Dangerous data: analytics and information behaviour in the commercial world

Luke Tredinnick (l.tredinnick@londonmet.ac.uk)

Research paper

Abstract: Data has become an increasingly important component in the contemporary business operations, epitomised by the rise of the Business Intelligence system, data analytics, and data visualisations. It has been associated with increased productivity and the development of new business opportunities. But the use of data is sometimes also associated with poor decision making, either because of the quality of the data on which decisions are made, or because of the ways in which that data is used. This paper explores the problem of dangerous data in commercial contexts: those situations where the use of data contributes to worse outcomes.

1. Introduction

This paper explores the dilemma of data in contemporary business contexts. Data has become vital component to planning and decision making in many contemporary commercial organisations, delivering improvements in efficiency and driving new business processes. Brownlow et al for example have argued that "many businesses are developing new business models specifically designed to create additional business value by extracting, refining and ultimately capitalizing on data" and describe data as "the new oil" (2015). These new models include the data driven business (Bulger et al, 2014) and data-driven decision-making, described by Yu et al "the process of using evidence and insights derived from data to guide the decision-making process" (2021). Data has been associated with significant increases in productivity (Brynjolfsson et al, 2011) and the development of new business opportunities alongside other benefits. But while data has undoubtedly transformed business in recent years, it can also become responsible for bad decision-making. Dangerous data describes those situations where the use of data contributes to worse outcomes.

In commercial contexts we now have access to swathes of data about every aspect of the business; data accumulates around costs, communications, business transactions, billing and around resource requirements. Sadowski argues that "companies are clamouring to collect data – as much as they can, wherever they can" (2019). Commercial and non-commercial organizations alike have become machines for generating ever more information as if manufacturing data were the primary driver of contemporary economic activity. This rise of data and its use to improve business processes has become associated with Business Intelligence defined as "the processes, technologies, and tools needed to turn data into information, information into knowledge and knowledge into plans that drive profitable business action" (Moss and Hoberman, 2004). Chang has described Business Intelligence as::

> The accurate, timely, critical data, information and knowledge that supports strategic and operational decision making and risk assessment in uncertain and dynamic business environments (Chang, 2006).

Business Intelligence solutions like Microsoft's Power BI suite and SAP Lumira have revolutionised the use of use of data in organisations of all kinds, integrating data sets from across the business into a single customizable interface, allowing desktop access by employees across the organisation, and integrating sophisticated AI driven analytics. They promise to release data from the silos associated with the legacy of different systems and software (Patel, 2019), and allow a truly integrated use of data across the organisation. Typical elements of the Business Intelligence system include reporting, dashboards, ad hoc query, search-based BI, online analytical processing, interactive visualization, scorecards, predictive modelling, and data mining (Chen et al, 2012).

While integral to strategic management, data and data analytics are not just impacting on senior management and strategic planning, but have become an everyday part of the working environment for millions of employees. Negash (2004) highlights the benefits of Business Intelligence Systems at every level within an organization at which decisions are made from senior management down, and suggests that:

> Business Intelligence systems provide actionable information delivered at the right time, at the right location, and in the right form to assist decision makers (Negash, 2004).

One important way in which employees encounter the new opportunities presented by software of this kind has been through the data dashboard with its integrated data visualisation that provide a snapshot of the organisation often through key performance indicators. Both dashboard solutions and data visualisation come with risks in the workplace, as this paper will explore below, but they have nevertheless become a familiar part of the work environment. We may be moving beyond the data dashboard in the 2020s towards configurable and personalised data analytics solutions, yet the underlying trend of data-driven business planning and decision-making looks set to stay. Indeed a recent report by Forrester suggests that data-driven companies are growing by over 30% annually (Hopkins et al, 2018), with the perceived benefits including (Sadowski, 2019):

- Profiling and targeting customers;
- Optimising systems and processes;
- Better managing and controlling the organization;
- Improved modelling and forecasting;

But does the granular oversight promised by our data-hoarding addiction actually drive performance, increase revenues, and grow the business, or does more data sometimes undermine operational efficiency?

Data-hoarding comes at a cost. The integration of Business Intelligence solutions with existing software and data sets may reduce the costs of collating, storing data, processing and re-presenting an organisation's existing data resources, but the cognitive overheads of exploiting that data are non-negligible. Cognitive overhead is widely defined in data science as the "logical connections or jumps your brain has to make in order to understand or contextualize the thing you're looking at" (Lieb, 2013); for it to become useful data has to enter business processes and becomes an additional factor in commercial decisions at every level. And data-hoarding comes with knock-on effects on the business and its culture including possibly undermining the trust between employees and employers through increased dataveillance (McParland & Connolly, 2020) and encouraging false productivity or the tendency to engage in work that meets targets but does not achieve any real outputs. While the benefits of data driven business are undeniable, the downsides are less well explored.

This paper addresses one key aspect of data proliferation in contemporary workplaces: what happens when we over-estimate the value of information in commercial operations resulting in a situation where more data generates worse decision-making. Bulgar et al (2014) note that "data analysis in ignorance of the context can quickly become meaningless or even dangerous"; the use of data can sometimes lead to unexpected outcomes. Dangerous data describes those cases where we relegate decision making to heuristic principles without necessarily fully interrogating the quality of the information on which those principles operate or its relevance to the decisions we are taking, cases where we rely on established principles without fully determining their applicability in specific circumstances. Nobody sets-out to make bad decisions in their work, but sometimes our decisions reflect those outcomes that can be justified, rather than those that are most desirable. Dangerous data provides the basis on which to defend bad decisions many times each day in organisations across the globe.

2. The problem of dangerous data

Dangerous data is data that contributes to worse decision-making. Of course the availability of the right data is fundamental to the efficacy of the choices we make in professional and commercial practice. Data can ensure that our decisions are made on a stronger basis that personal belief or professional judgement. But even the most reliable, high quality, and rigorously sourced data can also become dangerous when it is used in the wrong way or for the wrong tasks. And the data on which we base our decision is often far less well sourced and less well-understood than that at the point of use. Data becomes dangerous as a product of the contexts within which it is used, and there are several reasons for this; it may be incomplete, inaccurate, irrelevant, or otherwise misleading; or we may be applying it to the wrong situations, or interpreting it in the wrong way.

One facror is that the aggregation and retention of data about the organization has sometimes become an end in and of itself. The data-driven business has come to be perceived as a solution to the problems of organizational efficiency whether or not lack of data is actually a cause of those inefficiencies. Cukier (2015) has observed that:

> We fetishize data, we think that data is the answer. It's far from the truth. In fact, it's ridiculous, because the data is only a simulacrum of reality in the same way that a map is not a territory. And so while we need to use information and data to make decisions as we need to do, the data is always unfaithful, always unreliable, it always misleads, and you have to torture it until it confesses.

Quantitative information is often regarded with an air of impartiality and objectivity that distracts us from its highly contextual and situated origins; quantitative data packaged-up by the organisation in an accessible form and promoted as a key resource in all business processes becomes more difficult to contest. We seek "hard" data to justify our beliefs, and internal data is often easy to come-by, satisfying our data needs. But poor data can lead to poor decision making because data can take on the burden of evidence to the exclusion of alternatives, including personal experience, tacit knowledge, professional experience, and individual testimony. In an uncertain, competitive and mutable business environment, there is a reassuring certainty conveyed by the aggregation of data about the operation of the organization, a disarming illusion of control and oversight.

Data of any kind emerges from socially situated processes that are themselves imbued with assumptions and beliefs. Data generated within the organisation is only as reliable as the care which has gone in to collating it, and when data is automatically generated by processes whose principle purposes are something other that generating the data itself, then the quality and reliability of that data can become influenced by the social context within which it was produced. If for example we evaluate performance against data that has been generated for some other purpose such as resource management or work planning, then it is hardly surprising that individuals might prioritise the impact of that data on their careers over its accuracy. The purposes to which we put the data that we generate can influence the quality of that data. Kramer (2012) has written:

> Data is not reality, and conflating the two is dangerous. Nor is data neutral or objective. If we learned anything from the last fifty-odd years of cultural and social critique, it is this fact: that facts themselves are loaded with tints and hues and colorations, that they are porous and open to multiple, ambiguous interpretations, that they are only facts when their non-objective qualities are factored in.

In many cases dangerous data is worse than no data at all; knowing a little about something can be worse than knowing nothing at all because the little we know can prime our responses and discourage us from truly finding-out. The reason for is the ways in which humans think, the underlying motivations behind getting decision making processes right, and the cognitive biases informing what we do.

Sperber and Mercier (2017) have argued that human reason is an evolutionary response to the social nature of human life; when we reason we are not necessarily motivated by the desire to uncover the truth of a situation, but rather to construct effective arguments to persuade others of a particular truth that is to our advantage. Our decision-making processes are therefore not necessarily determined by the efficacy of the outcomes of those decisions, but also by our ability to justify those decisions to our social group, and the advantages which those decision confer. We are programmed to prefer convincing justifications and good arguments over the truth, and no matter how aware we become of this tendency, it is impossible to fully inoculate ourselves against the cognitive biases that inform out decision making.

Nickerson has observed that "there is an obvious difference between impartially evaluating evidence in order to come to an unbiased conclusion and building a case to justify a conclusion already drawn" (1998: 175). While the difference is obvious, it is not always easy to tell which of these we are engaged in; we can easily persuade ourselves that we are impartially evaluating evidence while inadvertently pursuing a particular agenda. This emphasises this risk posed by Sperber and Mercier (2017) contention that case-building of this kind is a fundamental part of human reasoning. Workplaces are social environments, but they are also often competitive environments where individuals collaborate towards particular ends but also pursued their own self-interest. The decisions we make in these kinds of contexts sometimes reflect the collective good, but at other times are more selfish. We may therefore prefer bad decisions that are easy to justify, or that have less direct negative consequences, over good decisions that are more difficult to justify.

It is precisely because of that innate tendency towards self-justifying arguments that dangerous data can lead us to bad decisions. It can act to depersonalise the decision making process, undermining collegiate work and team-building. It can inoculate people from the anxiety that their decisions may be subsequently questioned, second-guessed, or disputed, providing a justification for those decisions that discounts the wider social context of them. The proliferation of the production and retention of data in every aspect of our working lives, and the re-presentation of this data through business analytics tools such as Business Intelligence software, data analytics software, and data visualisations, provides potential justifications for decision making regardless of the quality of those decisions in their own terms. But the biggest problem of all is that most of us have a very poor intuitive sense of numerical data, and all of us intuitively over-estimate the significance of patterns. Therefore while the use of data has proliferated across all levels of commercial organisations, the ability of individuals to draw reliable inferences from that data has not kept up.

Our tendency to draw strong inferences from limited information is poorly adapted to the environment of data dashboards, large data sets, and data visualisations. While we have developed techniques to address these limitations to our cognitive processes and to correct for our poor intuitive senses, everyday business contexts militate against the robust analysis of the data that we use. If we are unwary we can be led astray by dangerous data in the most treacherous of way, not only drawing the wrong conclusions, but convincing ourselves in the process of the rigorous and impartial nature of our own reasoning. But if the data-driven business is here to stay, it is important that we understand some of the problems that our data-hoarding tendency can generate, and how those problems can be mitigated. The final parts of this paper will address some of the major things than go wrong with our use of data in commercial contexts, and the role of information professionals in addressing these issues.

3. The shape of dangerous data:

The ideal outcome is that data is used in organisation where appropriate to inform decision making, providing an evidence-base for those decision and helping to drive performance and efficiency gains, but that where data contributes to worse decisions this is recognised and mitigated. This ideal depends on data that is accurate, reliable, and many organisation expend considerable effort in trying to ensure this. But it also depends on the appropriate use of data, and that is harder to control. This section unpacks some of the key ways in which the use of data in decision making can go wrong, and what we can do about it. In broad kinds these risks are of three different kinds: problems with the data itself, problems with the purposes to which we put it, and problems with the inferences that we draw from it. All three play a role in contributing to dangerous data, but this paper focuses principally on the role of information behaviour and cognitive bias in understanding how individuals seek-out and exploit data in the workplace. It argues that the ways in which we generate and present data in commercial contexts can sometimes inadvertently contribute to bias in our uses of that data.

Cognitive biases describe common patterns in our processing of information. That is not to say that every time we process information we are subject to bias, or to imply an intention to misunderstand or mislead. Rather bas describes unconscious inclinations that “can affect how humans search for and process information” (Behimehr & Jamali, 2020). Arnott (2006) has identified 37 cognitive biases that have an impact on the development and implementation or decision systems but this paper focusses on just a handful of significant cases. Information behaviour on the other hand describes the habits individuals develop in their use of information, or the “complex human information-related processes that are embedded within an individual’s everyday social and life processes with evolutionary and developmental foundations” (Spink & Heinstrom, 2011: xvii). Important and well-studied aspects of this are information needs and information seeking, and these inevitably have an impact on how individuals use information systems. There is a close connection between information behaviour and cognitive bias, but nevertheless they also sometime are expressed differently both in research and in our working experience. While there are many ways in which our habits of action and thought can lead to inappropriate use of data, there are also a handful of key issues explored below.

3.1 Availability (cognitive bias): availability bias is probably the most significant cognitive bias in any kind of analytical or evaluative process. It is all-encompassing, difficult to recognise, and difficult to fully avoid in everyday contexts; availability bias sneaks up on us when we least expect it. While not noted for his contributions to epistemology, the US Secretary of Defence Donald Rumsfeld inadvertently touched on discrepancies in how we regard available evidence in formulating our understanding of an issue in a statement made in 2002 about the lack of concrete evidence for Weapons of Mass Destruction in Iraq. He suggested:

> As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult ones (Ratcliffe, 2017).

This was a widely discussed and sometimes mocked statement (Dunning, 2011; Logan, 2009). But while lacking clarity it highlights something important about the way in which we evaluate evidence: we tend to give greater credence to the information we are aware of and discount the potential significance of information that we may yet have to find. We go wrong when we fail to take account of unknown-unknowns.

Availability bias describes the tendency for people to over-estimate the significance of evidence that is more available to them, and under-estimate the significance of evidence that is harder to come by. More available simply means evidence that comes more easily to hand, either from memory, or by its ready accessibility. In a very practical sense this may include for example overestimating the significance of the initial results of a search, and stopping considering additional evidence too early. But availability bias also relates to the way in which we interpret the world; we are more likely to credit personal experience over the experience of others simply because personal experience is more resonate and therefore tends to come to mind more easily. We are most likely to credit recent experience as it tends to be more resonate. We are more likely to trust information generated within an organisation, than information that comes from outside of that organisation because we are likely to be more familiar with the former and have it closer to hand. We are likely to over-emphasise the significance of data that we use every day, over overlook or discount data that we use or encounter only occasionally.

Availability bias is a significant issue in the corporate use of data because decision-making will tend to be skewed towards that data which is most easy to access and with which we are most familiar, to the extent that individuals may not look for data to disconfirm their initial appraisal or look for data that is more appropriate for the task at hand in the first place. Indeed, individuals may discount data that is more evidential simply because they are less familiar with it. This is not simply a matter of the data that an organisation collates about its own activities, but also a matter of the subsets of that data which it chooses to prioritise. Often the data to disconfirm a particular interpretation may well be available to the organisation, but not fully considered in the decision-making process simply because it is less available to those making decisions. The choices that an organization makes in the ways in which it presents data to employees will influence what they do with that data subsequently.

Corporate dashboards and KPIs are a good example of the manufacturing of availability bias within corporate contexts. Both the selective presentation of data in a easy to use and visually appealing environment, and their ease of use and widespread availability create the conditions in which the data presented in a dashboard will be over-emphasised in corporate decision making, at the expense of data that in specific context may be more useful or relevant. This headline data becomes the known-knowns or Rumsfeld's statement, and skew subsequent decision making. While the underlying assumption in the implementation of data analytics and business intelligence systems is that users will where appropriate go behind headline-figures, to seek out the known-unknowns, and perhaps encounter the unknown-unknowns, the relative availability of these headline figures will invariably lead to an over-estimation of their significance whether or not individuals choose to go further.

This potential for bias is exacerbated when measures are selected by the organisation as of particular significance. It is important to measure outcomes, and well-selected KPIs can be an important tool in achieving this. But once established they can also drive decision making within an organisation, simply because of their high level of visibility and priority. Furthermore KPIs can outlast their useful function because once established organizations and groups can be reluctant to go beyond them, as they provide a useful measure against which to validate performance. One classic example of this is the Citizen's Charter, launched in the UK in the mid-1990s with the aim of improving public services and the accountability of public authorities. The charter included measuring public services against targets, for example response times to public enquiries. Instead of improving services these sometimes distorted the priorities of public authorities towards meeting those targets rather than addressing the underlying service-issues they were intended to measure, to the extent that the targets sometimes ended-up measuring only the service provider's ability to meet targets, rather than any underlying performance. In an analysis of the legacy of the Citizen's Charter, Deakin has written that:

> Targets deriving from information collected about service delivery have proved inflexible in practice, cramping and confining the activities of agencies and imposing excessive costs of collection (2009: .

While overall the Citizen's Charter was a mixture of successes and missed opportunity, it became widely associated with this distortion of public service priorities, and the use of targets and comparative measure of performance were rarely associated with improvement in services.

When we are dealing with data within the organization, therefore, it is important to ensure that we are not -over-relying on data that is easy to access or available, and are always looking beyond both our initial assumptions and existing data sets to question whether other sources of information are available that might make a material difference to the choice we are making, whether that information is from inside or outside the organisation. More importantly it is important that organizations think strategically about the presentation of data to employees, to avoid limited data sets becoming too over-familiar and over-relied upon. This may include the use of more flexible, personalised, or context-sensitive data interfaces that avoid re-enforcing the same measures repeatedly and reflect the underlying significance of the data itself. In recent years business intelligence systems have moved towards these more flexible, context driven, and personalized approaches. Finally it is important that Key Performance Indicators both measure something useful, and also are used in appropriate contexts, both of which should be kept under review.

3.2 Confirmation (cognitive bias): confirmation bias is our tendency to seek out and over-estimate the significance of information that re-enforces a pre-existing belief, and overlook, downplay or under-estimate the significance of information that confounds that belief. While it is possible to become aware of the values and beliefs that drive our interpretation of available evidence, at the same time it is difficult to remain cognizant of all of the ways in which we apply those values and beliefs in our work, since much of this is unconscious. Therefore confirmation bias is difficult to avoid, and something to which we are all prone much of the time. Furthermore knowing that we are prone to confirmation bias does not necessarily help us avoid it.

Confirmation bias is relevant to the use of data within organizations because data can be used to confirm existing beliefs, values, or messages. Thus while data can provide an evidential basis on which to challenge poor practice and transform business processes, it can also provide an evidential basis on which to defend poor practice and ward off unwanted change, or an evidential basis in which to undermine excellent practice and implement unnecessary or unwarranted changes. The efficacy with which data is used can often reflect where data analytics skills reside, rather than what the data itself reveals. Nickerson has highlighted the difference between impartially evaluating evidence, and building a case to justify a conclusion already drawn (1998: 175), and adds: :

> In the first instance one seeks evidence on all sides of a question, evaluates it as objectively as one can, and draws the conclusion that the evidence, in the aggregate, seems to dictate. In the second, one selectively gathers, or gives undue weight to, evidence that supports one's position while neglecting to gather, or discounting, evidence that would tell against it (Nickerson, 1998: 175).

This highlights the ways in which evidence can often become a means to confirm existing assumptions, and in the context of business systems can also become a means to justify existing behaviours or work processes.

When individuals come together in teams confirmation bias sometimes emerges in **groupthink,** in which social groups seek to minimise disharmony and disagreement by conforming to shared beliefs and values that largely go unchallenged. One of the way in which we can see to become aware of and address confirmation bias is to try to disconfirm something we hold to be true or believe. This evaluation can be built-in to business processes by, for example, ensuring that the evaluation of decisions takes place outside of the original group context within which those decision were taken.

3.3 Anchoring (cognitive bias): anchoring bias is the tendency for a single piece of information, very often the first information we might encounter on a topic, to provide an anchor against which subsequent information is evaluated. For example, if the first thing we read about a product is a glowing review, we are likely to anchor any subsequent information against that initial view.

Anchoring bias can become particularly strong when we tend to follow the same kinds of processes in researching a particular topic. For example when conducting an annual review, or undertaking standard research, the first step of the process can have an influence on how we interpret everything that follows. A member of staff who is performs well against the first criteria may escape attention for subsequent performance issues, where as a member of staff who performs poorly against the first criteria may have their other strengths overlooked. When considering job applications and interview performance, the first candidate may set an informal benchmark against which other candidates are evaluated. The order in which we process information is therefore important to the ways in which we form an opinion around that information. And this is another example of where corporate dashboards and business intelligence systems can create unintentional bias within the use of data in the organization. In simple terms the presentation of headline data frames the subsequent analysis of more detailed data sets. The structured nature of data aids data discovery, but also constructs a narrative about the relationship between its parts.

Ni et al (2019) have explored the role of anchoring bias in business intelligence systems, and suggest that “anchors are very common in BI use contexts and they can create significant challenges for the effective use of BI” (Ni et al, 2019). Nevertheless they also found that in some context BI systems can help mitigate the anchoring effect:

> the use of a BI system to support decision making did mitigate the effect of a spurious anchor. Using a BI system demands the expenditure of cognitive effort. A user needs to consciously think about the decision context and callup and analyse data relevant to the decision task. In this expenditure of cognitive effort decision-makers move from System 1 to System 2 processes. In the case of a spurious anchor, this effort is sufficient to negate the anchoring effect (Ni et al, 2019).

They recommend forewarning users about anchoring as an approach to tackling the potential bias. , Because anchoring bias in exacerbated by our habits, making a conscious effort to change the order in which we research a particular topic or to ignore obvious sources until towards the end of the process in one way to address the influence of anchoring bias. For example when conducting a review of particular teams or individuals against key performance indicators over a given period, we might choose to gather first rich information about that performance before considering KPIs to ensure that KPIs are not driving the evaluation. This has implications for the ways in which we design and implement data analytics systems in the organisation. For example we might choose to vary which data is highlighted in dashboards, and which kinds of data visualisations are used to avoid those visualisations from becoming anchors for subsequent analysis and use.

3.4 Principle of least effort (information behaviour): George Kingsley Zipf has exerted a quiet influence on information and library management through his eponymous “law” of frequency distributions. Born in Illinois in 1902 and interested in linguistics and philology, Zipf studied the statistical characteristics of different human languages. His observation that in any language a few words are used very often while very many are used on occasionally or almost never formed the basis of a more general rule: that in many instances the rank versus the frequency distribution in lists of items approximates to the same general principle. This is generalised in library and information management as the 80/20 rule of thumb: that 80% percent of the use of an information resource such as a library will be for just 20% of that resource, where as 80% of that resource will only be used 20% of the time. But it is for another related contribution that Zipf features in this paper: the formulation of the principle of least effort in his 1949 book of the same name, which explained his eponymous law as a consequence of reducing cognitive overheads in communication.

Although largely ignored for the first ten years after publication (Chang, 2015), the principle of least effort subsequently had a significant influence on the understanding of information seeking behaviour. Chang suggest that its influence has been particularly amongst research into information behaviours seeking to explain reasons for the selection and use of information (2015); indeed Bates (2017) has described the principle of least effort as one of the most established concepts in information behaviour research. Tubachi summarise the role of the principle of least effort on information seeking behaviour:

> With respect to information seeking, the principle of least effort postulates that the information seeker chooses a course of action that will involve most convenient search method for information seeking. The user will apply the searching tools that are most familiar and easy to use so as to find results. This happens in spite of the user having proficiency in technical searching (2018)

The ways in which we make data available to individuals within the organization will therefore tend to have an influence on the subsequent use of that data. While business intelligence systems, for example, make it relatively easy to access, process, and analyze data from a wide range of sources, that ease is not equally distributed across the entire system; headline data presented in dashboards will tend to have a pull effect receiving disproportionate emphasis because of the effort required to go beyond those headlines. That means that for example conclusions that are easy to justify by headline data about the functioning of the organisation, but that are falsified or problematised by analysis of the underlying data sets, may go unnoticed.

Related to this is Mooer’s Law, that information systems will tend to be used only when the pain and trouble for someone cause by having the information is outweighed by the pain and trouble of not having it. This is a kind of head-in-the-sand law, where sometimes it is better just not to know. It is worth quoting in full Mooer's observation about the cognitive burden created by information as it is still relevant for the role of data in the workplace:

> If you have information, you must first read it, which is not always easy. You must then try to understand it. To do this, you may have to think about it. The information may require you to make decisions about it or other information. The decisions may require action in the way of a troublesome program of work, or trips or painful interviews. Understanding the information may show that your work was wrong, or that your boss was wrong, or may show that your work was needless. Having information, you must be careful not to lose it. If nothing else, information piles up on your desk—unread. It is a nuisance to have it come to you. It is uncomfortable to have to do anything about it. Finally, if you do try to use the information properly, you may be accused of puttering instead of working. Then in the end, the incorporation of the information into the work you do may often not be noticed or appreciated (1996).

Thus he argues that "not having and not using information can often lead to less trouble and pain than having and using it" (1996). Summit has summarise this concisely: "Mooers' law tells us that information will be used in direct proportion to how easy it is to obtain" (1993: 16; cited by Austin, 2001). But there is more to Mooer's law than this – it suggests that people will tend to stop looking when continuing looking threatens to create additional problems, such as for example falsifying conclusions they have already drawn, or undermining a position they have already adopted. As we have already seen, on the whole when making decisions individuals look for reasons to justify those decision not information that might put those decisions under question. Both the cognitive overheads *and* the potential addition work created by looking beyond our initial assumptions tend to discourage individuals from digging too deep into data sets. The fact of making data available is not enough to drive improvement if employees are reluctant to engage with that data.

And this can drive a defensive use of data, where individuals make a greater effort to dispute a given interpretation than they do to confirm a given interpretation. Ironically perhaps the proliferation of data in the workplace has given individuals the resources to defend outdated practice, bad decisions, and problematic work areas by the selective use of that data, simply because individuals are likely to put more effort into using data in ways that is in their own interests than in using data in ways that is against their own interests, or simply neutral.

3.5 Grievous granularity (cognitive bias): relatively few of us could confidently select the correct test of statistical significance for any particular set of data, and subsequently reliably apply that test. More pertinently even if we could, in normal everyday business contexts we rarely do apply these kinds of significance tests unless they are absolutely required. It would be absurd to do so. Not only would the benefits be marginal in most cases, but also the time involved would be prohibitive. On the whole when we are considering data within an organisation, we tend to rely on intuitive notions of significance, and over-rely perhaps on apparent trends or apparently significant changes. We apply rule of thumb principle to our judgement, intuitively understanding that larger data sets are less influenced by random variation and are therefore generally more reliable, but that more care has to be taken in interpreting the significance of smaller sets of data. In many contexts this rule of thumb is sufficient. However in combination with the anchoring effect discussed above, the ability to use business intelligence systems to filter and drill-down into data sets can have significant consequences for the ways in which we rapidly evaluate data in real-time contexts.

One of the problems of data granularity is a form of extension neglect, particularly **sample size insensitivity**. We may know the importance of considering sample sizes, but at the same time neglect to fully account for or notice that factor in evaluating data subsets, particularly if our interpretation is already anchored by the originally larger data set from which our subset is drawn. University systems now commonly allow the analysis of student performance, for example, down to individual course or class levels against a wide range of demographic factors but at this level of granularity the data has no significance, and a single case can make a significant difference to perceptions of the data. The data is useful and significant only in aggregate. Universities are generally well-stocked with people trained in the importance of robust sample selection and the dangers of drawing inferences from unrepresentative data, and nevertheless the use of data of this kind in planning is commonplace. It is commonplace because it serves a social function. Similarly reporting of national surveys and auditing processes such as the National Student Survey or the Research Excellent Framework frequently places significance on small differences that can be accounted for entirely by differential sample sizes. These issues should be and are well understood and yet also routinely ignored because no matter how well we may understand the importance of significance, we still habitually overlook it when convenient, to our advantage, or when seeking to build a particular case.

Sample size insensitivity is not just a matter of significance, but also how we understand and interpret data. In general terms our intuitive sense of probability often does not reflect the real probabilities of unknown events. A good illustration of this is the famous Monty Hall problem, summarised by Gill (2011) as follows:

> Imagine you are a guest in a TV game show. The host, a certain Mr. Monty Hall, shows you three large doors and tells you that behind one of the doors there is a car while behind the other two there are goats. You want to win the car. He asks you to choose a door. After you have made your choice, he opens another door, revealing a goat. He then asks you whether you want to stay with your initial choice, or switch to the remaining closed door. Would you switch or stay? (Gill, 2011).

Most people's intuitive sense is that there is no advantage to changing your mind at this point, but that is incorrect: switching doors improves your chances of winning the car. If we imagine 1000 doors instead of 3, and opening 998 of them instead of 1, it becomes intuitively obvious that the car is highly unlikely to reside behind our original choice of the remaining two unopened doors. This is the same problem but the switch in frame of reference changes our intuitive understanding.

Sample size insensitivity is a particular issue when moving from larger data sets to smaller subsamples, and drawing inferences about those subsamples because of the anchoring effect that the initial data can have. For example when looking at aggregated growth trends and then drilling down to component parts of those trends, we might not recognise that the greater volatility within the smaller sample may simply be a reflection of the sample size and of no significance. But it is precisely the possibility of filtering data that is an important part of the power of Business Intelligence systems.

3.6 Dangerous data visualisation (information behaviour): the power of corporate data visualisation tools and business intelligence systems such as Microsoft's Power BI comes from their customisability and usability. Data visualisations allow complex data to be rapidly assimilated into work. Data visualisation communicate ideas in a powerful and direct way and that is their key function. But that power and persuasiveness of a good diagram or chat also comes with risks, as data visualisaton can both mislead, and anchor particular interpretations. The classic case study of the ways in which data visualisation can drive interpretation is the use of polling data in election leaflets, which have often exploited manipulated scales or selective data to skew understanding. While of course the intention of business analytics is not to mislead in this direct way, the inferences drawn from different kinds of data visualisation will tend to vary, particularly if those are not subject to robust analysis as is common in time-sensitive business contexts. Data visualisations are powerful, and communicate directly and rapidly, in ways that perhaps discourage fuller critical engagement with and consideration of the underlying data sets.

This is particularly true of the kinds of data visualisations that are used as a part of dashboards, simply because by definition these use arbitrary frames that do not directly reflect the uses to which that data will be put because the purpose of the dashboard is to provide a rapidly assimilable overview. For example our interpretation of an apparent trend might vary significantly when presented over two, three, five or ten years periods; the consideration of the timescale is important for identifying trends and understanding their significance. The arbitrary time scales used in dashboards may therefore have an influence on how that data is subsequently used. This is particularly because of the anchoring bias discussed above; the initial presentation of data has a disproportionate significance to subsequent decision making.

4. The role of information professionals in addressing dangerous data

The good news is that information professionals are in an ideal position to address the proliferation of dangerous data in commercial organisations. In recent years the BIR Annual Survey (Phillips, 2022; 2021; Carter, 2019) has revealed a demand for informant workers with business savvy, with creativity, and with analytical, evaluation, and presentation skills. The trend in the professional has been a move away from the information profession as a guardian and gatekeeper of information, organising the collection, and undertaking research to find high quality and relevant sources. In many contexts information professionals are now more actively involved in processing information, in presenting it within accessible formats, and in drawing inferences from the information that they uncover. The commercial sector has led the way in many of these changes, with commercial information centres becoming not just gateways to resource, but active in the generation of corporate knowledge. While information work increasingly involves data analytics roles, it is not merely in understanding data quality and significance that information professionals can have a part to play in improving the use of data in the organisation. It is also in understanding the social contexts within which information in generated, circulated, presented and used. Information science has taught us that the otherwise rational actors of classical economics do not always act rationally in relation to their information needs and behaviour. And it is in real-life information behaviour that the ideals of data driven approaches run aground.

Arnott (2006) emphasises the importance of moving decision-making from system 1 to system 2 processes. The terms system 1 and system 2 was popularised by Daniel Kahneman popular psychology book *Thinking Fast and Slow* published (2011), but the differentiation between two modes of cognitive engagement goes back several decades before this. According to Kahneman, system 1 cognitive processes are rapid, intuitive, and instinctive, where as system 2 cognitive processes are slower, more deliberative, and take more effort. Many of the decisions taken in commercial context rely on system 1 processes, but while system 1 is often surprisingly reliable, it is also plagued by the kinds of cognitive bias we have discussed above. Ni et al (2019) suggests that by moving corporate decisions from system 1 to system 2 processes more reliable and useful decision-making takes place. We cannot eradicate the influence of bias and behaviour, but we can shift focusses to ensure that bias and well-worn behavioural habits have a reduced influence on outcomes.

One way to do this is by basing decision in deliberative processes, and this is part of what underpins the rise of data and data analytics in the first place. The use of data moved decision-making out of professional experience and corporate instinct into a more deliberative process. But for data to be useful we also have to understand how people engage with it in real-world contexts, and that involves understanding the culture, context and practices of each individual organization. Many organisations put significant effort into ensuring the robustness of their data analytics, and expend considerable resources on cleaning-up messy real-world data sets and ensuring their validity; less effort is perhaps spent on understanding the ways in which data in produced and more importantly exploited. To draw an analogy, In 2004 I made the point that for many organisations, implementing intranet solutions was seen as the end of process, and few organisations considered how those systems would be used in practice to support business function and as a consequence intranets did not realise their potential (Tredinnick, 2004). The implementation of Business Intelligence Systems risks repeating aspects of this mistake, focussing on the system itself and data quality rather than the real-life contexts in which it is used and the benefits that accrue.

In the 1990s and early 2000s the idea of the information audit spread in information science and information management. The information audit was a process of understanding the information sources held by an organisation – both internal and external – and the business processes in which those sources were both generated and used. This combination of source and context was integral to the way in which information audits were designed to aid organizations, identifying under-utilised or ignored resources, bottlenecks in business processes, duplication of effort and redundance, and seeking to streamline business processes. As information has proliferated the idea of the information audit has become perhaps a little naïve, however the techniques of information auditing can also help us grasp the role of data withing the organisation. While the idea of data auditing has been around for a little while, in recent years it has most commonly become associated with GDPR compliance. This is undoubtedly an important consideration when thinking about the role of data within the organization, however auditing data within the organization can also help with understanding the processes within which the data that we rely on is generated, and ensuring the robust and appropriate nature of that data that accrues. More importantly it can help organizations to understand how data is actually used – not how it *should* be used – as a part of the everyday working processes. There are two aspects to this: ensuring that the data driven business is built with data in mind, and ensuring that data is used in ways that are genuinely productive.

Many organizations rely on data scraped from existing business processes, rather than re-designing business processes to ensure they generate data that is genuinely robust, reliable and useful for driving commercial decisions. Understanding the role of data within the business is a matter of understanding both the way in which it is produced. Ensuring the effectiveness of data driven approaches means ensuring that the processes which produce the data on which subsequent decisions are made are producing data of a kind and in a form which is genuinely useful to those subsequent business processes, and that is reliable in the ways that it *appears* to be reliable. Simply aggregating existing data sources is not necessarily enough to achieve this aim, and the ease of doing so distracts from the more difficult task of properly aligning process and outcome. One role of a data audit is to map-out this relationship to allow the redesign of business processes.

The second function of a data audit is however if anything more important – to understand how the data generated within and organization is actually used, the processes which it supports, and the decisions that it influences. The ways in which organizations plan to use data within the decision making process may not reflect the real-world uses of the resources that they make available. Indeed many of the truly beneficial uses may well be largely unrecognised. Understanding exactly how individuals incorporate different data sets into their working lives is an important part of understanding how to ensure that data becomes an integral part of business processes, works towards improving those processes, and of avoiding inadvertent misuses of data of the kind we have described above as dangerous data.

The ideal outcome of the data-drive organization is that data is used where appropriate to inform decision making, providing an evidence-base for those decision and helping to drive performance and efficiency gains. This depends on data that is accurate, reliable, and appropriately used. But it also depends on the social processes within which work is embedded, and the information behaviours that inform individual working lives. While a focus on data science can help ensure the robust, accurate and evidential nature of data withing the organization, to ensure its appropriate incorporation into business processes we need to go beyond that to understand the ways in which individuals relate to, and use the information around them as a tool, as a defence, and sometimes as a weapon. This paper has argues that the field of information science has already mapped this terrain, albeit in insufficient detail, through decades of research into information behaviour, and developed the methodologies to understand the ways in which information is exploited in the wild.

References:

1) Arnott, D. (2006). Cognitive biases and decision support systems development: A design science approach. Information Systems Journal, 16(1), 55–78.
2) Austin, B. (2001), Mooer's Law: In and Out of Context, *Journal of the American Society of Information Science and Technology,* 52 (8): 607-609.
3) Bates, M. J. (2015), Information Behaviour, in: *Encyclopedia of Library and Information Sciences, Fourth Edition,* London: Taylor and Francis, 2074-2085
4) Behimehr, S. & Jamali, H. R. (2020), Relations between Cognitive Biases and Some Concepts of Information Behavior
5) Brownlow, J., Zaki, M., Neely, A. & Urmetzer, F. (2015), Data and Analytics - Data-Driven Business Models: A Blueprint for Innovation The Competitive Advantage of the New Big Data World. Cambridge: Cambridge Service Alliance, available at: https://cambridgeservicealliance.eng.cam.ac.uk/system/files/documents/2015MarchPaperTheDDBMInnovationBlueprint.pdf [accessed: 12 February 2023]
6) Brynjolfsson, E., Hitt, L. & Kim, H. (2011), How Does Data-Driven Decision-Making Affect Firm Productivity and CEO Pay? Available at SSRN: https://ssrn.com/abstract=1819486 [accessed: 12 February 2023]
7) Bulger, M., Taylor, G. & Schroede, R. (2014), Data-Driven Business Models: Challenges and Opportunities of Big Data, Oxford Internet Institute, available at: http://www.nemode.ac.uk/wp-content/uploads/2014/09/nemode_business_models_for_bigdata_2014_oxford.pdf [accessed: 12 February 2023]
8) Chen, h., Chiang, R. H. L, & Storey, V. C. (2012), Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly*, 36 (4): 1165-1188.
9) Carter, D. (2019), New technologies and new data sources: The business information survey 2019, *Business Information Review,* 36 (3): 95-105.

10) Chang, E (2006), Advanced BI Technologies, Trust, Reputation and Recommendation Systems, The 7th Business Intelligence Conference, Sydney, Australia.
11) Chang, Y (2015), Influence of human behavior and the principle of least effort on library and information science research, *Information Processing and Management,* 52 (4): 658-669.
12) Cuckier, K. (2015), The Economist's Data Editor on Data Fetishism, *What's the Big Data,* available at: https://whatsthebigdata.com/2015/10/30/the-economists-data-editor-on-data-fetishism/ [accessed: 12 February 2023]
13) Deakin, Nicholas (2009), Citizens, Charters and Concordats : how the Major Government confronted the challenge of engagement with civil society in public sector reform, . Observatoire de la société britannique, n. 7: 49-61.
14) Dunning, D. (2011), The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance, *Advances in Experimental Social Psychology,* volume 44: 247 – 296.
15) Gill, R. H. (2011), The Monty Hall Problem, *International Encyclopedia of Statistical Science.*
16) Hopkins, B., Schadler, T. & McCormick, J. (2018), Insights-Driven Businesses Set The Pace For Global Growth, Forrester, available at: https://www.forrester.com/report/insightsdriven-businesses-set-the-pace-for-global-growth/RES130848 [accessed: 12 February 2023]
17) Kahneman, D. (2011), *Thinking Fast and Slow,* London and New York: Penguin Books.
18) Kramer, D. (2012), The Fetishization of Data, the data-reality conflation and the role of the digital humanities, available at: https://www.michaeljkramer.net/the-fetishization-of-data/ [accessed 12 February 2033]
19) Lieb, D. (2013), Cognitive Overhead, Or Why Your Product Isn't As Simple As You Think, *Techcrunch,* available at: https://techcrunch.com/2013/04/20/cognitive-overhead/ [accessed: 12 February 2023]
20) Logan, D. C. (2009), Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry, *Journal of Experimental Botany,* 60 (3): 712 – 714.
21) McParland, C. & Connolly, R. (2020), Dataveillance in the Workplace: Managing the Impact of Innovation, *Business Systems Research,* 11 (1): 106-124.
22) Mooer, C. N. (1986), Mooers' Law or Why Some Retrieval Systems Are Used and Others Are Not, *Bulletin of the American Society for Information Science and Technology,* 23 (1): 22-23.
23) Moss, L. & Hoberman, S. (2004), The Importance of Data Modeling as a Foundation for Business Insight, Teradata.
24) Negash, S. (2004), Business Intelligence, *Communications of the Association for Information Systems,* Vol 13, Article 15.
25) Ni, F., Arnott, D. & Gao, S. (2019), The anchoring effect in business intelligence supported decision-making, *Journal of Decision Systems*
26) Nickerson, R. S. (1998), Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, *Review of General Psychology,* 2 (2): 175 – 220.
27) Patel, J. (2019), BRIDGING DATA SILOS USING BIG DATA INTEGRATION, *International Journal of Database Management Systems,* 11 (2/3): 1-6.
28) Phillips, S. (2021), 'Never let a good crisis go to waste' – The 2020 Business Information Review Survey: Part I, *Business Information Review,* 38 (1): 8-14.
29) Philips, S. (2022), Consolidate gains, implement more change – The 2021 business information review survey: Part I, *Business Information Review,* 39 (3): 71-78.
30) Ratcliffe, S. (Ed.)(2017), *Oxford Essential Quotations; 5th edition,* London and Oxford: Oxford University Press.
31) Sadowski, J. (2019), When data is capital: Datafication, accumulation, and extraction, *Big Data and Society,* Jan-June: 1-12.

32) Sperber, D. & Mercier, H. (2017), *The Enigma of Reason: a new theory of human understanding,* London: Allen Lane.
33) Spink, A. & Heinstrom, J. (2011), *New Directions in Information Behaviour,* Bingley: Emerald Publishing Group.
34) Summit, R. (1993), The future of online information: challenges and opportunities, *The Electronic Library,* 11 (4/5): 233-235.
35) Tredinnick, L (2004), *Why Intranets Fail (and how to fix them),* Oxford: Chandos Publishing.
36) Tubachi, INFORMATION SEEKING BEHAVIOR: AN OVERVIEW, available at: https://www.researchgate.net/publication/330521546_INFORMATION_SEEKING_BEHAVIOR_AN_OVERVIEW [accessed: 12 February 2023]
37) Yu, S., Qing, Q., Zhang, C., Shehzad, A., Oatley, G. & Xia, F. (2021), Data-Driven Decision-Making in COVID-19 Response: A Survey, *IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS,* 8 (4): 989 – 1002.
38) Zipf, G. K. (1949), *Human Behaviour and the Principle of Least Effort,* Cambridge, Massachusetts: Addison-Weasley Press.