, Austrian Business Cycle Theory, 2016)

To properly assess historical data like the credit score we must first recognize the inherent limitations when applied to human behavior. This realization must be arrived at through an epistemology exercise like the above. Once the logical flaws are addressed and an a priori foundation is applied the information from the credit score can prove useful.

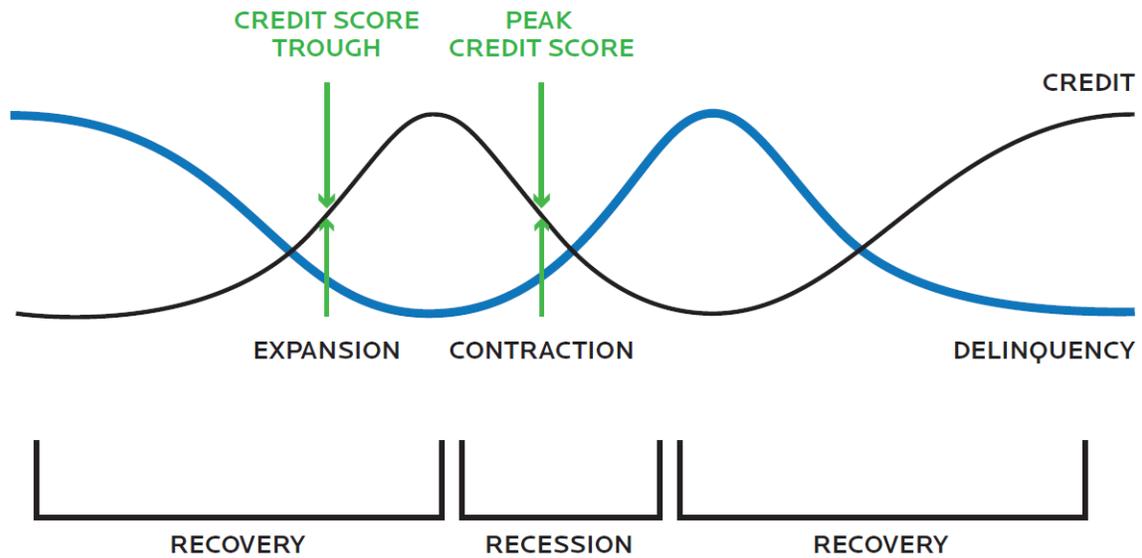
Austrian Credit Cycle Theory

The Austrian Credit Cycle Theory is an extension of the ABCT. Consumer credit underwriting and scoring is an empirical indicator of human action which means it is in constant motion and highly unpredictable. For that reason, the process of understanding credit scoring in the economy must be a deductive process. Credit cycle signals are deduced from the ABCT.

Austrian Credit Cycle Theory adheres to the expansion of credit through artificially low interest rates as a universal economic law that creates fluctuations in the business cycle and as a result in the consumer credit cycle. Since credit score is a human behavioral indicator it is unpredictable and can only follow, not lead, universal economic laws founded on human action. For that reason, credit scoring and delinquency lag the signals of the ABCT.

By establishing the a priori foundation of the ABCT the deductive process starts to illustrate that the credit score is less about payment behavior and more about the impact of ABCT on payment behavior. A spike in delinquencies does not arise independent of the ABCT, it is the direct result of the ABCT. Therefore, credit scores and delinquencies are a direct reflection of the stages of the ABCT.

The chart below illustrates the stages of the Austrian Credit Cycle Theory. When the economy is in a stage of recovery and credit expansion the credit score hits a trough at the early stage of the expansion. As the economy peaks and shifts into recession the credit score hits a peak at the early stage of the contraction. Delinquencies continue to climb into the early part of the new expansion.



During the ABCT the malinvestments from the credit expansion are exposed causing a correction. This starts a chain reaction for the economy. Consumers lose their jobs or their wages decrease and the lack of real savings creates the same exposure of malinvestment to consumers that businesses face. The empirical behavioral data that makes up the credit score is slow in responding and thus trails the stages of the ABCT. Credit scores are constantly in motion and the distribution is always shifting and changing in response to the business cycle. Without knowledge of this deduction the credit score is sending further misinformation into the market economy.

The credit score misleads the consumer no differently than the artificially low interest rates mislead the business owner. It is a misrepresentation of resources. Consumers are offered more credit during the expansion creating the illusion of excess resources (i.e. wages, savings). Credit is utilized by the consumer no differently than it is the businesses producing the goods and services.

The core issue with lending is the over reliance on behavioral data like the credit score. The primary driver of the credit markets starts and ends with the expansion of credit. As credit is expanding the positive feedback loop starts. Consumers are working off the new expansion and buying goods and paying their bills, thus, increasing their credit score. With no foresight to the nature of the ABCT this will inevitably end once the credit expansion stops. However, before malinvestment is exposed the consumer continues to get lower rates and more credit as their credit score improves.

The over reliance on credit score is a direct result of manipulation of the originary and market interest rates. When originary interest rates are properly set by discounting future goods against present goods there is less emphasis and need for behavioral data like the credit score. Without the

proper calculation of originary interest there is no way businesses nor consumers can accurately value goods and services. A true discount of future goods against present goods establishing accurate time preferences becomes a highly accurate depiction of resources and risk in the market economy.

One of the greatest examples of the problem of big data and empirical data is the 2008 subprime mortgage crisis.

Subprime Mortgage Crisis – Mispriced Risk

The subprime mortgage crisis was the combination of credit expansion and trusting empirical data, the credit score, to predict highly complex human behavior. The credit score methodology matched with artificially low interest rates drove intense demand for loans and credit expansion. Large tranches of consumer mortgages were securitized into pools primarily driven by consumer credit scoring. Every individual mortgage had its own set of complexities and that was compounded when they were pooled together.

The financial innovation of securitization and derivatives compounded the already faulty premise of their market theory. Even more pressure was put onto the credit scoring system. Securitizing pools of mortgages under the veil of diversification did not change the underlying epistemological errors, it only compounded the errors. Once credit expansion halted this exposed malinvestment and that 750-credit score that obtained a 4.5 percent fixed 30-year mortgage was now a 600-credit score and 100's of thousands of dollars under water on their mortgage.

Had a proper epistemological approach been taken to the credit markets this could have been avoided. By utilizing the a priori economic laws of the ABCT and applying the Austrian Credit Cycle Theory there could have been better foresight to the outcome. Sound theory would have driven the market perception rather than faulty data sets. Many Austrian economists utilizing the Austrian Business Cycle Theory predicted the 2008 mortgage crisis. By taking that one step further and overlaying the consumer credit score and delinquency rates a sound deductive process could have been in place to properly manage the credit expansion.

Unfortunately, nine years after the subprime mortgage crisis the real causes and issues are still largely missed. The regulators and legislators continue to make the same epistemological errors as the practitioners in the market. They are reacting to historical data and applying new protections and regulations based on that historical data. In this case, the most recent crisis. Nassim Taleb identifies this perfectly, "If the past, by bringing surprises, did not resemble the previous past to it (what I call the past's past), then why should our future resemble our current past?" (Taleb, Fooled By Randomness)

The measures to curb the damage and protect the economy will have no real impact and likely have further negative impact. We continue to treat the symptoms and not the virus. Banks lending to subprime borrowers was not the problem. Securitization was not the problem. Derivatives were not the problem. The problem was artificially low interest rates and credit expansion. Without the expansion of credit on and/or artificially low interest rates none of this is possible. The virus is central banking and the host is the financial markets. In this case the virus was strengthened by its

host through financial innovations which eventually led to an even worse demise. However, the host is not to blame though, it did not create the virus.

The central banking system was effectively telling borrowers that they were on a tight rope two feet off the ground when it was really 300 feet off the ground. This is the reality when interest rates are manipulated, they are inherently sending false information that allow individuals and businesses to take risks they normally would not take if they had correct information.

The over reliance on credit scoring due to misrepresented interest rates also exposed the systemic issues with induction in the credit markets. The mislabeled subprime mortgage crisis produced record foreclosures with prime borrowers. Wharton economists, Fernando Ferreira and Joseph Gyourko, conducted a study on foreclosures before, during, and after the subprime mortgage crisis. (Gyourko, 2015) Their study contradicted common knowledge that subprime lenders and borrowers were the core reason for the mortgage crisis. They describe their study findings below in an interview with Fortune.

“People have this idea that subprime took over, but that’s far from the truth. The vast majority of mortgages in the U.S. were still given to prime borrowers, which means that the real estate bubble was a phenomenon fueled mostly by creditworthy borrowers buying and selling homes they simply thought wouldn’t ever decrease in value.

We can draw two conclusions from this data. One is that your chances of being foreclosed upon in the past decade was more a matter of timing than anything else. If you were a subprime borrower in, for instance 2002, who bought a bigger house than a more prudent and creditworthy borrower would have bought, chances are you would have been fine. But a prime borrower who did everything right—bought a house he could easily afford, with a large down payment—but did so in 2006 would have had a higher chance of defaulting than the subprime borrower with better timing.” (Matthews, 2015)

According to their study, prime borrower foreclosures far outpaced subprime borrowers and recovery timing was also different. The subprime borrowers’ foreclosures peaked in 2008 while prime borrowers peaked two years later in 2010. All this data aligns with the ABCT and Austrian Credit Cycle Theory. This study did an excellent job outlining the limitation of the credit scoring system in this crisis by illustrating the specific timing and distribution of foreclosures and short sales. Credit score did not matter as it was always subordinate to the ABCT. It was going to be a lagging indicator to the movements of the ABCT. Without sound epistemology and theory lenders were blind to the fact that the 750 credit score they approved for a \$750,000 mortgage would foreclose in two years’ time. This became the black swan in our inductive hypothesis discussed earlier.

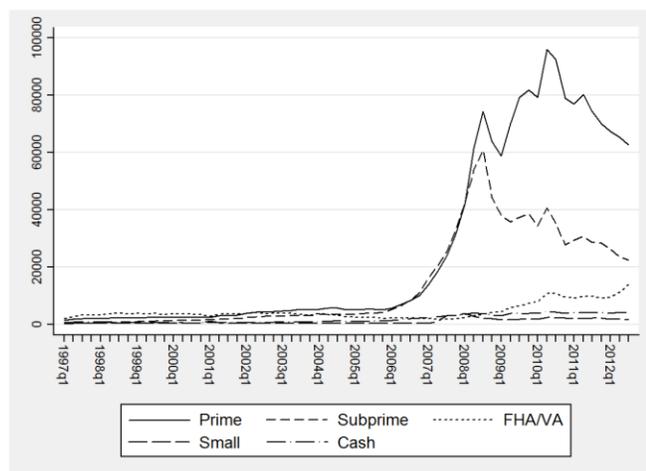


Figure 1: Total Foreclosures + Short Sales Over Time by Owner Type

Credit Score	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
300-499	6.6%	6.5%	7.1%	7.2%	7.3%	6.9%	6.3%	5.7%	5.6%	5.4%	4.9%	4.6%	4.7%
500-549	8.0%	8.0%	8.0%	8.2%	8.7%	9.0%	8.7%	8.5%	8.4%	8.1%	7.6%	7.1%	6.8%
550-599	9.0%	8.8%	8.7%	8.7%	9.1%	9.6%	9.9%	10.0%	9.9%	9.6%	9.4%	9.0%	8.5%
600-649	10.2%	10.2%	9.7%	9.6%	9.5%	9.5%	9.8%	10.1%	10.1%	10.2%	10.3%	10.3%	10.0%
650-699	12.8%	12.5%	12.1%	12.0%	12.0%	11.9%	12.1%	12.2%	12.2%	12.8%	13.1%	13.2%	13.2%
700-749	16.4%	16.3%	16.2%	16.0%	15.9%	15.7%	15.5%	16.0%	16.3%	16.4%	16.6%	16.9%	17.1%
750-799	20.1%	19.8%	19.8%	19.6%	19.3%	19.5%	19.6%	19.0%	18.9%	18.2%	18.2%	18.5%	19.0%
800-850	16.9%	17.9%	18.4%	18.7%	18.2%	17.9%	18.1%	18.5%	18.6%	19.3%	19.9%	20.4%	20.7%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Sub-Prime	46.6%	46.0%	45.6%	45.7%	46.6%	46.9%	46.8%	46.5%	46.2%	46.1%	45.3%	44.2%	43.2%
Prime	53.4%	54.0%	54.4%	54.3%	53.4%	53.1%	53.2%	53.5%	53.8%	53.9%	54.7%	55.8%	56.8%
Delinquency	1.55%	1.72%	2.55%	4.98%	9.00%	10.89%	10.41%	10.32%	9.00%	7.20%	5.63%	4.47%	3.70%
10/2 Spread	1.00%	0.20%	-0.12%	1.27%	1.71%	2.80%	2.78%	1.73%	1.65%	2.46%	1.33%	1.19%	0.55%

The above chart illustrates the constant motion and redistribution of credit scores over time. Specifically, this is a distribution of FICO credit scores (Dornhelm, 2018) over the last 12 years overlaid against mortgage delinquency and the 10/2 yield curve. (Reserve, 2018) The grey is the approximate timing of the recession. Prime credit scores peaked in 2008 during the recession and mortgage delinquency peaked in 2010 after the recession and into the early expansion. The yield curve is a good indicator for credit expansion and timing of the ABCT. Once the yield curve inverts this signals the contraction of credit and malinvestments are exposed. This typically leads to a recession and further contraction of credit.

As aforementioned, the Austrian Credit Cycle Theory adheres to credit and delinquency being subordinate to the ABCT. The movements in credit score and delinquency before, during and after the subprime mortgage crisis followed the ABCT and Austrian Credit Cycle Theory. This helps illustrate the power in a priori and deductive reasoning. Sound economics does not have to be overly complex or reliant on powerful algorithms, it needs sound epistemology and reasoning. What is important to note between the foreclosure and credit chart is the limitation of credit score to adequately predict human behavior.

By purely accidental circumstances, the subprime market depicts a healthier market. The credit score itself is not the reasoning. The higher interest rates that subprime borrowers are offered and the scarcity of capital for subprime borrowers illustrate a better reflection of resources and time preferences. Therefore, subprime borrower's behavior does not change drastically through the cycle. This was well illustrated in the subprime mortgage crisis. The subprime mortgage size grew but only by a few percent. When capital is scarce and market interest rates are accurately set it naturally balances the market by consumers making more informed decisions. The discriminatory credit system is in some ways protecting subprime borrowers but the narrative about the subprime could not be further off mark.

There is no denying that securitization and derivative products made this crisis much worse, but their sole purpose was the further expansion of credit. They created another layer of illusion around diversification and insurance that allowed for further justification of artificially low interest rates. Although the Wharton economists did recognize this was a matter of timing they did not conclude this tied to an a priori business cycle theory. Even though they correctly identified important nuances in the economy they missed the importance of epistemology in their findings.

The American people wanted a villain and that villain could not be the Federal Reserve. Instead, subprime lenders and borrowers became the convenient scapegoat in their place. This gave the

public a punching bag and allowed regulators and legislators the stage to play out their political theatre.

There is still not a realization of the faulty epistemology the entire system is working upon and without that realization the cycles will continue and most likely get worse. The virus goes undetected and will adapt and get stronger with time. We are now seeing this virus play out with the next generation in the form of student loans.

Student Loans Bubble – Education Arbitrage

The student loan market is another example of the limitations of induction in the credit markets. In lieu of credit score data the practitioners are utilizing job and salary data. The results and theory are the same, driving credit expansion through empirical data. The same epistemological issues apply with credit score as they do with employment and salary data being used to justify the student loan credit expansion.

The higher education credit expansion is largely based on misleading and/or incomplete data. Specifically, simple averages on employment, salaries, and student loan payback. This interpretation of the data is highly misleading due to the nuanced distribution of the data. It is also not uncommon to see entire subsets of data left out to prove a specific point. This incomplete and misleading empirical evidence leads to misinformation in the markets.

By getting an education the student is using means (loans) to achieve an ends (future gainful employment). The student leverages the means (loans) and/or current consumption of less gainful current employment to obtain greater means in the future. The artificially low student loan interest rates have done more to disrupt originary and market interest rates than any other category in the economy.

As a result, 50 percent of college graduates are in a job that does not require a four-year degree. (McGuinness, 2013) Furthermore, between 2010 and 2020 the Department of Labor predicts there will be 20 million new college graduates and only 7 million new jobs created that require a college degree. This means the majority of college graduates can obtain the same ends (employment) in the present moment without college, thus, leading to further malinvestment due to a misallocation of means. By applying the same market fixed interest rates to all degrees, the time preferences are false and making it seem like there are more means (gainful employment opportunities) available in the market than are.

Humanity is driven to gain means sooner rather than later which drives time preference and originary interest rates. We only sacrifice current consumption for the promise of greater future consumption, this is a praxeology axiom. Individuals would not prolong current consumption if they were not promised greater future consumption. This scenario only takes place when there is misinformation which is what we are currently witnessing in the higher education market.

Empirical data driven primarily from shallow interpretations or incomplete data is driving the justification for the student loan credit expansion. A simple empirical statement can drive intervention that creates unintended and unseen consequences for decades and sometimes generations.

For example, the following statement seems harmless but has a significant impact when taken as a literal prediction of the future. On average, students who go to college will make more money than those that do not go to college. For this reason, we need to make college available to more people through cheaper cost of borrowing. This simple observation has led to \$1.5 trillion in student debt with little to no improvement in earnings potential for most students.

Like the credit score, employment and education data are not static and subordinate to the movements of the ABCT. The labor and education markets will be entirely different five, ten, and twenty years from now which is why the discount of future to present goods is vital to setting time preferences. When this is not done then any empirical data being used will lag the ABCT and send further misinformation to the market.

A true discount of future goods against present goods has not been considered in higher education. This especially becomes a problem when college graduate supply outpaces the labor market supply which is what the market is currently experiencing. A simple empirical observation fixed at a certain point in time with specific variables can prove to be very dangerous when applied to future outcomes.

The results speak for itself, the cost of higher education has gone up over 1,300 percent in the last three decades primarily due to credit expansion. (Mislinski, 2016) However, the price premium which is increasing due to inflation is not properly offset by degree nor institution. Even though certain degrees and/or institutions will have radically different ends (salaries, job opportunities) the price premium is nearly the same across all education products.

This is driven by fixed market interest rates and leading to severe malinvestments. Unfortunately, this leaves many students with an ends (degree) that was not properly discounted. In the case where a future product is worth the same as a current product the discount and originary interest would be infinite but that is not being reflected in the current market interest rates. This has led to many students making decisions of malinvestment that they will carry throughout most of their adult life.

In looking at this problem from an epistemological perspective we must revisit the problems of big data laid out earlier in this paper. Specifically, the limitation of having all the data, the interpretation of the data, and the use of that data to predict future outcomes. In the case of higher education, we witness the impact of what a small deviation in interpretation of data can cause. The difference between an average and median or top quartile versus bottom quartile can be the difference in millions of people's lives. The compounding effect is significant over time.

The nuances in the data are not being discussed nor acknowledged in higher education. Specifically, not all degrees and institutions are created equal in the eyes of the labor markets. For example, an engineering graduate from the University of California, Berkeley can expect to be nearly \$1.1m better off after 20 years than someone who never went to college. In comparison, an arts graduate from Murray State University in Kentucky can expect to make \$147,000 less over 20 years than a high school graduate, after paying for their education. (King, 2014) This very example is not understood in the averages but worse is not reflected in the fixed rate student loans.

The other little secret in the data is the bottom quartile of performers in all categories up to a four-year degree are not out earning their peers that did not go to college. (Schrager, 2014) Those same

peers with less education now have higher home and auto ownership as well. The bottom quartile is subsidizing the top quartile and still half of the entire data set is not using the education for its intended use. Something an average from the 1990's could not have predicted.

What once started as an innocent and simple divulgence of data spun out of control. Without properly illustrating the risk of not finishing college or picking the wrong degree hundreds of thousands of individuals misallocated means. The very mechanism of ordinary interest rates is there to avoid these misinformed decisions and malinvestment. Had fixed interest rates been replaced by floating interest rates based on time preferences the supply of higher education would have naturally matched the supply of careers requiring a four-year degree. Simply put, resources would have been accurately allocated and price premiums would have shifted appropriately. The \$1.5 Trillion student debt crisis would be non-existent with a proper use of a priori epistemology.

The Big Data Narrative

Moving the Goal Posts – Big Data Gymnastics

When empirical data does not suffice for prediction which is inevitable in relation to human behavior the scientists go back to the drawing board to recreate a new magical set of data and explain away why it failed. Rather than addressing the logical errors and fallacies in the induction process the scientists go right back to same flawed process. The small nuances are implemented and back tested assuring users that the problems have been alleviated.

The credit score has a long history of big data folly. The FICO credit score has been adopted as the gold standard since 1989 but many don't realize that this score is in its 9th rendition since its inception. Other competing scores like the Vantage score have also emerged which is in its 4th rendition.

Each new scoring methodology is wrapped in shiny packing promising the utopia of all behavioral data. It does not matter how much back testing or stress testing is done it cannot account for present nor future data. This is the same futile exercise of the dog chasing its own tail. Further proof of the epistemological flaw of induction that past behavior is not always indicative of future behavior.

What inevitably happens is the data is used to fit a narrative or objective. For example, the Federal Housing Finance Agency (FHFA) is evaluating the approval of new credit scores outside FICO for government guaranteed loans. The following statement from the National Association of Realtors speaks to the objective.

“The National Association of Realtors is a strong supporter of utilizing newer, more predictive and inclusive credit scoring models, which we believe will responsibly expand access to mortgage credit and homeownership opportunities to more hardworking Americans, especially first-time borrowers and those who lack access to traditional forms of credit because of ‘thin’ credit files.” (Lea, 2018)

When the narrative needs to be adjusted from the perils of subprime lending to championing hard working Americans that are credit challenged it is accomplished through the impersonal use of big

data. It is these subtleties that get lost in an ocean of big data that has no philosophical foundation or theories.

The new Vantage 4.0 score promises all the new upgrades of big data, “less emphasis on medical collections, trended data, and machine learning.” (Clements, 2017) The objective remains the same, a justification of credit expansion and artificially low interest rates. The result also remains the same, the subordination to the ABCT.

The process and results are the same, an over reliance on empirical data for decision making. The very fact that data is being omitted or added on a regular basis illustrates the bias and weakness of the data to predict future outcomes. If the epistemological errors of data collection and representation are not challenged it leaves the door open for the data to be leveraged for nefarious reasoning.

Like the above example, data can be altered to fit a particular agenda without anyone taking notice to its true intentions or consequences. The French economist, Fredric Bastiat, knew the dangers of that which is seen, and that which is not seen. It is the unintended and unseen consequences that are not foreseen through sound theory create havoc in economies. One of the great unseen consequence of the big data process is the war on savers and pensioners.

The Big Data War on Savers and Pensioners

The third industrial revolution ushered in the perfect storm for the conflict between centralized and decentralized governance structures. The evolution of the internet and computer became the central focus for the ensuing conflict and an important inflexion point for society. Centralized governance structures could utilize this technology advancement for greater centralized control over economies and policy. In contrast, the decentralized power structures could leverage the new technology advancements to flatten governance structures through decentralization of information and commerce.

The first maneuver made by the centralized governance structure was decoupling the US dollar from gold in 1971. The internet combined with pure fiat currency open the door for a new frontier on monetary and political philosophy. The field of economics and monetary policy turned to the computer to grab power and resources. That power was wielded to control money and the economy through central banks and econometrics. The scientific management of the economy became the narrative and reality in the 1960's, many claimed the business cycle dead. Mark Thorton documents the folly of the 1960's well in the following statement.

“Okun was the chairman of President Nixon’s Counsel of Economic Advisors from 1968-1969. Right before the crash he described the economic expansion as “unparalleled, unprecedented, uninterrupted.” Okun believed that the economy was on a new dramatic departure from the past...After declaring the business cycle dead, he went on to demonstrate that research on the business cycle was now a thing of the past and that a “new” approach to the economy had replaced it” (Thorton, 2018)

The credit expansion of the 1960's led to unprecedented growth followed by crippling the economy through record inflation and eventually stagflation. The first experiment in utilizing computational power and scientific modeling for controlling economies was an abysmal failure. Okun's words about the business cycle being dead are still spoken today and the next iteration of this narrative is taking place via big data.

At its core the conflict is a war on savers and pensioners, anyone with capital through savings becomes the target of attack through centralized and scientific control of economies. The masses unknowingly get hit with the silent tax of inflation and their savings are devalued through the expansion of credit via artificially low interest rates.

The narrative often turns to ancient philosophical and religious wisdom claiming the harm of usury interest. This narrative is important in justifying the scientific management of the economy through artificially low interest rates. The masses become the victim in this narrative but what is conveniently missing from the narrative is the shift in creditors and debtors in the last two-hundred years.

There was a time when wealth was concentrated by a select few allowing for abuse of savings but those days have passed. With financial innovation over the past 200 hundred years the players in the capital markets have shifted. Creditors are everyday workers and small business owners, they are fire fighters and teachers who have pension plans and small business owners that have bank deposits and 401K plans. Mises recognized and articulated this well in Human Action.

“Public opinion has always been biased against creditors. It identifies creditors with the idle rich and debtors with the industrious poor. It abhors the former as ruthless exploiters and pities the latter as innocent victims of oppression. It considers government action designed to curtail the claims of the creditors as measures extremely beneficial to the immense majority at the expense of a small minority of hardboiled usurers. It did not notice at all that nineteenth-century capitalist innovations have wholly changed the composition of the classes of creditors and debtors. In the day of Solon the Athenian, of ancient Rome's agrarian laws, and of the Middle Ages, the creditors were by and large the rich and the debtors were the poor. But in the age of bonds and debentures, mortgage banks, savings banks, life insurance policies, and social security benefits, the masses of people with more moderate income are rather themselves creditors. On the other hand, the rich in their capacity as owners of common stock, of plants, farms, and real estate, are more often debtors than creditors. In asking for expropriation of creditors, the masses are unwittingly attacking their own particular interests. (Mises, Human Action)

The capital markets are distributed through the masses but consolidated into large institutional investors like pension funds. As a result of this depersonalization, many do not think of these institutional investors as the teacher or fireman down the street. This causes a costly disconnect in the markets and there is a dark reality in the global pension system not being addressed. Currently, the global pension system which is made up of the public employees, unions, and private employees is projected to be \$78 trillion underfunded, most of these liabilities go unnoticed since they are off balance sheet calculations. (Group)

Zero interest rate policy (ZIRP) and negative interest rate policies penalize savers and make it nearly impossible for savers to earn adequate returns. This means the very teacher, policeman and local union member that is being victimized as a debtor being taken advantage of is simultaneously being hurt by reducing the markets ability to drive returns for their savings. The state legislator who does not understand complex markets plays hero by sticking it to the predatory creditor while they ignore the gaping underfunded hole in their public employee pension system. The very process is short term thinking, we are cutting off our noses to spite our face.

As illustrated with the subprime mortgage and student loan crisis the interference and manipulation of ordinary and market interest rates has significant unseen and unintended consequences. When it comes to highly complex systems like human behavior and economies it is best to approach with caution and humility. Those in the realm of science studying complex systems and chaos theory understand this better than most. The slightest shift in variables can send a shock wave of changes down the line. Murray Rothbard spoke to the emergence of chaos theory later in his life.

The upshot of chaos theory is not that the real world is chaotic or in principle unpredictable or undetermined, but that in practice much of it is unpredictable. And in particular that mathematical tools such as the calculus, which assumes smooth surfaces and infinitesimally small steps, is deeply flawed in dealing with much of the real world. Chaos theory is even more challenging when applied to human events such as the workings of the stock market. (Rothbard, 2011)

The butterfly effect from chaos theory taught us that a butterfly flapping its wings in Brazil can cause a hurricane in Texas. Why? We have absolutely no idea why due to the unpredictable nature of causality in complex systems. For this reason, we must push back on the limitations of big data and inductive processes that can effect millions of individuals. No technology nor individual has the foresight nor capacity to truly understand this complexity.

If big data and its use is not challenged it will be leveraged to try to control economies and ultimately continue to hurt savers, creditors, debtors, and pensioners. Impartial and incomplete data will be used to fit within agendas. Without a sound understanding of epistemology behind technology the masses will be giving up their future for the sake of the present power struggle. Time preferences will continue to be miscalculated and over consumption and malinvestment will continue on the backs of hard-working savers and pensioners.

The Future of Big Data

Big Data Discrimination – The Social Score

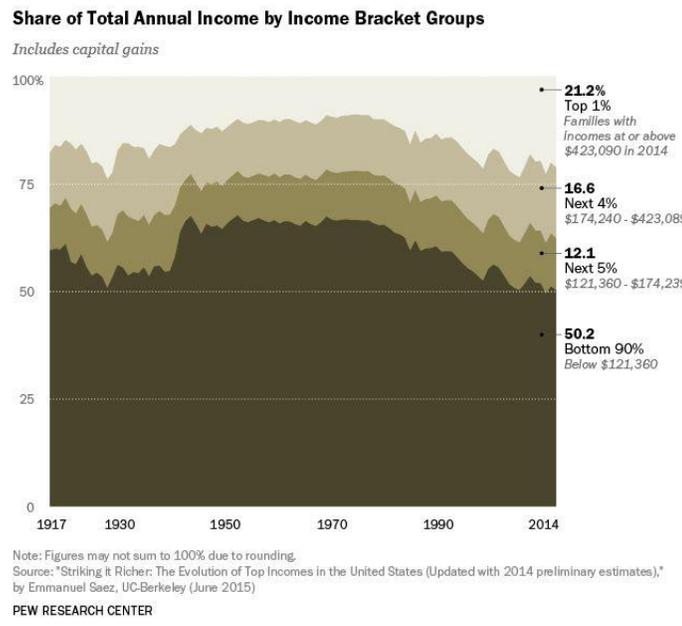
By its very nature the credit score is a discriminatory scoring system and should be viewed as a precursor of what is to come with future big data developments. The use of credit scoring is not inherently bad, and the score does have insightful and predictive components but when that data is used to suppress choice it becomes something entirely different.

Creditors want to evaluate their risk by knowing as much information as possible, but the use of credit score should begin and end there. However, when legislative and judicial systems utilize this behavioral data for wholesale discrimination against large subsets of a population it creates economic segregation and bias within economies.

There has been a 100-year attack on creditors and savers at the local, state and federal level. This attack has been leveraged through legislative, judicial, and regulatory action. The main premise has been built around usury and the limitation of interest rates. This narrative helps to justify zero interest rate policy and project savers and creditors in a negative light. It is an easy narrative to present to the populous and one in which can be villainized easily. Yet it stands as one of economics great ironies, the very basis of trying to control and suppress interest rates does far more harm than many would imagine.

Limiting a creditors ability to properly price risk within a market through repressive usury limitations effectively eliminates liquidity for a large subset of society. It also gives an elite group of society a competitive advantage over the rest by funneling scarce resources to that socioeconomic class first.

The Austrian Business Cycle Theory outlines the expansion of credit and its importance in the business cycle. Low interest rates ensure the expansion of capital in this process and components like the credit score ensure that this expansion of credit is allocated to a certain subset of society.



The individuals or groups of society that get low cost capital first benefit the most, they can profit early before price and asset inflation hit the marketplace. The Pew Research Center illustrated this point in the corresponding chart. The top 1 percent gain significant wealth over the bottom 99 percent during times of low interest rate environments.

Artificially low interest rates and expansion of credit hit the lower credit scoring class twice. Initially with the loss of cheap capital before prices rise and the loss of purchasing power through inflation. All of the above is accomplished through usury restrictions and credit scoring discrimination. This is a system of control and the credit scoring system is the justification for that system. The credit score becomes a reflection of who an individual is and how they will act in the future. It gives ultimate reign to positions of authority to project what they believe is best for the individual based on that score. It is far easier to depersonalize and discriminate against a number than it is an actual human being. This all played out in the global theatre during the subprime mortgage crisis. As we look to the future the question becomes how big data will be used to expand scoring model justifications beyond credit.

Companies like Facebook and Google are already looking to monetize their big data for behavioral scoring techniques. For example, Facebook was granted a patent to use their social networking platform for behavior decisioning like lending or email lists. The scores created by big technology companies will be sold as highly predictive indicators for many different outcomes but what is lost in this discussion is the epistemology behind this inductive process.

This will inherently go from a tool to be used for predictive aspects in the market economy to a tool that can be used to limit choice and direct resources to the elite few. Human history is paved by a path of humans using force to gain control over scarce resources. The medium in which it is accomplished shifts, but the objective remains the same and one of the newest weapons in that struggle is big data.

The credit score is just one way in which big data is used to discriminate and limit choice within our economy. When you take this to its logical conclusion one could imagine using big data scoring for many different avenues, some exciting and some frightening. Unfortunately, one of the worst-case scenarios is no longer an exercise of imagination, it is being played out with over one billion people in China.

The Chinese government has mandated that the entire population will be on their social credit scoring system by 2020. This is not an ordinary credit score, it goes well beyond just payment behavior. According to Brookings, "The system resembles an American credit score, but more than just low credit limits and high interest rates, a poor Chinese social credit score can lead to bans from travel, certain schools, luxury hotels, government positions, and even dating apps." (West, 2018) This is an example of a post big data world, one in which centralized private and public institutions can utilize individuals or groups behavioral data for or against them. All camouflaged in the mystery of big data yet permanently flawed by the errors of induction.

The credit scoring system in the United States has shifted and changed over the years and there are many different variations of credit scores. The question becomes who gets to decide what makes up a social credit score and why? And when that score changes and why? Over one billion people in this world are soon to be subject to a small group of individuals interpretation of empirical data about them. Keep in mind this is a small group of individuals that are not all knowing and riddled with biases and flaws like everyone. When you start to think about the false positives and false negatives that will occur in relation to humans with a social scoring system this discussion becomes sobering.

We learn by making choices and the consequences of those choices create an important feedback loop that becomes highly important to forming future human behavior. This process is often accelerated by those that are willing to take risks and try new approaches. In a world where social credit scores limit choices we are cut off from the very oxygen that drives humanity, it is a world bound to misallocate scarce resources and oppress millions of people. The Mises Institute eloquently states the vital importance of self-ownership and choice below.

“What a person chooses depends on how he values his opportunities. Since his valuations are both subjective and unobservable to other people, the decision-maker has superior knowledge about them. This enables him to make better decisions for himself than anyone else can. If a dictator "really" knew that a person would not like or benefit from 30 out of

the 50 opportunities, then by forcing the person to choose among the 20 remaining ones, he might (other things equal) make the person better off. But the dictator doesn't have this information. As a rule, another person, especially a distant person, an acquaintance, or someone who does not keep in constant touch with the decision-maker, cannot gauge the costs and benefits of that person's acts as well as the person making the choices.

Furthermore, when we realize that a great many of our acts connect with other acts of ours and of other people in a web that extends out in space and backwards and forwards in time, we see that it is practically impossible for someone else to make us better off by making our decisions for us. For example, we might have long-term plans to write the Great American novel. We decide to pass time traveling around and meeting Americans in what seems to be an aimless fashion. An outside observer cannot know what we are up to. Better to choose from 50 available opportunities than have someone else who supposedly knows better allow us to choose from 20 of these 50, or worse, select a choice for us." (Institute, Mises Institute, 2005)

The social credit score serves as an sobering reminder of what is possible when epistemology and sound reasoning is not addressed in a society. The result is a concentration of power and resources through mediums like big data. This new weapon can and will be used to eliminate free thinking and choice. Brookings briefly hit on the likely purpose of the social credit system, "the program's rollout may have materialized from a desire to stop potential political unrest and uphold the power of the regime." (West, 2018) This score is the illusion of fairness through the lens of technology and its purpose extends beyond individuals and looks to groups and associations. Botsman unveils more details behind the scores use.

"Posting dissenting political opinions or links mentioning Tiananmen Square has never been wise in China, but now it could directly hurt a citizen's rating. But here's the real kicker: a person's own score will also be affected by what their online friends say and do, beyond their own contact with them. If someone they are connected to online posts a negative comment, their own score will also be dragged down." (Botsman, 2017)

Without all past, present, and future data the use of incomplete data to judge and predict human behavior is bound to miss its mark. Regardless of whether that intention is genuine or not the process of utilizing big data or technology to limit choice should be vehemently challenged. That challenge must start from the foundation of epistemology and that ground must not be given up. Those brave challengers will be the difference between a world based on a global social credit score or one that respects individual choices and values.

Dataism – The New Collectivism

The emergence of big data has brought new philosophy, and some would argue religion to the forefront of thinking. Dataism is the new foundation for this thinking and its forefathers are quickly making a case for the importance and transformative power of big data and free flow information. This new thinking is in its infancy, but the founding literature is quite clear in illustrating the path this will go.

Dataism: The future evolution of technology wherein data become more valuable than humans. Dataism goes beyond techno-humanism, which envisions technology making people smarter. Dataism implies that the data are more important and that without the data, people become helpless and somewhat useless. (Magazine, n.d.)

There is no greater advocate for this new philosophy than Israeli historian Yuval Noah Harari. One of his most recent books, *Homo Deus*, gives a detailed account of Dataism and its likelihood of taking over humanity. Yet, Harari calls for unaltered freedom of information to create a system he coins as the internet-of-all-things.

“If life is the movement of information, and if we think that life is good, it follows that we should deepen and broaden the flow of information in the universe. According to Dataism, human experiences are not sacred and *Homo Sapiens* aren’t the apex of creation or a precursor of some future *Homo Deus*. Humans are merely tools for creating the internet-of-all-things, which may eventually spread out from Planet Earth to pervade the whole galaxy and even the whole universe. This cosmic data processing system would be like God. It will be everywhere and will control everything, and humans are destined to merge into it.”
(Harari, 2017)

Harari goes into detail outlining that human experiences have no value, we are purely bits of information in a system. Our only meaning is giving information to that system. It is a full dismissal of the importance of billions of individual subjective valuations and their pursuit for desired ends. In the view of the Dataist, this is a meaningless exercise if we do not share that information with the processing system. Harari is bold in making an objective claim on what represents value for billions of humans.

In one sentence Harari claims life is good and follows that by defining what is best for billions of individuals. The plea to the greater good of the internet-of-all-things is eerily similar to statements made in the past by fascist and communist minds. This religion is quickly evolving and developing guiding commandments rooted in the gathering all information. What’s most disturbing is that information should be gathered regardless of it being voluntary or not.

“Dataism isn’t limited to idle prophesies. Like every religion, it has its practical commandments. First and foremost a Dataist is out to maximize data flow by connecting more and more media, and producing and consuming more and more information. Like other successful religions, Dataism is also missionary. Its second commandment is to link everything to the system, including heretics who do not want to be plugged in. And “everything” means more than just humans. It means everything. Our bodies, of course, but also cars in the street, refrigerators in the kitchens, chickens in their coops and trees in the jungle – all should be connected to the internet-of-all-things.” (Harari, 2017)

This is a continuation of the China social score and it is a future that will sacrifice human rights for the acquisition of information. Harari even goes a step further to ensure that his audience does not confuse our human right to freedom of expression with freedom of information.

“We mustn’t confuse freedom of information with the old liberal value of freedom of expression. Freedom of expression was given to humans and protected their right to think

and say what they wished. Freedom of information, in contrast, is not given to humans. It is given to information.” (Harari, 2017)

Freedom of information is a powerful concept and it is easy to romanticize about its possibility. However, once information starts to take precedent over human rights and natural law it must be viewed with great caution. Harari is framing a case for freedom of information against a strong history of censorship. It’s an emotional appeal and it takes a little deeper dive to understand Dataism’s true intention behind the veil of free flowing information.

“People just want to be part of the data flow, even if that means giving up their privacy, their autonomy and their individuality. Humanists art sanctifies the individual genius, so a Picasso doodle on a napkin nets million at Sotheby’s... But a growing number of artistic and scientific creations are nowadays produced by the ceaseless collaboration of everyone. Who writes Wikipedia? All of us.” (Harari, 2017)

Free flow of information and the internet-of-all-things by everyone, willing and unwilling, is the new collectivism wrapped in a shiny new technology package. Dataism is going at the heart of property rights and making a case that information supersedes individual human property rights. It is making a new case for shared property through information and claiming that it is best for humanity even though it will likely end humanity. This claim for the consolidation of resources for the greater good is an old appeal being recycled as it has many times in the past.

“If humankind is indeed a single data-processing system, what is its output? Dataists would say that its output will be the creation of a new and even more efficient data-processing system, call the internet-of-all-things. Once this mission is accomplished, Homo sapiens will vanish.” (Harari, 2017)

What continues to go missing in the discussions around Big Data and Dataism is epistemology. The literature and thinking reveals its own inherent flaws in epistemology. The believers are quick to claim the all-knowing and all-powerful nature of Dataism yet subtly insert its inability to be deterministic. Harari goes into great detail to make a case for Dataism and its ability to control the world and our lives yet fully acknowledges its inability to predict the future. In one section Harari labels Dataism as all-knowing and all-powerful while in the same chapter states it is not deterministic.

“We cannot really predict the future, because technology is not deterministic. The same technology could create very different kinds of societies...The rise of AI and biotechnology will certainly transform the world, but it does not mandate single deterministic outcomes.” (Harari, 2017)

The Economist labeled data as the most valuable commodity in the world and there is no question that information and data are powerful resources that should be used to satisfy our ends. However, this resource is not all-knowing nor all-powerful. Like all scarce resources, humanity has been fighting or cooperating to utilize them for their respective ends for the test of time. Unfortunately, resources can be used for the desired ends of power and I think much of what we are seeing with Big Data and Dataism is about controlling the most important resource in the fourth industrial revolution.

As a historian, Harari, should know the struggle for scarce resource better than most but he chose to take a different approach in how he presented the future. He chose a future that celebrates shared property regardless if that is voluntary or involuntary. A future where we should freely give up all our information to companies like Google and Facebook so that we can better serve society. A future where algorithms tell humanity what to do but those very algorithms and companies are controlled and created by humans with biases. Algorithms with incomplete and impartial information will be directing humanity on what job they should take, who they should marry, and what worldview they should have.

The new world of Dataism will be led by supposed all-knowing and all-powerful algorithms that cannot predict the future. Yet, these algorithms will tell us what's best for us in exchange for all our information and anonymity. In doing that we should find our true purpose and meaning by freely giving our information to large private and public conglomerates. And when that is all done according to Harari we need to answer the following question.

“What will happen to society, politics and daily life when non-conscious but highly intelligent algorithms know us better than we know ourselves?” (Harari, 2017)

My answer to this question is that it will likely be similar to what humanity went through during the terrible 20th century. However, this time rather than authoritarian Fascist and Communist leaders it will be algorithms taking their place. Its straight forward for humanity to attack a charismatic political leader gone array but how does a society attack an unconscious and impersonal algorithm when free flow of information goes array?

Regardless of the promise of Dataism, algorithms will never know us better than we know ourselves for the very simple epistemological errors outlined in this paper. The danger is believing the statement to be true and freely giving up our rightful individual property, our private information, into a centralized system(s). Dataism is nothing more than the same narrative of collectivism which thrives to control scarce resources through a small elite group of individuals. The resources and technology have changed but the philosophy and epistemology have not.

Conclusion

Big data is nothing more than another empirical tool, a tool that can be wielded in many ways like any tool. It is not the purpose of this paper to define how this tool should be used but only to outline the epistemological and philosophical under pinning's of this tool. For the same reason we do not use a screw driver to hammer in a nail is the same reason we should be applying reasoned and sound outcomes for the intended use of technological tools.

Even though we do not use screwdrivers to hammer nails for sound reasons it does not prevent nor should it prevent others from using this tool for that purpose. We must think of big data in the same manner, the tool itself is not wrong or right but how the user leverages that tool is something we should be aware of and understand.

Awareness becomes key in understanding whether use is implicitly impacting individuals' actions and liberties. The line must be drawn when the screwdriver goes from hammering nails to being used as a weapon to harm others. The same is true of big data as a technological tool when that line

is crossed. The key to knowing when that line has been crossed starts and ends with sound epistemology.

As governments and technology companies start to aggregate and control larger sets of data we must have challengers in our society that are willing to blow the whistle on the misuse of technological tools. In an era where individuals willingly give away their rights to privacy for convenience we must be aware that this might be a path that we cannot come back from. This naivety towards giving up personal data also risks impacting other individuals in the process that do not wish to cross that line. Understanding the tool is vital in our battle against the misuse of technological tools.

Ludwig Von Mises and many of the great minds from Austrian economic and libertarian thought understood the importance of epistemology and the power of deductive reasoning. Their ground work in theory applies now as it did then, it is our job to take that foundation and interpret it against the shifts in culture and technology. Theory and a priori truths do not change with technology and culture and for that reason we must start with a foundation and understanding of knowledge regardless of technological or cultural changes otherwise we are bound to fall prey to collectivism and authoritarian regimes.

References

- Botsman, R. (2017, October 21). *Big data meets Big Brother as China moves to rate its citizens*. Retrieved from Wired: <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>
- Clements, N. (2017, March 31). *3 Big Changes To Credit Scores That Will Impact Your Wallet*. Retrieved from Fortune : <https://www.forbes.com/sites/nickclements/2017/03/31/3-big-changes-to-credit-scores-that-will-impact-your-wallet/#7ed560a75a67>
- Dornhelm, E. (2018, September 24). *US Average FICO Score Hits 700: A Milestone for Consumers*. Retrieved from FICO: <https://www.fico.com/blogs/risk-compliance/us-average-fico-score-hits-700-a-milestone-for-consumers/>
- Group, C. (n.d). *Pension System*. Citi Group.
- Gyourko, F. F. (2015). A NEW LOOK AT THE U.S. FORECLOSURE CRISIS. *National Bureau of Economic Research*.
- Harari, Y. N. (2017). *Homo Deus*. Harper Collins.
- Hoover, K. D. (2006). *Causality in Economics and Econometrics*. Durham: Duke University.
- Hume, D. (2017). *A Treatise of Human Nature*. Pinnacle Press.
- Institute, M. (2005, September 6). *Mises Institute* . Retrieved from The Economics of Self Ownership: <https://mises.org/library/economics-self-ownership>
- Institute, M. (2016, October 2). *Austrian Business Cycle Theory*. Retrieved from Mises Wiki: https://wiki.mises.org/wiki/Austrian_business_cycle_theory

- King, K. (2014, April 4). *Is College Worth It?* Retrieved from The Economist:
<https://blog.ecu.edu/sites/dailyclips/blog/2014/04/04/is-college-worth-it-the-economist/>
- Lea, B. D. (2018, January 8). *Credit score changes in mortgage market could increase homeownership opportunities.* Retrieved from Fox Business : <https://www.foxbusiness.com/markets/credit-score-changes-in-mortgage-market-could-increase-homeownership-opportunities>
- Liran Einav, J. L. (2013). The Data Revolution and Economic Analysis. *NBER Innovation Policy and the Economy Conference*, 12.
- Magazine, P. (n.d.). *Dataism*. Retrieved from PC Mag:
<https://www.pcmag.com/encyclopedia/term/69512/dataism>
- Matthews, C. (2015, June 17). *The Subprime Mortgage Crisis Wasn't About Subprime Mortgages.* Retrieved from Fortune : <http://fortune.com/2015/06/17/subprime-mortgage-recession/>
- McGuinness, W. (2013, January 29). *Half Of Recent College Grads Work Jobs That Don't Require A Degree: Report.* Retrieved from HuffPost:
https://www.huffingtonpost.com/2013/01/29/underemployed-overeducated_n_2568203.html
- Mises, L. V. (1933). *Epistemological Problems of Economics*. Ludwig Von Mises Institute .
- Mises, L. V. (n.d.). *Human Action*. Mises Institute .
- Mislinski, J. (2016, December 31). *Brace Yourself: Our Latest Look at Student Debt.* Retrieved from Advisor Perspectives:
<https://www.advisorperspectives.com/dshort/commentaries/2016/12/31/brace-yourself-our-latest-look-at-student-debt>
- Reserve, F. (2018, October). *10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity.* Retrieved from FRED: <https://fred.stlouisfed.org/series/T10Y2Y/>
- Rothbard, M. (2011, June 5). *Chaos Theory.* Retrieved from Mises Canada :
<https://www.mises.ca/chaos-theory/>
- Schrager, A. (2014, July 15). *College Graduates Don't Always Outearn High School Grads.* Retrieved from Bloomberg: <https://www.bloomberg.com/news/articles/2014-07-15/is-college-worth-it-some-high-school-grads-earn-more>
- Science, W. I. (1998). Quantum Theory Demonstrated: Observation Affects Reality. *Nature*, 871-874.
- Taleb, N. (2010). *The Black Swan*. Random House .
- Taleb, N. (n.d.). *Fooled By Randomness*.
- Thorton, M. (2018). *The Skyscraper Curse: And How Austrian Economics Predicted Every Major Economic Crisis of the Last Century*. Auburn: Mises Institute .
- Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives* , 4.

West, J. K. (2018, June 18). *China's social credit system spreads to more daily transactions*. Retrieved from Brookings: <https://www.brookings.edu/blog/techtank/2018/06/18/chinas-social-credit-system-spreads-to-more-daily-transactions/>