Remedies Against Bias In Analytics Systems

Abstract

Advances in IT offer the possibility to develop ever more complex predictive and prescriptive systems based on analytics. Organizations are beginning to rely on the outputs from these systems without inspecting them, especially if they are embedded in the organization's operational systems. This reliance could be misplaced if the systems contain bias. This could even be at best unethical and at worst illegal. Data, algorithms and machine learning methods are all potentially subject to bias.

In this article we explain the ways in which bias might arise in analytics systems, present some examples where this has happened, and give some suggestions as to how to prevent or reduce it.

We use a framework inspired by the work of Hammond, Keeney and Raiffa (1998, reprinted 2006) on psychological traps in human decision-making. Each of these traps "translates" into a potential type of bias for an analytics-based system.

Fortunately, this means that remedies to reduce bias in human decision-making also translate into potential remedies for algorithmic systems.

Keywords: algorithms; artificial intelligence; bias; decision-making; psychological traps

Word count: 11,388

Introduction

Forecasts and predictions are central to running any kind of organization, whether in business or the public sector. The availability of big data - indeed, ever bigger data - and new analytics techniques enable a greater range of approaches to forecasting and prediction than ever before. Modern information technologies also deliver the resulting predictions far more rapidly. However, in order to be useful, it is essential that these predictions are sufficiently accurate or exact. There are two different

dimensions to accuracy/exactness: precision and bias (Edwards and Finlay, 1997). Precision recognizes the reality that even the best predictions will hardly ever be 100% accurate (except when predicting events with yes/no outcomes). So, some degree of error is unavoidable. Precision is a measure of this random uncertainty, or random error. Bias, on the other hand, is a systematic error. Dictionaries offer two definitions of bias: "the fact of preferring a particular subject or thing" and "the action of supporting or opposing a particular person or thing in an unfair way" (Cambridge Dictionary, 2018). These show that bias may be accidental, or it may be deliberate.

Bias can appear, for example, as a result of incorrect use or interpretation of metrics, from lack of analysis of the distribution of the variables, or from an inappropriate design of experiments and sampling process. People's first choice for describing the behaviour of a variable is generally to use the arithmetic mean/average. However, in many cases a different form of average is needed, such as the geometric mean, mode or median. It is also possible that the sample that is used is not representative: for example, traffic and transport patterns are very different during school vacations. In the world of big data, it is easy to think that simply having more data is the solution for understanding the phenomenon. However, while the data may be big in volume, it may all be from the same source and/or have the same type of intention/orientation, as can happen when social media are affected by the strategies of communication used.

Deliberate bias is usually linked with the "unfair" part of the dictionary definition, but it is not necessarily a bad thing: monitoring systems in chemical plants that give warnings, using analytics techniques, are typically designed to alert staff before the issue reaches critically dangerous levels; while in many countries it is a legal requirement that the speed indicated by an automobile's speedometer must never

be less than the actual speed, but may be greater. Equally, in the decades when election opinion polls were most commonly carried out by telephone (roughly the 1950s to the 2000s), it was standard practice to adjust the raw results to reflect the fact that supporters of certain political parties, such as the Democrats in the USA and Labour in the UK, were known to be less likely to have a telephone than supporters of other parties.

In far more contexts, however, bias is problematic, and the effects of algorithmic bias can be very complex. Teachers in Houston, Texas, USA have challenged the legality of the use of Education Value-Added Assessment System (EVAAS) scores to judge individual teachers (including dismissing them from their positions), on the grounds that teachers had no way to ensure that EVAAS was correctly calculating their performance score, nor was there a way to promptly correct a mistake (Amrein-Beardsley, 2017). The latter point has much wider significance: because of the comparative nature of the method used in EVAAS, any error (in data or in the algorithm) in calculating the score for one teacher would potentially affect the scores for all other teachers in that district.

Zhao, Wang, Yatskar, Ordonez, & Chang (2017) give an example of gender bias in the imSitu visual semantic role labelling dataset, which is widely used in research on classification of images. Only 33% of the people in pictures in the dataset related to the activity "cooking" were male. However, a predictive analytics system developed with a training set from this dataset classified the gender of just 16% of people in the test dataset as male. The data (the sample of photographs) were biased, but the algorithm amplified that bias.

Facial recognition software is currently accused of bias, especially against people of colour. The American Civil Liberties Union carried out a test that matched

28 members of the US Congress with mugshots on a database of criminals, including a disproportionate number of people of colour (Snow, 2018).

Alternatively, the facial recognition software may not be showing bias, but just incompetence - lack of precision, in our terms. In two tests in the UK, the Metropolitan Police (covering most of London) identified 104 people with the aid of the software, but only 2 of them were actually on their wanted lists. The South Wales Police actually went on to intervene in 46 cases, even though 31 of them turned out to be innocent people (Fearn, 2018).

The safety monitoring systems mentioned earlier are an example of conscious bias; when this type of bias is present, people using the system know about it. To be precise, *some* people must know about it; we will return to this point later. Our main concern in this article is with bias that people do not know about. This may be because it is unconscious bias, or because the people who consciously put the bias in the system have not told the people who need to know about it.

Prior work on bias in decision making by analytics systems has concentrated almost exclusively on either descriptive accounts, and/or the "black box" issue of lack of explanation (O'Neil, 2016), and/or treating bias as a technical issue. We could not find any work that used the extensive prior literature on bias in human decision making. The main contribution of this paper is to translate that research into the analytics context. We will be looking at sources of bias, how to avoid or prevent bias, how to identify bias, and how to compensate for bias. After a short section on definitions, we first present a list of the major sources of bias in decision-making, linking them to the phases in analytics system development. We then go on to offer some remedies to reduce or prevent bias, first at the strategic level (applying to all

analytics work in the organization) and then at the level of the individual analytics project or study. We close the article with our concluding remarks.

Descriptive, predictive and prescriptive systems

A commonly-used categorization proposed by Robinson, Levis, & Bennett (2010) divides analytics systems into three types: descriptive, predictive and prescriptive.

Descriptive analytics systems answer the questions "what has happened?" and "what is happening?"

Predictive analytics systems answer the question "what could happen in the future?" Note that this is carefully worded: *could* happen rather than *will* happen.

Prescriptive analytics systems answer the question "what should the organization do?"

The difference between predictive and prescriptive analytics systems may be more a result of how the organization chooses to fit the analytics systems' outputs into its business processes rather than a property of the analytics techniques themselves. Thus from the data scientist's perspective, there may be little difference between the two types.

We will concentrate more on predictive and prescriptive analytics systems in later sections. These two categories have a more direct link to action. The main function of descriptive systems is typically monitoring, so in analytics terms they help determine that action is needed, but not what that action should be. Nevertheless, it is important to guard *all* categories of analytics system against bias. It might be thought that descriptive analytics systems would not suffer from bias, but this is not the case; as we will see in later sections, even raw data can be biased.

Phases in the analytics system development process

In what follows, we will use one of the most highly-cited versions of the analytics process (Jagadish et al., 2014). Their "pipeline", as they call it, has five phases: acquisition (including filtering); information extraction and cleaning (technical and formatting issues); data integration (matching sources, choices on how to store data - for people and applications); modelling and analysis; and interpretation (including checking assumptions). We will abbreviate these to acquisition, extraction, integration, analysis and interpretation respectively.

Jagadish, et al. (2014) point out that "Many people unfortunately focus just on the analysis/modeling step" (p.88). As will be seen, this article offers a different view. All the sources of bias we identify potentially concern the acquisition and interpretation phases, and together with analysis these phases account for all but one of our examples.

Psychological traps in human decision-making

Bias can be considered from two different perspectives: statistical/mathematical and human/psychological. We regard these as complementary. Theoretical and practical understanding of decision making comes from studying what human decision makers do, while our understanding of what analytics techniques and systems do is essentially statistical/mathematical. Thus the former provides the context for the latter, and therefore the psychological aspects of decision-making are where we will start. Our discussion in this section is based on the framework of Hammond, Keeney, & Raiffa (1998). They identify the six most common types of psychological traps in decision-making that can be sources of bias. Other authors present even longer lists; for example, Lebowitz and Lee (2015) came up with 20. We choose to use that of

Hammond, Keeney, and Raiffa because they have been researching human decision-making for several decades and their work is the most cited of its kind.

For each of the psychological traps we identify the resulting bias, present brief examples of data analytics practice that materializes that bias and drives the organization to decisions that are possibly inappropriate, and indicate which of the five phases of the process are involved, with an annotation such as *(interpretation)*. A major source of bias is making an inappropriate choice about whether models are used in a deterministic or stochastic manner. The viewpoint of this article is that either may be appropriate, and that the analytics professional must decide how to tackle a particular problem in order to yield the least biased and most practicable solutions.

Sources of bias in analytics systems

It tends to be assumed that human decision makers are the best models of decision making that we can access, simply because of the lack of alternatives; who (or what) else could we use? Yet research shows that human decision makers have their limitations. The most basic of these is that humans do not always look for the best (optimal) decision. Instead, as Herb Simon (1969) explained, we *satisfice* - we seek a decision that is "good enough", and when we find one, we stop looking and go and implement that decision. This is also known as the theory of bounded rationality. The idea that humans do not make completely rational decisions is still a contentious one, especially amongst economists, even though Simon won the 1978 Nobel Prize for Economics for his work. We do not have space to go into the arguments here. Bounded rationality of human decision-makers is, however, widely accepted by both scholars and practitioners in business and management. Ironically, it is often more difficult to devise and programme an analytics algorithm to find a satisficing solution

than it is to produce an algorithm to find an optimal one.

Analytics systems do have some advantages over human decision makers when it comes to bias. The detail and completeness of data in analytics systems should mean that hindsight bias - believing past predictions and forecasts were closer to what really happened than they actually were - is not an issue. Unlike human decision makers, where recent research has shown that hindsight bias can even exist when we do not know exactly what the real outcome was (von der Beck, Cress, & Oeberst, 2018). However, many other sources of bias are common to both humans and analytics systems.

The notion of humans as imperfect decision makers is also at the heart of Keeney and Raiffa's work, including their framework based on six types of decision-making traps. These traps, slightly modified, give us the sub-headings for the remainder of this section, as shown in Table 1.

TAKE IN TABLE 1 ABOUT HERE

Bias towards the first information received

This source of bias is an example of a phenomenon known as anchoring. People tend to be overly influenced by the first relevant information that comes to them about an issue. This information then *anchors* all their subsequent thinking about the issue, even affecting whether or not they seek any more information at all.

Anchors come in many forms. They might be a number, a stereotype, or "what happened the first time", which could be as much about the context as the actual decision made. If your first foreign vacation was a nightmare, would that bias you against all future foreign vacations? It might well do.

Strangely enough, in the modern era of near-ubiquitous information technology, when the information relates to the past, the first information received is most likely to be that relating to the most recent period. It is generally easy to find a product's sales figures for yesterday, last month or last year, and even for today, the current month or year (so far). It usually takes more effort to find the figures for the period before that. We are all used to this; that's how online banking systems typically give us our account information, as opposed to the paper statements of previous decades, where the reader's eyes were typically drawn towards a starting point (say) one month ago. The current presentation is great for seeing how much money we have now; not so good for recognising trends *(interpretation)*.

Even algorithms can suffer from anchoring. Where a method involves a large number of iterations, gradually converging on the final solution, there is a whole area of the subject covering research into whether or not the starting parameter values affect the final solution produced by the system *(analysis)* (see e.g. Griffiths and Higham, 2010). Small changes in one do not necessarily lead to small changes in the other; those changes can be very large. This latter phenomenon gave rise to a whole new field in the form of catastrophe theory (Poston and Stewart, 1978).

A very important variant of this form of bias is using only the data that the organization already has, rather than attempting to search for any more *(acquisition)*. Cristina Cabella of IBM said in an interview with *Computer Weekly* that "research shows that data scientists spend around 60% of their time cleaning and organising data before they can use it because for decades, organisations have been collecting whatever they could without any consideration of how that data would be used" (Ashford, 2018). This stresses the additional issue that the purpose for which data are collected may affect which data are collected, and the details of the data that are

collected. This is similar to the statistician's distinction between primary data, which are data specifically collected for the issue at hand by those working on it, and secondary data, which have already been collected by someone else (Moore, McCabe, & Craig, 2017). There is a tendency to assume that organizational data must be primary data, but that is by no means always the case. This is especially true if the definitions and assumptions have not been documented *(acquisition* and *interpretation).*

An example of how important this is comes from the data used to describe home energy consumption using sensors. A sample online dataset (Candanedo, 2017) shows there are at least two important elements to be aware of in the data gathering process. On the one hand, which scales to use for temperature, pressure, humidity, etc.; and on the other, time and location of measurement. Both elements can be a source of bias. If changes can occur very rapidly in time, then data needs to be observed frequently enough. Equally, in this dataset the weather data came from the nearest weather station, at Chievres Airport. But airports, for obvious reasons, are often located in areas of open country, and so can have specific local atmospheric conditions. It is only too easy to just "grab" a dataset without checking its assumptions *(integration).*

In analytics practice there are many examples of this anchoring phenomenon; bias towards the first information received. For instance, the first way that people learn to summarize a variable's value is the arithmetic mean. So that is what people tend to do first, and often this can result in a situation like the old joke that having your head in the fridge and your feet in the oven produces on average a comfortable temperature for your body. More seriously, we also learn to "deal with" outliers or unusual values by treating them as not normal (in both the everyday and statistical

senses) and ignoring them. Yet they can have an important influence in the analysis, for example if the outliers are at the beginning of the period that is being analyzed. Analytics models of time series, for example, have had to be adapted for financial time series because of the important weight that price volatility (a type of uncertainty) has in the decisions. GARCH (generalized autoregressive conditional heteroskedasticity) models (Brockwell and Davis, 2016) were developed to provide more predictive power about these volatility effects closer to the time point of decisions *(interpretation)*.

A simpler example of this bias is often found in thinking about queuing processes. Bias towards the first information may lead people to model the process as deterministic when in fact there is considerable variation. Just because the first event takes a long time in service, it does not mean that all following events will take a similarly long time. That is why in queuing theory models, Poisson and exponential probability distributions are used to represent arrival events and service times as a stochastic process (Shortle, Thompson, Gross, & Harris, 2018).

Keeping the status quo

In most organizations, and indeed in life as a whole, it is easier to leave things as they are than to introduce a change. After all, change brings risks as well as improvements: people do not like change, it causes disruption (by definition) and it does not always work out as planned. Change management is a field of research in itself.

Another form of this particular bias is to do something, but to opt for the same choice as the last time this sort of decision came up, without actually giving serious consideration to any alternative choices. Again, this is a common human behaviour; quite a lot of people always spend their vacations in the same place!

The most general form of this bias is assuming that the future will be too much like the past, especially by assuming that any change will be gradual and incremental rather than sudden and/or radical (Huber, 2004). The IT industry has some legendary examples of this:

"I think there is a world market for maybe five computers." Attributed to Thomas Watson, then president of IBM, in 1943.

"There is no reason anyone would want a computer in their home." Attributed to Ken Olsen, founder of DEC (Digital Equipment Corporation), in 1977.

"Two years from now, spam will be solved." Attributed to Bill Gates, founder of Microsoft, in 2004.

Missing a "tipping point", where radical change occurs, is often a recipe for business failure, as the cases of Pan-Am (the arrival of budget airlines), Kodak (digital photographic technology - which, ironically, they held an original patent for - replacing film) and Digital Equipment Corporation (the eventual dominance of personal computers) demonstrate.

A more subtle, but ubiquitous form of keeping the status quo is given by former Google "Design Ethicist" Tristan Harris. "Determining the choices on a menu in a piece of software restricts what users can do - so who controls the choices?" Whoever that is, enforces what they see as the status quo (Harris, 2016). Software design has an even broader influence. How are options presented to a user? The selection of variables, the order, and even their number can all limit what the user can do, even if this is unintentional.

This is why there is a key difference between checking a box to opt in and checking a similar box to opt out. As well as this binary selection, the scales chosen

and the order of items in lists can affect the user's understanding of the meaning of a construct and even drive them to incorrect interpretations *(interpretation)*.

Far from it being beneficial to keep the status quo, being aware of the presence of events that modify the behaviour of the data can be extremely important for data analytics. For example, after 9/11, the volume of passengers in the airlines market dropped substantially, yet the seasonality was preserved. In a detailed analysis, the industry appeared to go back almost five years in revenues while still carrying the same level of costs (Blunk, Clark, & McGibany, 2006).

Another situation that overturns the status quo is when the analysis period includes transformations such as mergers and acquisitions or significant new products. Changes need to be considered explicitly in the analysis in order to avoid the bias of making decisions under the outdated historical parameters *(analysis)*.

A more technical example of keeping the status quo in analytics practice is when there is a possibility of overfitting: fitting models resembling a particular specific dataset without the opportunity to generalize. In particular, in machine learning algorithms there is the risk of tailoring the models so closely to the specific data sets that it reduces the possibility to generalize them. Samples are needed for training the model, for validating the model, and for testing the model. These sets need to be selected from different time periods if possible *(acquisition)*.

Alternatively, we can obtain statistical indicators of correlations among items to describe a concept, such as quality, but possibly the items are not quality related but simply show the same direction of results. As an early statistics lesson says, correlation does not equal cause and effect.

As mentioned in the previous section, the lack of understanding of variation in behaviour can create the expectation that data and results in the future should be

highly close to the past and present conditions. To sum up, perhaps the most common mistake in prediction is to assume that the future will be too much like the past (Huber, 2004). Predictive analytics is no exception to this.

Justifying past choices

This source of bias refers to making new decisions that continue to justify the decisions already made. In other words, refusal to admit that you might have made a mistake. More subtly, it can be a reluctance to admit that your policies have been taking the organization in a wrong direction.

The simplest example is when there are good results in organizations but nobody has a clear answer to the question - why? Sears is an example of a company having a complete integrated direct marketing strategy that nevertheless did not adapt the organization to electronic business, while Blackberry (RIM Corp.) created the digital keyboard but did not use it in their products because of the success they were having with the physical keyboard. Another example comes from product development in the automobile industry where some manufacturers continued producing big-engined cars that were limited in fuel efficiency. The values of customers changed faster than these car makers did.

The widest form of this problem is known as *groupthink* (Janis, 1982), where a group refuses even to consider evidence that disagrees with its current position, which could be classed as either *(acquisition)* or *(interpretation)*.

A good example of this in the United Kingdom was the behaviour over several years in the 1970s of British Leyland (later Austin Rover), then one of the UK's two largest automobile manufacturers and responsible for the highly successful Mini, amongst many other models. Independent consumer surveys regularly found that the

build quality and reliability of British Leyland cars was significantly worse than that of their main competitors. The company equally regularly denied this and tried to trash the results, claiming that the surveys, one of which typically included over 10,000 responses each year about British Leyland cars alone, were not representative. Any statistician would realize that the company's reaction was nonsense, since the number of responses was an order of magnitude greater than would be found in a typical survey such as an opinion poll. Nevertheless, the groupthink of British Leyland's management literally would not allow them to accept what the survey results were telling them. Only in 1980, when the company were negotiating a collaborative agreement with Japanese manufacturer Honda to produce cars together, did the comments of the Honda management (which British Leyland had to respect, to secure the agreement) finally convince them that there really was a problem with quality and reliability, and something needed to be done about it. As a result, the first product of this collaboration, a model called the Triumph Acclaim, had the lowest percentage of warranty claims for any British Leyland model (Wikipedia, 2004).

Groupthink can be a widespread problem, yet it has obviously been recognised for centuries; for example, it is an important element in the well-known Hans Christian Andersen fairy tale, "The Emperor's New Clothes". [Spoiler alert: that story also depends on deliberate deception, and the relationship between the group leader and the group's other members.]

In analytics, Amazon had to withdraw a machine learning system for sorting through job applications. It was designed to replicate existing hiring practices, using 10 years of past data, but "learned" a bias against women applicants based on that data *(acquisition)* (Cook, 2018). Possibly Amazon's past practices needed closer scrutiny.

An even more extreme example, where analytics were deliberately used to mislead in order to justify past choices comes from the Enron scandal. As Zajechowski (2017) points out, "there is little question that the data that was being provided to shareholders was largely fictionalized". This illustrates our point that deliberate bias may only be known to some people. As McLean and Elkind (2003) explained (p.239), "Certainly there were executives who worried about the company's financial maneuvers, but the vast majority of people who worked for Enron simply assumed that the Global Finance team and Enron's accountants at Arthur Andersen - not to mention the stock analysts and credit analysts - knew what they were doing and that there was nothing for them to worry about."

Supporting instinct

This source of bias is based on Hammond, Keeney and Raiffa's "confirming evidence trap". It's the tendency to search for information that agrees with your instincts (which are themselves sources of bias) rather than information that contradicts or challenges them. Hammond, Keeney and Raiffa advise (p.124) "Don't surround yourself with yes-men", which is a term, less common now than in 1998, that itself manages to include *two* forms of bias - who has ever heard the term "yes-women" used? Nevertheless, it's still good advice.

There is a great deal of debate as to the effect social media have had on this type of bias. Certainly the spread of social media has made it easier to be exposed only to similar opinions to one's own (Vicario et al., 2016). The jury is still out as to whether or not this really does reinforce them *(acquisition)*. In an organizational context, constructive disagreement is recognized as an important element in innovation. As Simons, Mintzberg, & Basu (2002) observed, "Companies need to be

lean and mean. 'Lean and mean' is back in fashion these days. It's a mantra for getting in shape after the recession, just in time for the recovery. 'Lean' certainly sounds good--better than 'fat.' But the fact that 'mean' sounds good is a sad sign of the times." The multi-criteria of the decision-making process and the definition of driving in the correct direction can create controversy from which management needs to obtain the positive to create value and competitive advantage.

Analytics in practice has to deal with the use of theoretical rules/statistical tests/evaluation methods in order to improve the accuracy of the process and the acceptance of the results it produces. There are rules of thumb that people use, such as a correlation coefficient that is greater than 0.7 indicates high correlation, the p-value has to be less than 0.05 to consider variables as significant in a model, or more complicated when the indicator of model goodness is a number that incorporates many aspects of the modelling process such as the R-squared often used to judge the fit of models with "good" values greater than 0.7 (Hair, Black, Babin, & Anderson, 2010; Wang and Thompson, 2007) *(interpretation)*.

Analytics practice shows that in most cases the analytics professional requires a combination of rules/methods to determine if the model is useful and can help to make better decisions. The bias potentially appears in the interpretation of the results and the application of the criteria to decide what is good or bad in a transparent way. The force of instinct/intuition holds a lot of importance because the grey areas may be subject to bias. In the financial market the product Collateralized Debt Obligations (CDOs) was created based on the combination of assets (mortgages in different cities) and the instinct of those working in mortgage markets pointed to having independent (non-correlated) mortgage city-markets (Buffet, 2016). The models were created based on the assumption of independence of the assets in different cities; the reality

showed that the assets were correlated and the correlation adversely affected the outcomes for the investors *(interpretation)*.

Problem-solving methods affecting decisions

This is related to what Hammond, Keeney and Raiffa call the "framing trap". There is an extensive literature on problem structuring (Pidd, 1996; Rosenhead, 1989) and how it links to subsequent phases of the decision-making process. However, what many decision makers fail to realize is that the "structure" of the problem is in the heads of the people thinking about it. It is not a fixed property of the problem itself. Thus an inappropriate approach to structuring and/or solving the problem is likely to lead to an equally inappropriate solution. There is almost certainly nothing in an analytics technique that can indicate if it is appropriate to use it; the techniques exist in the world of statistics, mathematics and computation, not the world of business and decisions. Use of inappropriate methods due to lack of knowledge of others is the source of the old saying "if all you have is a hammer, everything looks like a nail".

In a context of uncertainty, human judgment can affect the choice of the way to solve a problem, to define if a result is appropriate or if a heuristic is good enough to follow. As Tversky and Kahneman (1974, p.1126) commented "The illusion of validity. As we have seen, people often predict by selecting the outcome (for example, an occupation) that is most representative of the input (for example, the description of a person)."

An extreme example of this source of bias comes from medical research, where for many years (far too many) clinical trials used only adult male subjects (Schiebinger, 2003), even if the drugs and procedures would as a result then be approved to treat women. The reasons given for choosing only male subjects actually

related to the possibility that female subjects might affect the results. Despite this "clue", no notice was taken of by far the greatest bias, which was assuming that there was no difference in treatment effectiveness between male and female subjects *(acquisition)*. The outcome is that much medical "knowledge" of how to treat female patients is still based on research results that may not be valid for them.

We mentioned menus and other forms of display as a potential source of "keeping the status quo" bias, but the detail of how analytics results are displayed can also be biased by the choice of method. Histograms are a common way to visualize data, but with a numerical variable the class boundaries (the "edges" of the columns) are an arbitrary choice, and so open to bias. This can be avoided, as Edwards and Finlay (1997, p.98) point out: "the procedure to prevent unwanted bias creeping in is for the class boundaries to be decided upon prior to the analysis of the data - or better still, prior to the data being obtained." Good practice, but how many people, even analytics professionals actually do this? Instead, how often are the class boundaries determined automatically by the software being used *(analysis* and *interpretation)*?

There are other sources of bias in the mathematical/statistical analytics world. Professionals have adopted "what if?" analysis and the creation of scenarios as means to review possible outcomes for decisions. The participation of several experts for estimating parameters of the models is a way to enrich the process. However, bias in the parameter selection can come from taking values from low, average or high results when the variable is actually showing a positive or negative trend, or it is at an inflection point, etc. Bias can also stem from a predefined interest in including values associated with previous decisions. This fault is often associated with not having a full spectrum of possible results (sample points) to study. For example, always analysing the net present value in budgeting-projects decisions using the same interest rate, or

the same ratios and rates of change in the pro-forma financial statements can reduce the opportunity to observe results where the outcomes are qualitatively different. In this particular case sampling methods can help: Monte Carlo simulation has become an important technique to help in determining input values to models *(acquisition* and *interpretation)* (V. Clark, Reed, & Stephan, 2010).

If the analytics method being used relies on estimates of probabilities, then bias may arise from estimates that are poor or even invalid. One of the fundamental rules of probability theory is that the probability of two events A and B both happening, P(A∩B), cannot be greater than the probability of one of them happening, P(B), say, irrespective of whether A happens or not. But studies by Tversky and Kahneman (1983) gave examples where as many as 85% of subjects estimated P(A∩B) as greater than P(B), especially when event A was much more likely to happen than event B.

Modern analytics practice operates more broadly than just with numerical data. The advance of Big Data and the relevance that it has for so many human activities have created a wave of enthusiasm that people want to ride. In many cases, the cost of the new data is reasonable and the technology needed to use it is accessible. Nevertheless, the issue of using the appropriate data remains an aspect to deal with in the analytics process. For instance, in social media, tweets are created in high volume for multiple topics and in different circumstances. The use of the data because of its volume does not reduce the risk of having data that misrepresents the population and thus biases the apparent solution to the problem at hand. Incorrect sampling and/or incorrect review of the data available can lead to conclusions that are only acceptable for some groups. It is too easy for unethical people to use tweets for creating perceptions/images/lies, with consequent negative effects on the users of data

and the interpretation of that data for problem resolution *(acquisition* and

interpretation).

Lacking forecast feedback

We live in an uncertain world. In business, each decision we face only happens once,
so we only observe one outcome. We can never be 100% sure what would have
happened if we had made a different decision. Fortunately, many decisions are so
similar to other decisions that we can treat them as instances of the same decision.
Even so, the feedback we obtain may be limited. For example, sometimes the
feedback itself is a biased sample. Take the case of a bank considering applications
for loans. If the application is granted, then the bank will eventually have full
information about whether or not the loan was paid back on time - or at all. But if the
application is *not* granted, the bank's data stream ends there. The bank will never
know if it would have been a better decision to accept that loan application
(acquisition).

Forecasts leave the manager in a marginally better place to get feedback than
decisions. Many forecasts don't change the world, so it is fair to make a comparison
between the forecast and what actually happens. Though as with many of the other
traps, care needs to be taken with this assumption. The sales forecast for September
that we made in August should not change the actual value of sales in September,
should it? But what if it's B2B (business to business) sales, where all the sales have to
be negotiated, and the company is strongly target driven, and the forecast is seen as
the target…well, once the forecast for the month is about to be reached, wouldn't the
sales staff maybe ease up? It depends on whether or not they receive bonuses for each

sale, of course; but if they do not, it's a possibility. The phrase "self-fulfilling prophecy" is well-known.

Less well-known outside the organizational studies literature is the opposite phrase "self-defeating prophecy", but feedback is equally important to avoid it. As computer science professor Michael I. Jordan puts it, "it's not a problem if Netflix delivers similar recommendations to millions of users, but if a financial services firm recommends the same stock to a large number of clients, it can artificially alter markets" *(interpretation)* (Burns, 2018). We find it interesting that he uses the word "artificially". From our perspective, this would be a very real alteration, with a very real rise in the price of that stock.

A source of bias mentioned in the previous section, assuming a series is stationary when in fact it has a trend, can be made worse by the absence of forecast feedback. Edwards and Finlay (1997, p.300) again: "Where such a trend is present, the forecast will always tend to lag behind it, being too low if the trend is an increasing one, and too high if the trend is decreasing. This is a type of bias, and the forecast needs to be corrected to compensate for it." *(analysis)*

It is, however, not just the complete absence of forecast feedback that is a source of bias, but the lack of sufficiently detailed or accurate feedback. The assumptions in the modelling process are crucial for what to do and how to interpret results, and feedback is needed to validate them. Inappropriate assumptions of normality, or correlations or time effects can reduce the quality of the analytics process. To assume that a variable's mean, median and mode are equal, affects analysis in risk management where metrics such as VAR are based on percentile analysis. Misuse or misinterpretation can modify the risk taking positions and all this can arise because of incorrect selection of loss distributions. Similarly, the metric to

use for identifying the minimum error (MAD, MAPE, MSE, etc.) or goodness of fit of a model can affect what is regarded as a significant result or a valid model (Gilliland, Tashman, & Sglavo, 2015) *(acquisition* and *interpretation).*

Finally, the analytics process supports the decision-making process in the quest to achieve organizational targets. However, sometimes what the results of the analytics process reveal is that the targets cannot be achieved. It is not uncommon for results like these to go no further (Lin, 2014) (and see our later comments about dark data), but the organizations need to know how far they are likely to be from the original targets *(interpretation).*

Remedies to prevent or reduce bias in analytics systems

It is natural to think of remedies against bias at the level of the individual analytics project. There are remedies at that level such as feedback, already mentioned and with more to come later in this section. However, for the success of analytics in the organization over the longer term, it is more important to put in place remedies at a strategic level. These will influence all analytics projects and indeed all decision-making, and so we will discuss those first. There are six, as shown in Table 2: ensure diverse inputs, standardize terminology, document assumptions, centralize data, use tools appropriately and test, test and test again.

TAKE IN TABLE 2 ABOUT HERE

Ensure diverse inputs

The benefits of diversity apply to the whole organization, but above all under this heading, ensure that there is a diverse team working on everything connected with analytics. As with most areas of computer science, analytics professionals, especially those working in AI (artificial intelligence), tend to be young and male, and people of

colour are under-represented (Hayes, 2018; West, Whittaker, & Crawford, 2019). There is plenty of research evidence that diverse teams produce better outcomes than homogeneous teams (e.g. Brodbeck, Guillaume, & Lee, 2011) though diverse teams tend to start more slowly, and people working in the computer science and analytics fields are not renowned for their patience! Diversity certainly needs to cover age, gender and ethnic origin, but may need to cover many other aspects such as background, nationality (especially for multi-national organizations), training, family income group and even regionality within a country.

A diversity of inputs should help reduce all the sources of bias relating to what might be termed the "mental baggage" of the decision-maker, especially a single decision-maker: bias towards the first information received; keeping the status quo; justifying past choices; supporting instinct; and problem-solving methods affecting decisions.

Standardize terminology

Andy Goldin, director of data and analytics at PwC South East Asia Consulting, emphasised the importance of having standardised metrics and terminology surrounding data within an organisation in an interview with *Computer Weekly*, "I don't mind if the whole organisation is using data terms incorrectly, as long as it is used in the same way" (Tao, 2018).

That really sums it up. How can you rely on any analysis if you cannot agree what you are actually talking about?

The situation is admittedly not as bad as a generation ago, when one of the authors found that the "an order" meant *five* different things in a single organization. Depending on whom you were talking to, it could be: a machine that a customer

wanted; an instruction to production to build a machine; raw materials that were requested from a supplier; an instruction to deliver a machine; or an instruction to go out and service/repair a machine. Enterprise Resource Planning systems such as SAP have gone a long way towards standardizing terminology within large organizations, but there remains much to be done.

This is even more of an issue when the analysis cuts across organizations. What, for example, does the financial year or fiscal year mean? For limited liability companies in the USA it matches the calendar year. Limited companies in the UK can choose, with most selecting either the calendar year or the UK tax year, which runs from 6 April one year to 5 April the next. And that's the Gregorian calendar year, not the Jewish, Muslim or any other calendar year…

The main benefits of standardizing terminology are threefold: reducing bias towards the first information received and justifying past choices, because of the need to look at both of these in terms of the standard terminology; and reducing problems from the problem-solving method used because of the effort needed to perform the standardization, which forces departments to talk to each other. This is a subtle but vital benefit of this remedy.

Document assumptions

Failure to document assumptions turns what might have been primary data into secondary data. What *was* meant by financial year when this set of data was collected? How were missing values in the data dealt with? Does a student count as having started their course (and therefore having left without completing it) if they moved in to the university's residences but went away again without actually enrolling on the course?

Documenting assumptions avoids arguments about what items of data "really" mean. It therefore acts as a remedy against all of the "mental baggage" biases, and also helps to guard against lack of feedback, since the process of documenting assumptions requires closure of the loop - looking at what happened - to check that the assumptions in use are reasonable. Or, if they were documented a long time ago, checking that they remain reasonable.

It is also necessary to make sure that the assumptions are visible when people access the data or use the model: linking interpretation back to acquisition. Admittedly, this will not ensure that the users necessarily *read* them, but that is the best that can be done. No system is foolproof. Even forcing the assumptions to come up on the screen and requiring agreement to continue is not guaranteed to work: when did you last read the detail of a Terms and Conditions screen or pop-up before clicking or tapping "Yes, I agree"?

Centralize data

One of the authors of this article worked with a public sector organization providing social care that had four different sets of data on who it actually employed. All four were different, and none of them was completely correct. Not surprisingly, the closest to the real situation was the payroll system data, as it is in the employee's interest to make sure that payroll data about him/her is correct; that remains a very useful rule of thumb for where to find the most accurate data about employees if there is more than one system.

As the era of big data has made it much easier technologically to centralize data than previously, it might be thought that these problems would be disappearing. However, this is so far from being the case that the term "dark data" has been coined

for all the data that is not held in organizational systems and whose existence may be unknown (Martin, 2016), extending its original meaning of data from scientific experiments - typically those with negative results - that were never published in the literature. For example, the improved information technology that facilitates data centralization also makes it easier for people to take and keep their own copies (possibly unauthorized) of organizational data, which they use instead of accessing the corporate system every time. Corrections and amendments to the official data will probably not appear in these unofficial copies.

Centralizing data helps guard against bias towards the first information received, keeping the status quo and supporting instinct. The mechanism here is partly related to the benefits of ensuring diversity and standardizing terminology; no one person or department "owns" the data, so they cannot act to obscure it or its consequences. Centralizing data also encourages feedback, because it has the potential to make all data more widely available throughout the organization.

Use tools appropriately

Analytics tools can either be bought-in ready-made, custom developed by analysts (either in-house or working for external suppliers), or a combination of the two. Tools are our friends for obtaining results, but they have to be used with a lot of care because there are possibilities of errors, bias and misinterpretations. Software suppliers and consultancies are currently offering a wide range of AI or analytics solutions and architectures. However valuable these might potentially be, with all due respect to their providers, there is no way that they can possibly cover all analytics projects that an organization might need to do, without at the very least a substantial degree of in-house customization. Tools often come with predefined algorithms. The

greater the use of "built-in" functionality, the greater the risk of running into bias from problem-solving methods affecting decisions.

The analytics professional should always review the models and the way they are built inside any tool, and be able to justify their choices to the decision-makers who need the support. Not surprisingly, in advanced levels of analytics the analytics professional typically prefers to create their own tools (Chambers and Dinsmore, 2015), as this is easier to scrutinize.

A strategic remedy that managers can and should consider is to use different methods to approach the same issue. For example, if there is an interest in detecting outliers, the approaches might be descriptive, using clustering, or based on some rules regarding standard deviations, or even a deep learning black box approach. Comparison of the results from different methods can help to identify if there are hidden misconceptions or bias in the analysis. Another benefit is that in predictive analytics, a combination of forecasts from different methods is usually superior to a single method. This has been demonstrated for over 100 years (Galton, 1907), but is still often ignored or even denied (Hyndman, 2018). Some organizations do remember this lesson, though; lift/elevator manufacturer ThyssenKrupp uses machine learning systems to predict the occurrence of faults, combining the outputs of several open source machine learning algorithms to produce an overall result (L. Clark, 2018).

Appropriate use of tools is also an excellent remedy against the bias of supporting instinct, since any consideration of the tools forces more detailed scrutiny of what is being done with them.

Test, test and test again - check against gold standard if possible

We use this phrase to emphasise the importance of testing at all stages of the analytics

process, and of continuing to test after the system has been implemented. The ideal test of a predictive analytics system is what is known as a gold standard test: checking the prediction against a known outcome. For example, researchers at Stanford University (Esteva et al., 2017) have developed a neural network-based analytics system that classifies images of skin lesions that might be cancerous. Although part of the research involves comparing the performance of the analytics system with that of human dermatologists, the crucial gold standard element is that both are then compared against the results obtained from a biopsy of the potential cancer.

If no gold standard is available, then testing against the opinion of human experts is the best alternative, but it is important to realize that those human opinions - even of experts - may be subject to any of the forms of bias mentioned here.

With a predictive or prescriptive analytics system, it is also vital to keep on testing after the system has been implemented. Returning to ThyssenKrupp, Hyun-Shin Cho, their head of digital operations, observed that "models also get stale. We need to retrain them constantly because the action we take while maintaining the assets will make them behave differently in the future" (L. Clark, 2018). And the world may also change through actions taken by others.

Additional complexity can arise in models given the level of granularity of questions that people want to answer. In some examples, regression models are developed based on the whole data set, but once the data set is partitioned the models need to be revisited. Similarly, there are situations where the data can be used as it is, with the level of aggregation that is available, but equally there are cases where the data needs to be reorganized and studied according to the new definitions of variables in that reorganization. Finally, in terms of testing, another remedy is to create time series and follow-up means for organizational performance indicators/metrics, in

order to be more precise in the way that the organization measures performance and how it can define and achieve acceptable results and attainable goals. Again, this is a strategic issue, not one restricted to a single analytics project.

Having presented these six remedies at the strategic level, the next section considers the main project-level remedy, which is to use a standardized process/method.

Use a standardized process/method

At the level of the individual analytics project, by far the most important remedy is to use a standardized process/method for all analytics projects. We are coupling the terms process and method together to stress that this needs to cover both what the analytics team is supposed to do (method) and how it actually does it (process). However, while this can potentially act as a remedy for all the sources of bias we have identified, it is equally important to realize that it can itself be a source of bias, particularly in terms of problem-solving method and feedback..

This standardized process/method must cover both the steps leading up to the outcomes of the predictive or prescriptive analytics system, and what the organization does with them to take action. As well as the Jagadish, et al. (2014) "pipeline" and similar methods, there are other versions of the process that are less linear (see for example Edwards and Rodriguez, 2016).

The process/method should be based on permanent feedback, every step requiring the review of the current and previous steps. Thus the adaptation of the analytics process is based on an ongoing review.

Guidelines are needed for project/experiment design, sampling, evaluation of assumptions and obtaining and responding to ongoing feedback. As mentioned

already, another aspect to keep in mind that affects bias in the analytics process is the use of tools. Technological tools sold as "analytics software" (and sometimes hardware) come with predefined algorithms. Do they really fit what the organization needs? How well documented are the assumptions behind them? Even in the case of the most general analytics tool, spreadsheets, there are many wrong practices in developing models, such as huge nested formulae that are impossible to review. One mistake in a spreadsheet can lead to errors of hundreds of thousands of dollars (European Spreadsheet Risks Interest Group, 2003).

Any tool used for an analytics project has its own specific set of issues. For example, with classification or regression trees, what kind of algorithms are implemented? In multiple regression it is possible to control the entering and leaving variables in models, and the partitions of data that are generated. Other more general issues include which algorithms are used for optimization, what type of methods/distances are used for defining similarities etc.

A complex issue to tackle in the analytics process is when different approaches to the same problem yield different results, and people do not know why the solutions, apparently valid, are not pointing to the same results. The analytics team - including the user managers - need to clarify how the models were determined, how the steps were predefined, and what was actually done.

External pressure to avoid bias

Governments are starting to suggest or even mandate some of the remedies we have been discussing. Article 22 of the European Union's General Data Protection Regulation (GDPR) relates to automated decision-making and profiling about an individual. It requires the individual's explicit consent (by an "opt in" not an "opt

out", as mentioned earlier), and also that the organization introduces simple ways for them to request human intervention or challenge a decision. The presence of bias might well put an organization in breach of these requirements, although with a "black box" analytics system it could be difficult to spot.

A recent UK government report on AI (Anonymous, 2018a) states (p.12) "We are concerned that many of the datasets currently being used to train AI systems are poorly representative of the wider population, and AI systems which learn from this data may well make unfair decisions which reflect the wider prejudices of societies past and present." and therefore recommends that (p.12) "[Researchers and developers] need to ensure that data is pre-processed to ensure it is balanced and representative wherever possible, that their teams are diverse and representative of wider society, and that the production of data engages all parts of society." and that "Alongside questions of data bias, researchers and developers need to consider biases embedded in the algorithms themselves --human developers set the parameters for machine learning algorithms, and the choices they make will intrinsically reflect the developers' beliefs, assumptions and prejudices. The main ways to address these kinds of biases are to ensure that developers are drawn from diverse gender, ethnic and socio-economic backgrounds, and are aware of, and adhere to, ethical codes of conduct."

In the health and care sector, the UK government has now gone further, by issuing an official code of conduct on the use of data-driven technology in that sector (Anonymous, 2018b). Note however, that at the time of writing this is guidance rather than mandatory; the difference between those two terms matters a great deal in the UK public sector. This guidance includes the requirement to document and follow a Standard Operating Procedure, as suggested earlier.

Concluding remarks

Literature on human decision making identifies six common sources of bias. The main contribution of this article is to translate these to the context of decision making by analytics systems, and offer a set of remedies to reduce these biases. Whereas much prior literature in analytics treats bias as a technical issue, most of our discussion is centred on the phases in the analytics process that link most closely to the business: acquisition and interpretation.

We have presented the six sources separately. However, in the analytics process these sources of bias can come in combination. For example, keeping the status quo or justifying past decisions is often linked to anchoring to the first information received and the way that instinct is used and justified. Similarly, issues of bias in data display appear under more than one heading.

The remedies for the sources of bias include using an appropriate analytics project process. The risk of thinking that merely having higher volumes of data will solve the problems is a real one. The organization needs to follow systematic steps with ongoing feedback and review.

Interpretation requires business understanding of the organization. Bias in the methods used in analytics can be avoided by stringent testing and evaluation of partial results. For operational implementation, the organization needs to develop two things: stress testing methods; and a method to evaluate and review the development of models. Testing also needs to continue after the systems are operational.

The main limitation of this article is that it is based on the six most common biases in human decision making according to Hammond et al. (1998); there are other types of such bias that may be relevant to analytics systems. It is also open to any bias in the authors' own acquisition and interpretation activities.

The most important area for future research is empirical work to test the effectiveness of the remedies in practice, and to see if any other important sources of bias are present. Secondly, our focus in this article has been on the development and initial use of the analytics system. Further research is needed into how the continuing use of an analytics system affects the amount, or the type, of bias in its decisions. A third important area is to attempt to construct a causal model linking observed bias back to its sources.

As a final positive point, remember that analytics can also be used to test for human decision-makers' bias, for example in hiring practices (Thibodeau, 2018).

References

Amrein-Beardsley, A. (2017). Breaking news: A big victory in court in Houston. Retrieved 23 May, 2019 from http://vamboozled.com/breaking-news-victory-in-court-in-houston/.

Anonymous. (2018a). *Government response to House of Lords Artificial Intelligence Select Committee's Report on AI in the UK: Ready, Willing and Able?* (CM9645). Her Majesty's Stationery Office.

Anonymous. (2018b). *Initial code of conduct for data-driven health and care technology*. Her Majesty's Stationery Office Retrieved from https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology.

Ashford, W. (2018). AI is key to keeping IBM compliant with GDPR. *Computer Weekly*. Retrieved 29 June 2018.

Blunk, S. S., Clark, D. E., & McGibany, J. M. (2006). Evaluating the long-run impacts of the 9/11 terrorist attacks on US domestic airline travel. *Applied Economics, 38*(4), pp. 363-370. doi:10.1080/00036840500367930

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (Third ed.) Berlin: Springer.

Brodbeck, F. C., Guillaume, Y. R. F., & Lee, N. J. (2011). Ethnic diversity as a multilevel construct: The combined effects of dissimilarity, group diversity, and societal status on learning performance in work groups. *Journal of Cross-Cultural Psychology, 42*(7), pp. 1198-1218.

Buffet, M. (2016). How do CDOs and CDSs influence the crisis of 2008. *Lingnan Journal of Banking, Finance and Economics, 6*. Retrieved 19 April 2019.

Burns, E. (2018). Tech experts weigh in on the AI hype cycle. *Computer Weekly*. Retrieved 1 September 2018.

Cambridge Dictionary. (2018). Bias. Retrieved 24 May, 2019 from
     https://dictionary.cambridge.org/dictionary/english/bias.

Candanedo, L. (2017). Appliances energy prediction data set. Retrieved 23 May, 2019
     from https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction.

Chambers, M., & Dinsmore, T. W. (2015). *Advanced Analytics Methodologies:
     Driving Business Value with Analytics* Upper Saddle River, NJ: Pearson
     Education.

Clark, L. (2018). AI elevates predictive maintenance for Kone and ThyssenKrupp.
     Computer Weekly. Retrieved 13 September 2018.

Clark, V., Reed, M., & Stephan, J. (2010). Using Monte Carlo simulation for a capital
     budgeting project. *Management Accounting Quarterly, 12*(1), pp. 20-31.

Cook, J. (2018). Amazon scraps 'sexist AI' recruiting tool that showed bias against
     women Retrieved 23 May, 2019 from
     https://www.telegraph.co.uk/technology/2018/10/10/amazon-scraps-sexist-ai-
     recruiting-tool-showed-bias-against/.

Edwards, J. S., & Finlay, P. N. (1997). *Decision Making with Computers: the
     Spreadsheet and Beyond* London: Pitman.

Edwards, J. S., & Rodriguez, E. (2016). Using knowledge management to give
     context to analytics and big data and reduce strategic risk. *Procedia Computer
     Science, 99*, pp. 36-49. doi:10.1016/j.procs.2016.09.099

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S.
     (2017). Dermatologist-level classification of skin cancer with deep neural
     networks. *Nature, 542*, pp. 115-118. doi:10.1038/nature21056 Retrieved from
     http://dx.doi.org/10.1038/nature21056

European Spreadsheet Risks Interest Group. (2003). Spreadsheet mistakes - news
     stories collated by the European Spreadsheet Risks Interest Group. Retrieved
     23 May, 2019 from http://www.eusprig.org/horror-stories.htm.

Fearn, N. (2018). Police facial recognition systems are 98 per cent inaccurate, says
     research. Retrieved 23 May, 2019 from
     https://www.computing.co.uk/ctg/news/3032401/police-facial-recognition-
     systems-are-98-per-cent-inaccurate-says-research.

Galton, F. (1907). Vox Populi. *Nature, 75*, pp. 450-451.

Gilliland, M., Tashman, L., & Sglavo, U. (Eds.). (2015). *Business Forecasting:
     Practical Problems and Solutions*. Hoboken, NJ: John Wiley & Sons.

Griffiths, D. F., & Higham, D. J. (2010). *Numerical Methods for Ordinary
     Differential Equations: Initial Value Problems* London: Springer.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data
     Analysis* (7th ed.) New Jersey: Pearson Prentice Hall.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). The hidden traps in decision
     making. *Harvard Business Review, 76*(5), pp. 47-56.

Harris, T. (2016). How Technology Hijacks People's Minds—from a Magician and
     Google's Design Ethicist. Retrieved 03 August, 2018 from
     http://www.tristanharris.com/essays/.

Hayes, G. (2018). What does a data scientist REALLY look like? Retrieved from
     https://www.kdnuggets.com/2018/11/data-scientist-look-like.html

Huber, G. P. (2004). *The Necessary Nature of Future Firms: Attributes of Survivors
     in a Changing World* Thousand Oaks, CA: Sage Publications.

Hyndman, R. J. (2018). A brief history of time series forecasting competitions.
     Retrieved from https://robjhyndman.com/hyndsight/forecasting-competitions/

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Exploring the inherent technical challenges in realizing the potential of Big Data. *Communications of the ACM, 57*(7), pp. 86-94. doi:10.1145/2611567

Janis, I. L. (1982). *Groupthink* Boston: Houghton Mifflin Company.

Lebowitz, S., & Lee, S. (2015). 20 cognitive biases that screw up your decisions. *Business Insider UK*. Retrieved 06 August 2018.

Lin, N. (2014). *Advanced Business Analytics: Integrating Business Process, Big Data, and Advanced Analytics* Upper Saddle River, NJ: Pearson Education.

Martin, E. J. (2016). Dark data: Analyzing unused and ignored information. *Econtent, 39*(5), pp. 6-8.

McLean, B., & Elkind, P. (2003). *The smartest guys in the room* New York: Portfolio.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). *Introduction to the Practice of Statistics* (9th ed.) New York: W. H. Freeman.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* New York: Penguin Random House.

Pidd, M. (1996). *Tools For Thinking: Modelling in Management Science* Chichester, England: John Wiley & Sons Ltd.

Poston, T., & Stewart, I. (1978). *Catastrophe Theory and Its Applications* London: Pitman.

Robinson, A., Levis, J., & Bennett, G. (2010). INFORMS to officially join analytics movement. *OR/MS Today, 37*(5), p 59.

Rosenhead, J. (1989). *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict* Chichester: John Wiley & Sons.

Schiebinger, L. (2003). Women's health and clinical trials. *Journal of Clinical Investigation, 112*(7), pp. 973-977. doi:https://doi.org/10.1172/JCI19993

Shortle, J. F., Thompson, J. M., Gross, D., & Harris, C. M. (2018). *Fundamentals of Queueing Theory* (5th ed.) Hoboken, NJ: John Wiley & Sons.

Simon, H. A. (1969). *The Sciences of the Artificial* (1st ed.) Cambridge, MA: MIT Press.

Simons, R., Mintzberg, H., & Basu, K. (2002, June). Memo to CEOs: The five half-truths of business. *Fast Company, 59,* 117-121.

Snow, J. (2018). Amazon's face recognition falsely matched 28 members of Congress with mugshots. Retrieved 23 May, 2019 from https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28.

Tao, A. L. (2018). How F1 and others are moving beyond descriptive analytics. *Computer Weekly*. Retrieved 30 July 2018.

Thibodeau, P. (2018). Gartner analyst sees limit to tech's ability to fix hiring bias. *Computer Weekly*. Retrieved 31 August 2018.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty : Heuristics and biases. *Science, 185*(4157), pp. 1124-1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), pp. 293-315. doi:10.1037/0033-295X.90.4.293

Vicario, M. D., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences USA, 113*(3), pp. 554-559.

doi:10.1073/pnas.1517441113 Retrieved from https://www.pnas.org/content/pnas/early/2016/01/02/1517441113.full.pdf

von der Beck, I., Cress, U., & Oeberst, A. (2018). Is there hindsight bias without real hindsight? Conjectures are sufficient to elicit hindsight bias. *Journal of Experimental Psychology Applied*doi:10.1037/xap0000185

Wang, Z., & Thompson, B. (2007). Is the Pearson r^2 biased, and if so, what is the best correction formula? *Journal of Experimental Education, 75*(2), pp. 109-125.

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race and Power in AI*. New York: https://ainowinstitute.org/discriminatingsystems.html

Wikipedia. (2004). Triumph Acclaim. Retrieved 23 May, 2019 from https://en.wikipedia.org/wiki/Triumph_Acclaim.

Zajechowski, M. (2017). The lessons we can learn from bad data mistakes made throughout history. Retrieved 26 May, 2019 from https://www.smartdatacollective.com/lessons-can-learn-bad-data-mistakes-made-throughout-history/.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2017) *Men also like shopping: Reducing gender bias amplification using corpus-level constraints*. Paper presented at the EMNLP, Copenhagen, Denmark. http://markyatskar.com//publications/bias.pdf

Table 1: Our sources of bias shown against the six traps of Hammond, et al. (1998)

Our sources of bias	*Hammond, Keeney and Raiffa's traps*
Bias towards the first information received	The Anchoring Trap
Keeping the status quo	The Status-Quo Trap
Justifying past choices	The Sunk-Cost Trap
Supporting instinct	The Confirming-Evidence Trap
Problem-solving methods affecting decisions	The Framing Trap
Lacking forecast feedback	The Estimating and Forecasting Traps

Table 2: Listing the sources of bias against the strategic-level remedies

Sources of Bias	Remedies
Bias towards the first information received	Ensure diverse inputs Standardize terminology Document assumptions Centralize data Test, test and test again
Keeping the status quo	Ensure diverse inputs Document assumptions Centralize data
Justifying past choices	Ensure diverse inputs Standardize terminology Document assumptions Test, test and test again
Supporting instinct	Ensure diverse inputs Document assumptions Centralize data Use tools appropriately Test, test and test again
Problem-solving methods affecting decisions	Ensure diverse inputs Standardize terminology Document assumptions Use tools appropriately
Lacking forecast feedback	Document assumptions Centralize data Test, test and test again