European Review, Vol. 27, No. 3, 357–377 © 2019 Academia Europæa
doi:10.1017/S1062798719000048

Big Data and Us: Human–Data Interactions

BARRY C. SMITH

Institute of Philosophy, School of Advanced Study, University of London, Senate House, Malet Street, London WC1E 7HU, UK. Email: Barry.Smith@sas.ac.uk

The growth of continuously generated, large-scale datasets, and new analytics to handle them, has created expectations, in some quarters, that new insights can be generated that will help us address the biggest challenges that face us as a species and therefore can shape future societal outcomes. It is hoped that these new technologies will lead not just to new discoveries but also to new questions and thinking that will deliver significant scientific advances. Perhaps there will be some genuine scientific advances but since many of the challenges that face us reside in the human world and depend upon how humans behave, we need to turn to the humanities and the social sciences as well as the natural sciences and look at the role Big Data could play there in adding to, or shaping, our future. And here, what concerns us is not just the assumptions guiding the new analytic techniques for data mining, data merging, linking and analysis, born out of smarter AI algorithms, it is a more fundamental issue about the constraints and limitations of the kind of data inputs and outputs being appealed to in Big Data systems and whether they are well served to provide an understanding of the human world.

First Things First: Big Data, What Is It?

Big Data is defined as datasets growing so large in volume, and with such velocity that that they cannot be handled by traditional analytic and data management tools. And it is not just the volume and velocity of data increase, but its variability that we must consider, especially in the humanities. In many sciences, even very large datasets are well-behaved and regular in format, and it is just the nonlinear speed of their growth that demands handling methods. However, radically new tools and technologies are needed most when dealing with domains beyond the sciences, where the data are messy, dynamic and varied. We produce more blogs, emails and comments on social media websites than ever. We use our smart phones to take and upload photographs, to record and transmit, perhaps unknowingly, information about

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

our location and interests. Google logs the websites we browse, and Apple the music we download. Credit card companies log the purchases and trips we make, and all this information, including the additional storage and analysis of it by companies, is piling up into datasets that someone, somewhere, wants to mine for vital predictive patterns of social or musical trends, purchasing intentions, transport schedules, group behaviour, climate impact, health and well-being. Unless you have chosen to be a recluse, or are simply left out because you live on the wrong side of the digital divide, you are part of Big Data already.

Huge amounts of data are continuously generated and it is the aim of the analysers of Big Data to offer something exhaustive in scope, where n = all, and where 'the data are flexible and scalable'. This vast proliferation of data of all kinds affects the way information is gathered, organized and used:

> Traditionally, data analysis techniques have been designed to extract insights from scarce, static, clean and poorly relational datasets, scientifically sampled and adhering to strict assumptions (such as independence, stationarity, and normality), and generated and analysed with a specific question in mind.[1]

By contrast:

> The challenge of analysing Big Data is coping with abundance, exhaustivity and variety, timeliness and dynamism, messiness and uncertainty, high relationality, and the fact that much of what is generated has no specific question in mind or is a by-product of another activity.[2]

Machine-learning algorithms have been developed to deal with this new situation. Deep learning algorithms in particular offer the promise to 'automatically mine and detect patterns [in the data] and build predictive models and optimize outcomes'.[2]

Big Data: What is it Good For?

The promised advances in the knowledge gained by making use of Big Data tend to focus on key examples, such as retail purchasing, email logs, barcodes and travel pass scanning, social media posts or responses. These can be cross-referenced to provide massive relational datasets. This continuous, dynamic accumulation of information can be used to study the movement and actions of people in large cities in order to guide future urban planning or retail decisions.

Who is Gathering So Much Data and How Do They Want to Use these Data?

In 2012, Wal-Mart reported generating 2.5 petabytes of data every hour related to transactions of over a million customers,[3] and Facebook reported processing 2.7 billion 'Like' actions and 300 million photo uploads per day.[4]

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

We now know that Facebook has also given access to other players to use large relational datasets, and for those who know how to analyse these data – and have access to the data – it offers a huge amount of new information with which not only to study behaviour and attitudes, but also to attempt to influence them. Data harvesting from Facebook user profiles, without the users' permission, and reportedly without the knowledge of Facebook, has been done, allegedly, by the company Cambridge Analytica in an attempt to influence outcomes in elections.

In other domains there is cluster analysis of how income and health are related, predictions of crime hotspots, better ways to schedule flight times, to design food products and target goods and services to groups of consumers. However, beyond the specific examples of success, which are mostly commercial – although some are to do with crime-fighting in cities – there are still questions about the efficacy of the widespread and often overblown claims of being able to predict and influence people's behaviour.

Beyond these recognized concerns, there is still far too little foundational thinking about Big Data and the value of its widespread application to different facets of the human world. Research in the humanities often concentrates on specific, small-scale objects and their peculiarities, such as people, places and things, often studied not in the abstract but in their historical context. It deals with acts and utterances that stand in need of interpretation, or art works we aim to understand. Its paradigms have not been quantitative research on large, well-structured datasets, but unique and particular pieces of information that might reveal human intentions or emotions. Can the tools of Big Data help us address the questions that arise for the humanities, and, more broadly, can they help us with the human problems at the heart of some of the grand challenges we face as a species, such as conflict, environmental damage, healthy living and food security?

At first, there appears to be no reason, in principle, to think they cannot be of help. After all, there is a growing body of highly developed interdisciplinary research involving two or more disciplines that span the divide between the sciences and the humanities. Genuine interactions between researchers in the humanities and the sciences have led to outcomes that neither could have envisioned alone. New methods involving carbon dating and genetic tracing have answered questions that have preoccupied archaeologists and historians about how people lived, what contact they had with other societies, what they ate and made. Another example is the way in which the neurobiology of the emotions, mental illness, and decision making has been broadened by contributions from historians and philosophers. Art historians working with materials scientists have been able to cast further light on the methods used by artists to pursue their singular visions.

Can many of these issues and others like them be usefully explored by combining research in the humanities, not just the digital humanities, with developments in AI, computer science, biotechnology, the life sciences and engineering?

It is sometimes suggested that the methods of the humanities and its way of addressing the nature of the human world are at odds with the Big Data revolution. But if Big Data cannot be repurposed to address, in new ways, the traditional

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

questions in humanities research, it may still create new research agendas for humanities researchers in the digital age. Research in the humanities may change when more of the objects of study have been digitized, or are themselves born digital, or involve the Internet of Things.

Big Data for the Humanities?

Perhaps, surprisingly, the humanities bears some of the hallmarks of Big Data, especially with respect to the three 'V's mentioned above: 'Big data are high volume, high velocity and/or high variety information assets that require new forms of processing'.[5]

More than many other branches of intellectual life, the humanities reveals high variability among its data sources. These may be texts, paintings, sculptures, ruins, historical sites, films, recorded interviews, artefacts of all sorts. In fact, we have always had Big Data in the humanities because we have had libraries, archives, museums and galleries to serve as the storehouses of human knowledge and experience. To handle all these items, large-scale bibliography and cataloguing were needed, while for collections, curating was needed to provide meaning and give context to the items on view. These practices in the humanities have fostered skills that are much needed in the digital sphere to ensure that what is digitized are meaningful and usable data.

Even when we restrict attention to texts alone, there's a need for Dig Data analytics to handle massively large document records. A key example is the Arts and Humanities Research Council funded Big Data for Law Project in the UK. There are an estimated 50 million words in the statute book, with over 100,000 words added or changed every month. No legislator, policy maker, or end-user can be expected to read the statutes or even all of the words of an act that is due to be amended by parliament. Big Data technologies may prove invaluable here to provide adaptable online tools to search for recurring themes or sift for relevance. This is possible for legislation as long as the data, the tools and the techniques are trusted and inspire confidence. The data in question are legal texts – statutes – that have been carefully digitized in such a way as to ensure the legal data are manageable and useable.

So far so good, but how should we understand the challenges the Big Data revolution promises to bring to traditional ways of garnering knowledge? What humanities disciplines such as history, philosophy, art history, literature, as well as the digital humanities, provide is a fundamental engagement with the assumptions behind the gathering of information, with its nature, and how it is used to generate knowledge. In assessing the likely impact, potential advances or downsides of Big Data in the humanities I will point to some of the key issues that need to be raised. I will now address these in turn.

What are the Data in Big Data?

A key problem in discussions of Big Data is that the term 'data' is ambiguous – a fact too frequently overlooked. Data can mean data structures – formatted bytes, bits of

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

code that are not in themselves intrinsically meaningful. Or, data can mean support for a conclusion: in this sense, data are only data when they are evidence for a hypothesis, a source of useful and telling information. It is something akin to the latter notion that people usually have in mind when they say they hope for insights from Big Data, but it is often just data structures in the former sense that are proliferated by the automatically-generated amassing of information. It is all too easy to assume that we have data in the sense of evidence for a hypothesis – even for a hypothesis we haven't yet thought of – when we collect massive amounts of data and have new computational techniques for processing the data. Yet we are very far from that point when we digitize and archive material. That is not enough to get to the next stage. We need to *datify* the collected material and *datafication* goes beyond digitization. Once you have digital objects standing for the real, you can do things with them that you couldn't do before. It's here that we find the cutting-edge projects. New tools that link contents from vast stores of digitized newspapers to show patterns of convergence in nineteenth-century thinking about key events, spanning several continents, made possible by, for example, knowing which newspapers writers in regular correspondence were reading. By being able to see what newspapers Wordsworth, Coleridge and their contemporaries read on a single day, we may begin to know better which issues were topical, thus enabling us to understand from the letters they exchanged on those days how they interpreted the world around them, contrasting this, perhaps, with how others across the country were reading these events. This opportunity creates new questions and offers new methods of inquiry in the humanities. But it is possible only if the data have been organized in such a way as to allow the data to be searched, classified and transformed into knowledge. The data may be too messy to adduce anything much at first. And the old adage of AI and computer science – 'garbage in-garbage out' – still applies. So what we need is a new generation of digital curators who will be a valuable part of the story. The curating skills of sifting, understanding, organizing and interpreting will become vitally important in the new digital age. For unless we take care to curate the information we are storing, there may be little of use we can get out of it. Most of the information we gather by mere digitizing will become landfill.

Big Data as the End of Scientific Theorizing?

Not everyone sees things this way and here there is quite a division between researchers in the humanities and researchers in digital technologies. While many of the former believe that unguided data-driven, or data-sifting, approaches don't yet generate real insights, many of the latter think it is an advantage of Big Data to generate claims we did not know we were looking for. Without a hypothesis, they think we can generate a conclusion that will provide an answer to a hypothesis we haven't yet dreamed of. But how will they suggest themselves and their relevance to our inquiries?

It was Charles Darwin who prefigured the view that data are only data for a theory: 'How odd it is that anyone should not see that observation must be for or

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

against some view if it is to be of any service'.[6] The alternative view, widely reported by Chris Anderson in a 2008 *Wired Magazine*, suggests we are at the end of theorizing and hypothesis testing:

> Petabytes allow us to say: 'Correlation is enough.' ... We can analyze the data without hypotheses about what it might show ... statistical algorithms find patterns where science cannot ... Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There's no reason to cling to our old ways.[7]

He is not alone. Kitchin helpfully assembles a list of quotes heralding this new post-theorizing vision:

> ... scientists no longer have to make educated guesses, construct hypotheses and models, and test them with data-based experiments and examples. Instead, they can mine the complete set of data for patterns that reveal effects, producing scientific conclusions without further experimentation.[8]

> ... mining Big Data reveals relationships and patterns that we didn't even know to look for.[9]

In a paper entitled, 'No questions asked: big data firm maps solutions without human input', Clark talks about 'digital serenity', which works without human biases: 'Instead of waiting to be asked a question or be directed to specific existing data links, the system will – undirected – deliver patterns a human controller might not have thought to look for.'[10]

Like these writers, Kitchin assumes we are seeing the emergence of a new scientific paradigm in the sense of Thomas Kuhn, one where we longer need scientific models to explain the phenomena in a given target domain or have to ask the appropriate questions. He accepts that 'science is entering a fourth paradigm based on the growing availability of Big Data', which enables an entirely new epistemological approach for making sense of the world; 'rather than testing a theory by analysing relevant data, new data analytics seek to gain insights "born from the data".'[2]

These are eye-catching claims but there are two cautionary points to make. First, let us suppose that the patterns in the data are getting at something that is really there in the phenomena themselves; patterns, as Dyche says, that get at something we didn't know we were looking for. The key question is whether, if we do have genuine patterns emerging from large, messy datasets without supervision or many prior assumptions, we will care about these patterns. Aren't the patterns that interest us necessarily linked to human perceptions, needs and interests? Consider the following case. Suppose we did have a question that traditionally we struggled to answer: for example, art historians want to know whether drawings from the cinquecento were by Raphael or from the school of Raphael. Many scholars argue that the fineness and skill of the cross-hatching indicate that some drawings are by Raphael himself. But this makes it true by definition that the best drawings are by the master. Instead, imagine that we scan all candidate drawings and use a deep learning

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

algorithm to find patterns in them. The algorithm may well distinguish Raphael from the school of Raphael, but left unchecked to continue searching for patterns it may answer questions it wasn't even asked, since it may partition the sample drawings into Raphael in the morning, Raphael in the evening, or Raphael when tired or after a bout of heavy drinking. The point is that we don't care about these distinctions and yet they may be generated automatically. It is always humans that have to call a halt or decide how discriminating they wish to be.

Second, why suppose that patterns thrown up or generated automatically will always correspond to something significant or constitute a genuine discovery? When the range of data to which an algorithm applies is extensive it is not inevitable that it will make significant discoveries. Spurious correlations abound.[11] As they say, you don't find a needle in a haystack by increasing the size of the haystack. Pace Anderson, correlation is not enough. Correlation isn't causation. When a building is ablaze there will be a correlation between the size of the fire and the number of firemen attending. As the fire increases so does the number of firemen, but the increase in firemen doesn't cause the increase in the size of the blaze. Without understanding the situation or the workings of the system concerned we cannot move from correlations to causal dependencies.

Despite this obvious distinction, enthusiasts of Big Data are often tempted, as the quotes above show, to suggest that when we have correlation between two phenomena on a vast enough scale there is greater reason to assume that we must have causation, and therefore a potential causal explanation of the phenomena in question. The idea is tempting. Why would things turn out that way, regularly and repeatedly, over so large a domain unless there was an underlying causal mechanism to ensure the regularity continues? But bigger doesn't always mean better. In fact, the larger the dataset the greater the risk of spurious correlations. For example, the divorce rate in Maine over a decade is highly correlated (99.2%) with the per capita consumption of margarine. The total revenue generated by arcades over a decade is highly correlated (98.5%) with the number of computer science doctorates awarded in the US. And the per capita consumption of chicken over a decade is highly correlated (89.9) with total US oil imports. So when more data leads to more evidence of correlation, we cannot assume we have stumbled on a significant result. Instead, we need hunches about which correlations are worth pursuing, how they fit into bigger and bigger scientific frameworks and cohere with other findings, and this relies on humanly instigated, intelligent searches. Even without a prior hypothesis there are always priors. Just supposing more data will throw up the answers isn't always helpful.[12]

Treating patterns thrown up by Big Data analytics as merely a new tool for generating new possibilities, only some of which may be worth attending to, is a fine ambition. In this way it can serve as a form of cognitive enhancement for humans. Music composition, for example, is a case where algorithmically generated patterns can extend the range of possibilities a composer can draw on. But in the end, his or her selection of the patterns, or minor revisions to them, still plays a crucial role in the acceptability of the machine-augmented composition.[13]

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

Big Data or Better Data: Prediction or Explanation?

It may be objected that correlation is enough if we can create machine algorithms that match all the relevant existing data points in a system, and not merely that they generate predictions that can be confirmed by further data sampling. Why would it matter if we are not tracing the causal forces at work in the target system, like a weather front, or bacterial spread, as long as we can predict the behaviour of the system? This argument has some force and needs further examination. In effect, it is a plea for the value of prediction over explanation. Do we want an explanatory account of the internal workings of the system or do we just want a program that will predict its behaviour and allow us to intervene or simulate it? The idea that it is possible to have predictions without explanation is what connectionist networks provided in the late twentieth century.[14] These were AI programs mimicking neural networks that could, after supervised learning, predict rules for, say, regular verbs past tense endings and the irregular exceptions in English, without making use of a prior rule, or offering any underlying explanation for how the patterns in the data were arrived at.[15] More impressively still, the gradual acquisition of past-tense formation in English exhibited the same U-shape learning trajectory that children show in their acquisition by at first over-regularizing the rule for pluralizing verb endings, saying 'go-ed' instead of 'went', then coming to correct it in subsequent usage.[16] However, we don't learn much about how a child – or for that matter the system – performs the task: they simply do so after enough trials and learning. Merely mimicking the performance of children's language acquisition does not prove that the system gives us any explanation of how children do it. Machine learning could provide yet another route to the same goal and little insight will have been gained. That doesn't matter if all we want is a system that will perform behaviourally in a reasonably similar way to human language users. It all depends on what the goals are for this kind of technological advance.

Often, there is a trade-off between explanation and prediction: mere heuristics can do a pretty good job at capturing rough and ready generalizations about human behaviour without giving explanations in terms of the underlying states and causal processes responsible for that behaviour. Philosopher Daniel Dennett has spoken about the strategy of taking up the intentional stance when trying to figure out what a chess playing program is going to do. This means attributing to it beliefs and desires in order to figure out what it's trying to do and how to beat it. However, there are no beliefs and desires in the machine. At the design level, there is an algorithm composed of pattern matching and action rules that is really responsible for what the machine will do, and at the implementation and physical levels there is a binary code corresponding to electronic on and off switches. Dennett thinks that since we don't know all these details it is more efficient for us to take up the intentional stance we use to predict and respond to the behaviour of agents when dealing with AI systems. In thinking that the chess-playing program wants to get its Queen out early, the human opponent can adapt to its behaviour and predict reasonably well how the computer will behave. For Dennett, the intentional stance means adopting rational

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

decision theory to predict behaviour of the system by attributing beliefs and desires that give reasons for the system acting as it does. However, according to Dennett, neither humans nor computers have beliefs or desires: they only approximate to the principles of rational decision theory. It is not actually inner states like beliefs and desires that causally propel behaviour of either human agents of machine algorithms. The attribution of such states is just a reasonable, instrumental heuristic that observers use to organize and predict what a system, including systems like us, do.

However, philosopher Donald Davidson holds a very different view of the role interpretation plays in the rational explanations of people's behaviour. Ordinarily, we appeal to beliefs, desires and intentions to make sense of what people do and say and Davidson thinks that it is only when a creature's behaviour succumbs to rational explanation in terms of beliefs, desires and intentions that it counts as an agent and its behaviour can count as an action. The beliefs and desires people have are the ones we can attribute to them in the course of making the best rational sense of their lives and conduct. For Davidson, the principles of interpretation determine what does and doesn't count as a rational agent. Rational explanations of people are post hoc and partial, they can be revised in the light of new bits of behavioural evidence in order to make best overall sense of the person's total life and conduct. In offering post-hoc interpretations of their motives to explain their behaviour there is no attempt to predict their behaviour before the fact. According to Davidson, there are no laws of rational psychology that will enable us to predict with unswerving accuracy what people will do next. People are like the weather, we can make sense of their behaviour after the facts are in but cannot be sure to predict every event in advance. The hope of Big Data is to change this situation. But even if we get prediction, we don't necessarily get explanatory purchase. So when we are right about what the traffic flow on a given evening is in a large city, or correctly predict what people will buy, we are not guaranteed to know why we are right or why people do this. In what way will this help us to understand the human world and how it may, suddenly and unpredictably, change and depart from the erstwhile patterns? Are we content merely to predict without understanding for as long as it works?

Advocates for the new digital technologies will point out that it's not just the size of the datasets that matters: '. . . in the era of big data, more isn't just more. More is different'.[7] Let's pursue this line of thought a little. What makes the data different and how does that address human needs and interests?

Researchers Danah Boyd and Kate Crawford suggest in one of their six provocations for Big Data that bigger data are not necessarily better data; nevertheless, they point out that the claim for better data comes from the relationality or networked nature of that data:

> There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem. Big Data is notable not because of its size, but because of its relationality to other data . . . Big Data is fundamentally networked. Its value comes from the patterns that can be derived by making connections between pieces of data, about an individual, about individuals in relation to others, about groups of people.[17]

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

So what matters here is social computing, that is, the computer-facilitated study of human social behaviour, human experience, and human culture, mediated by information that takes in its context. Never before has such a rich profusion of data from social media been available for those capable of mining and analysing it. It's true this type and the range of information is different. But how authentic is the glimpse of people's thoughts, experiences and attitudes generated from the combining of data on social media supposed to be? Surely we need to remember, as Lev Manovich says, that: 'Peoples' posts, tweets, uploaded photographs, comments, and other types of online participation are not transparent windows into their selves; instead, they are often carefully curated and systematically managed.'[18]

Although, perhaps we are expecting too much. The data provided are not unbiased. But perhaps predicting people's behaviour in such a way as to recommend websites to visit, ways to assess their health and recommend treatments, sell them goods, or find them books, films or music of interest is good enough: 'Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.'[7]

There just has to be enough pattern in people's behaviour to predict further behaviour in a way those people themselves could not discern. Consider how creative machine algorithms for handling Big Data can be when it comes to choosing the things we care about.

Curated by You or Curated for You?

Once it was noticed that purchase patterns could point to associations – if you like X, you'll like Y – without explaining the basis of these judgements or why those patterns should hold, there came an opportunity to exploit them at scale. There is no causal claim here, just good prediction. And if predictions work for us we will endorse the recommendations we're given. We add 'Likes' to Facebook, or we purchase the recommended items. Currently, existing algorithms are worryingly good at predicting what else we will like on the basis of what we have endorsed so far, with the odd hopeless departure from successful cases. As artist, musician and social commentator, David Byrne points out:

> Social media recommendations, of course, are a form of curation. The internet memes and viral eruptions, the algorithms and whims of the herd, sweep us along and these, too, increase the value of the things they recommend. Are they unbiased? Not so much. They're often trying to sell you something, or gather information on you that they can then sell to someone else, so trust in that world is a rare commodity.[19]

Of course, as Byrne reminds us: 'One has to choose carefully who to trust for recommendations. We are curating the curators.' How we do this is a good question for we need to predict how these systems will continue to work and whether, however well they have performed so far, they will diverge from us in unexpected ways.

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

Nonetheless, these new tools provide: 'something valuable, beyond making a seemingly random selection based on simple criteria such as top-selling, most emailed or newest (though those, technically, are forms of curation, too)' (Refs. 2,19).

Although, however good these digital technologies are, will they feed our need for curiosity? What about 'our urge for the creative, the quirky, the ironic, the exclusive'. They will not show us how or where to find it because: 'In contrast to all of the above, what I and other "experts" offer is surprise.'[19]

The point about surprise is well-taken. People will often behave in reliably predictable ways: what will amuse them, bore them, frighten them, which brands they will like, who they do and don't admire. And just so long as they continue to exhibit these patterns in behaviour it will not be so hard to judge what they will choose or like. By their actions you shall know them, and what people do is a *reasonable* indicator of who they are and what moves them. But every now and then they will surprise us by departing from what they have done up to now, and although they are doing something unexpected we can still make sense of it. A person's character is a function of their behaviour. Every now and then they can act out of character, departing from the pattern they have exhibited to so far, and that surprising choice could be triggered by truly creative thinking that subverts the wisdom of the crowd: 'The value of encountering an idea, an artist or a writer outside the well-trodden and machine-predictable paths' resides in their judgements being 'a little more biased, skewed, pre-edited and peculiar than what those herd-based and algorithmic services come up with. I'd love to think that what I'm doing serves this need in a way that big data and learning algorithms cannot.'[19]

The same is true when choosing what wines you'd like to drink. Predicting what people will like to try next on the basis of what others who have purchased the same bottle went on to buy and recommend won't always lead to exciting new finds or extend one's range. Each person's wine journey needed this leap to something new that a human not a machine recommends. Byrne is surely right: Big Data 'favours memes over masterpieces' and 'looks for trends not causes'. If anything, Big-Data driven predictions can be a little too safe, a little too predictable. As the joke goes, a machine algorithm walks into a bar and the bartender asks, 'What will you have to drink?' The machine algorithm replies, 'What's everyone else having?' The wisdom of the crowd is not always going to lead to better, more exciting choices.

So, although

> Big Data analytics might provide some insights, it needs to be recognized that they are limited in scope, produce particular kinds of knowledge, and still need contextualization with respect to other information, whether that be existing theory, policy documents, small data studies, or historical records, that can help to make sense of the patterns evident.[2]

What Big Data and learning algorithms leave out at present is the special role human experience plays in a creative mind in filtering out or finding meaning and significance as a way to guide successful choice and behaviour. This is not to deny that there can be room to combine this ability with the use of digital tools

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

in order to give us access to datasets on an unprecedented scale. Such datasets can provide possible search spaces that creative minds can exploit. They could be a tool for creativity, as in the case of musical composition, but so far they can't replace it, and the reasons for thinking they can't are interesting and surprising, as I will now go on to show.

To sum up where we have got to so far, I offered two objections to Big Data digital technologies that are believed by some to replace scientific theorizing. The first is that the patterns in data automatically generated need not be patterns that have relevance for us. The second is that bigger datasets are more susceptible to spurious correlations. Neither is decisive. Beyond that, we considered whether prediction rather than explanation had some virtues when working on such a large scale. But even here, the usefulness of algorithms predicting what we may like to see, or read or listen to, are rather safe and will not surprise us and move us on in the development of our tastes. But now I want to turn to what I consider the most fundamental problem for the big data revolution: the failure to take into account the multisensory basis of human experience, emotion and decision-making and why generalizations and predictions about us that leave this key dimension of our experience out will always fall short.

What we Don't Find in Big Datasets

The danger of over-claiming on behalf of the new digital technologies lies in thinking that the searchable space of possibilities Big Data creates can contain all the information we need, or usually rely on, when making creative decisions or expressing preferences. This is not the case. The dream of Big Data is to be exhaustive: to capture all the information for a given domain from which significant and meaningful information can be mined. But where do we gather the data from and how are the patterns analysed on the basis of those data presented to us? According to Kitchen: 'though Big Data may seek to be exhaustive, capturing a whole domain and providing full resolution, it is both a representation and a sample, shaped by the technology and platform used, the data ontology employed and the regulatory environment, and it is subject to sampling bias.'[2]

There will be sampling bias, but even if that could be corrected there is a more fundamental problem with the claim to exhaustively capture features of the human world: of experience, memories, perceptions, emotions and actions; namely, that which the digital harvesting of particular ranges of data and the computational techniques used to analyse them leave out. This is the mass of information from the sensory world that constitutes a large part of the way humans understand the world around them and themselves.

We must remember that the data digitally captured are only stand-ins for what they give us data about. It is a selection of data chosen and formatted to suit existing technologies that can be accessed or used with the tools we've created. What data to collect and the tools used to transform is decided within the context of existing

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

scientific theory. It is important to remember that 'data are not natural and essential elements that are abstracted from the world in neutral and objective ways and can be accepted at face value; data are created within a complex assemblage that actively shapes its constitution'.[20] What we should be cautious of, and what I want to address head on concerns the input and output side of digital technology. What is the full range and extent of the data we are collecting, and how can the analysed data we collect be presented to us?

On the input side, we 'datafy' everything we can, but does this include everything that we might need and attend to? And on the output side, how should we expect to interact with the data generated? This is a new set of problems for the Human–Data Interface, part of a much-needed Human–Data Interaction research akin to that carried out in Human–Computer Interaction departments.

The problems concern not just the points of interface between humans and data, but the connections between what human agents easily consume and handle in their fast, fluid dealings with the world: something that is being extensively studied by the cognitive neuroscience of perception, cognition, emotion and action. We are sensing, thinking, feeling and acting beings and many interactions within the human body and between the human body and the external environment shape and constantly modulate the experience by which we take in and understand the world around us and ourselves. Cognitive neuroscience has increasingly shown the way that subtle and surprising interactions between these systems sustain our conscious experience of the world. Recent findings have shown why the way we ordinarily think about how we perceive the world around us is not how it works. We assume most of the information we take in is simply through our eyes and our ears, and yet the senses don't work in isolation but are always engaged in continuous cross talk with inputs from one sense affecting the workings of another. So it is a cognitive oversight to believe that collecting data in a digitally convenient audio-visual way will enable us to exhaustively capture all aspects of our pasts and store them for the future. It is highly unlikely that gathering information this way will capture the information we rely on to make sense of ourselves, others and our environments. That's the input side. But equally, on the output side, if we are to engage frequently with the findings of continuously collected, dynamic waves of data, it will not always be appropriate to interact with the data in a purely visual way. Let's look at both the output and input side of this problem.

Human–Data Interaction

How is the increasing amount of complex data that can be served up to us when large unstructured datasets are sieved for new findings, made available to us? To make sense of how we can go from large unstructured datasets to meaningful insights, a lot of faith is placed in data visualization. But visualization techniques alone are not the answer. Just as vision science tends to dominate perceptual neuroscience and psychology, so it tends to dominate ideas in computer science about how best to access complex information. And yet vision is not always the best medium to pick

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

up all types of information. Besides, visualization is not neutral: it involves a way of pre-digesting and packaging the information for human users. Seeing patterns in data may work in some domains, while in others it may be better to hear data, or feel data. To understand why is to understand more about how the brain and the nervous system pick up their own data and process them.

Some information, in particular spatial information, is best absorbed visually. However, information with a temporary dimension, and where we must pay attention to different levels of input simultaneously, is best absorbed through audition. Sonification of data may well work best when people have to process the temporal dynamics of a system, or temporal change. Equally, musification may be best when we need to present multiple pieces of information simultaneously. An example of the latter is shown in a collaboration between Stanford neurologist Josef Parvizi and music composer and computer scientist Chris Chafe, who musified the EEG signals from the brains of patients with epilepsy. The idea to convey the sounds of a brain in normal state, in seizure and post-seizure came from neurologist Parvizi having heard the Kronos Quartet playing a piece of music called *The Music of the Spheres*, based on signals sent back to earth from a remote probe. Parvizi wondered whether it was possible to capture the music of the brain, and contacted his colleague Chafe to work on the problem. Chafe's solution was to provide a voice for the EEG signal coming from one hemisphere, and a different voice for the signal from the other hemisphere. As the signal oscillated so did the voice. When both are played together, representing the different, but co-occurring, activity of each hemisphere of the brain, it is possible for naïve participants to judge, with around 80% accuracy, whether they are listening to a brain in a normal or an abnormal state. Thus, without training in EEG, the majority of people can distinguish instantly between brains in a healthy state and brains entering seizure. The presentation of data in what Parvizi and Chafe are calling a brain stethoscope turns out to be the best way for us to take in and understand the information we can record from the brain.[21]

In other work, Chris Chafe sonified the ripening of tomatoes in a greenhouse. When they ripen, tomatoes give off CO_2 at different rates depending on how fast they ripen, and by attaching sensors that measure the amount of CO_2 released from each plant, then sonifying the signal, you can hear the tomatoes ripening as you walk through the greenhouse. By this technology, Chafe is making the invisible perceptible: a form of sensory extension.

The general point here is that we should not be so visuo-centric. So far, we have been reliant on an almost exclusively visual perspective in studies of perception. And yet we continually use many more of our senses to perceive and make sense of the world. As we walk around, we see and hear, but also feel our bodies and smell our surroundings. The inputs from the other senses are produced in tandem with visual and auditory input, which is why our experiences are always multisensory. We see and hear the environment around us while simultaneously smelling it, feeling the ground under our feet, and perhaps having a lingering taste of coffee in the mouth. And these experiences don't happen in separate parcels: they are part of the unified conscious experiences we have, moment to moment, and we often don't notice how

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

that experience is modulated by ambient sounds or smells, and yet we know from work on dining in restaurants how much those unnoticed signals can affect experience and behaviour when it comes to eating.[22]

There are senses that you don't know are operating and yet they shape your current experience. You can sense the positions of your limbs without having to look at or feel them. This is proprioception. This is part of your sense of balance that comes from the workings of the ear canals: the vestibular system that tells you whether you are going up or down in an elevator, or whether you are moving forward of backwards. The sense of balance is always on but you are unaware of it until it is switched off, as in zero gravity, or due to dysfunction of the vestibular system. Another constant, though hidden, contribution to your everyday experience is your sense of smell. You may be unaware of the smell of the room if it is familiar to you, but you would notice its absence if it was not sensed, as those who lose their sense of smell report, when speaking of feeling alienated from their immediate surroundings. People think they smell only when they sniff at something. But that's not true. They smell because they breathe, as psychologist J.J. Gibson put it. Smell is always on, subtly modulating our conscious experiences of familiar people, places and things.

We are constantly bombarded with sensory inputs that interact or combine to give rise to a rich range of multisensory experiences occurring in a unified conscious space. Many of the sensory experiences we think of as unitary are in fact integrations of several sensory inputs. Take tasting, which we think we are just getting from the tongue. In fact, what we call tasting is always the result of touch, taste and smell. All the tongue can give us is salt, sweet, sour, bitter, savoury (umami), and metallic. And yet we can 'taste' peach, pear, strawberry, mango. We don't have mango receptors on the tongue. These flavours are perceived by the nose and the odours that travel from the mouth to the olfactory epithelium in the nasal cleft. It is when odours are pulsed to the nose from the mouth when swallowing that we get a big release of flavour.[23] Texture matters too. When we remove the texture of – say – a cucumber by pureeing it and forming a jelly, only 8% of the population can tell what they are eating. The identification and acceptability of a food depends on its mouthfeel.[24]

In future, and in particular when the Internet of Things becomes a reality, we won't need to consider seeing data, hearing data, and feeling data with data haptics. Perhaps we should be smelling data for recipe cues. As we've seen in the work of Chris Chafe, we might make the invisible perceptible by using any of the above means to sense what we usually only read about or measure. There have even been tests with data edibilization: 'encoding data with edible materials, as a novel approach to leverage multiple sensory channels to convey data stories'. The advantages of data edibilization are described in terms of 'attractiveness, richness, memorability, affectiveness, and sociability'.[25]

The reason for considering all these ways of accessing data is clear when we turn to the input side and focus on the kind of information our sensory systems and brains automatically gather and integrate, and how rich this information is as

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

compared with the way the digital domain artificially segregates it. Let's look at this issue more closely.

Multisensory Data Interactions – The Input Side

Making sure we gather meaningful data that will allow us to extract significant findings at a later time means using data that are accurate, relevant and tailored to the nature of the human mind and human experience. The data that we have considered gathering so far, even with remote sensing, have failed to recognize the multisensory nature of everyday human experience.

Our fundamental contact with the world comes through our senses, shaping knowledge of the world and our own bodies, and to understand what goes on here we have to understand the senses, how they work, how many there are, and how they enable us to perceive the world and ourselves.

For far too long we have laboured under a faulty conception, believing that we have just five senses: sight, smell, hearing, taste and touch. But why think we only have five? What of the sense of balance, the workings of the vestibular system and proprioception just mentioned? What about feelings of pain, hot and cold, itching, knowing whether your fingers are clenched or stretched, knowing how the fabric of your clothes feels? Why think of this as a single sense of touch? These multiple senses are working all the time to tell you about the world around you and about yourself. There is interoception, our inner senses that detect subtle changes in our bodily states: our viscera, our breathing, the changing rhythms of our heartbeats. All of these changes can modulate your emotional state and the accuracy of your perception and cognition.

The role interoception plays in emotion and cognition is only just beginning to be understood. Every time the heart beats it sends a signal to the brain and those who can use that channel, unreflectively, can regulate their emotions and make better decisions. Some people can guess with incredible accuracy how many heartbeats they have had in a minute. This is known as the Shandray test. A pulse oximeter is put on the finger to record beats and the subject has to say how many heartbeats they have had after a minute. Many say they are just guessing or not able to do it, and yet some of them, unknown to themselves, are entirely accurate. This accuracy without awareness can guide perception and cognition. For instance, work by Sarah Garfinkel and Hugo Critchley shows that when signals are presented on the heartbeat – at systolic discharge when the heart pumps out blood – they are usually inhibited unless they are fear signals, which are amplified for the person. While at diastole, when the heart is refilling, people make better perceptual discrimination judgements.[26] Snipers intuitively know to take their shots between heartbeats. Can this information also influence decision-making? The answer is yes. In a remarkable study by the same group, city traders in London were ranked by their profit and loss scores. They were then measured as to their interoceptive accuracy on the heartbeat test. Those with highest profit and loss scores, i.e. those who made the best split-second decisions when

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

looking at a bank of computer screens, were those who had the highest interoceptive accuracy scores.[27] The traders speak about relying on gut instinct, but they are actually accessing the heart-brain communication channel. This aspect of their split-second decision-making could not be simulated by machine learning algorithms.

Far from relying on five senses, contemporary neuroscientists have suggested that we may have as many as 22 or 23 senses. There is good reason to think that we may have two senses of smell: externally focused smell – orthonasal olfaction – happens when we inhale, enabling us to detect things in the environment such as food, predators or danger; and an internally generated smell – retronasal olfaction – when odours travel from the mouth to the nose and are exhaled. This sense of smell enables us to assess the quality of what we have eaten, enabling us to know whether it is good or bad, to be consumed or expelled. Associated with each sense of smell is a distinct hedonic response. Orthonasal olfaction gives rise to the pleasure of anticipation; while retronasal olfaction gives rise to the pleasure of reward. Anticipation is not always matched by reward. The enticing aromas of freshly brewed coffee are never quite matched by the taste, or rather smells rising from the mouth to the nose. There is always a little disappointment in how good the coffee smells outside and how it is experienced on the palate.[28]

In addition to the proliferation of the senses, another revision neuroscientists are making to ordinary thinking about the senses is how much the senses interact and combine at both early and late stages of processing to produce rich perceptual experiences of our surroundings. We simply take in the rich and complex scene without giving much thought to how the different contributors produce the whole experience. We give little thought to how smell provides a background to every conscious waking moment. People who lose their sense of smell can be plunged into depression and show less sign of recovery a year later than people who lose their sight. This is because familiar places no longer smell the same, and people no longer have their reassuring olfactory signature. Also, patients who lose their smell believe they have lost their sense of taste. When tested, they acknowledge that they can taste sweet, sour, salt, bitter, savoury, and metallic. But everything else, missing from the taste of what they are eating, is due to retronasal smell. As mentioned above, touch contributes to sauces tasting creamy, and other foods tasting chewy, crisp or stale. At first, the only difference between potato chips, which 'taste' fresh or stale, is a difference in texture. The largest part of what we call 'taste' is in fact smell in the form of retronasal olfaction, which is why, as mentioned above, people who lose their ability to smell say they can no longer taste anything. Taste, touch and smell are not merely combined to produce experiences of foods or liquids, rather the information from the separate sensory channels is fused into a unified experience of what taste and food scientists call flavour. What we call the taste of menthol lozenges involves a minty aroma, a slightly bitter taste and a cool sensation in the mouth. Without adding in all of these components we don't have the 'taste', or more accurately the flavour, of menthol.

Flavour perception is the result of multi-sensory integration of gustatory, olfactory and oral somatosensory information into a single experience whose components we are unable to distinguish. It is one of the most multi-sensory experiences we have

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

and can be influenced by both sight and sound. The colours of wines and the sounds food make when we bite or chew them can have large impacts on our resulting appreciation and assessment, and irritation of the trigeminal nerve in the face will make mustard seem 'hot' and menthol 'cool' in the mouth without any actual change in temperature. So far, these chemical senses cannot be captured digitally and yet they make a large part of our everyday experiences are humans.

In sensory perception, multi-sensory integration is the rule not the exception. In audition, we don't just hear with our ears, we use our eyes to locate the apparent sources of sounds in the cinema where we 'hear' the voices coming from the actors' mouths on the screen, although the sounds are actually coming from the sides of the theatre. This is known as the ventriloquism effect. Similarly, retronasal odours detected by olfactory receptors in the nose are experienced as tastes in the mouth. The sensations get relocated to the mouth because oral sensations of chewing or swallowing capture our attention, making us think these olfactory experiences are occurring in the same place. The McGurk effect shows how the eyes and the ears collaborate to 'hear' what someone is saying. If you look at a mouth making the sound 'ga' while hearing an auditory input of the sound 'ba' what you experience and take as hearing is 'da', a sound you neither saw or heard but which is intermediate between 'ba' and 'ga'. So what we think of as our visual experiences are not the product of our eyes alone. For another example, when you get on a plane and sit down, you can see everything in front of you in the cabin, But when you look again when you're in the climb you will 'see' the front of the cabin looking higher than you are. Of course it is. But how can it look that way? You are in exactly the same optical relation to everything in the cabin that you were before. And yet it looks different. This is because the fluid in your ear canals tells your brain that you are tipping backwards and together with the visual information from the eyes, the brain creates a vestibular system adjusted visual experience, something a camera just can't capture.

Other collaborations among the senses are due to cross-modal effects, where stimulation of one sense boosts activity in another. We say vanilla smells sweet, although sweet is a taste not a smell, and pure vanilla has no sweetness in it at all. Industry knows about these effects and exploits them. Aromas in shampoo can make your hair feel softer; and red-coloured drinks are experienced as sweeter than lime-green drinks even if the sugar content is the same, while drinks with a light green colour 'taste' sour. In this and many more ways, our senses cooperate to create the experience of the world around us, including those we think are just cases of seeing or hearing. In many of these interactions vision will dominate; but not in every case. For anyone unlucky enough to have disturbance in their vestibular system they will feel the world is spinning, although cues from the eyes and the body should be telling them everything is still.

What we experience is always the result of how the brain works out what is going on from the combined input of all our senses. It is in this way that our senses cooperate and help us make sense and get around in the world. And since the world we inhabit is not a sensory, but a multisensory world, it will be important in our future ambitious attempts to collect and process large, unstructured datasets – which represent as exhaustively as

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

possible our environment, past and present – to take on these multidimensional aspects of our surroundings and our multisensory reception of them.

The multisensory perception and integration that shapes human experience, our understanding of the world and ourselves, presents a huge challenge to the vaulting ambition of the new digital technologies. For the Big Data revolution to achieve its aim of exhaustibility it will need to capture such information about human experience if it is to provide us with meaningful data for the future.

Acknowledgement

This work was supported by a grant from the Arts and Humanities Research Council, AH/N50452X/1.

References

1. H.J. Miller (2010) The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, **50**(1), pp. 181–201.
2. R. Kitchen (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, **1**(April–June), pp. 1–12.
3. Open Data Center Alliance (2012) *Big Data Consumer Guide*. Available at: http://www.opendatacenteralliance.org/docs/Big_Data_Consumer_Guide_Rev1.0.pdf
4. J. Constantine (2012) How big is Facebook's data? Quoted in R. Kitchen (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, **1**(April–June), pp. 1–12.
5. Gartner, IT Glossary, http://www.gartner.com/it-glossary/bigdata/
6. F. Darwin and A.C. Seward (Eds) (1903) *More letters of Charles Darwin. A Record of his Work in a Series of Hitherto Unpublished Letters* (London: John Murray), Vol. **1**, p. 195.
7. C. Anderson (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 June. Available at: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
8. M. Prensky (2009) H. sapiens digital: from digital immigrants and digital natives to digital wisdom. *Innovate: Journal of Online Education*, 5(3), Article 1. Available at: https://nsuworks.nova.edu/innovate/vol5/iss3/1
9. J. Dyche (2012) Big data 'Eurekas!' don't just happen. *Harvard Business Review*, **18**, Blog. 20 November.
10. L. Clark (2013) No questions asked: big data firm maps solutions without human input. Wired UK. URL: http://www.wired.co.uk/news/archive/2013-01/16/ayasdi-big-data-launch
11. Thanks to Anil Seth for making this point when presenting an earlier version of this paper.
12. Kitchin in effect, concedes this point: 'It is one thing to identify patterns; it is another to explain them. This requires social theory and deep contextual knowledge. As such, the pattern is not the endpoint but rather a starting point for additional analysis, which almost certainly is going to require other data sets, see R. Kitchen (2014) Big Data, new epistemologies and paradigm shifts. Big Data & Society, April–June, p. 8, emphasis added.
13. See P. Dahlstedt, this issue.

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

14. D. Rumelhart, G. Hinton and R. Williams (1986) Parallel distributed processing: explorations in the micro-structure of cognition. In: D. Rumelhart and J. McClelland (Eds), *Psychological and Biological Models* (Cambridge, MA: MIT Press), vol. **2**, pp. 318–362.
15. D. Rumelhart and J. McClelland (1986) On learning the past tenses of English verbs: implicit rules or parallel distributed processing? In: J. McClelland, D. Rumelhart and the PDP Research Group (Eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Cambridge, MA: MIT Press).
16. K. Plunkett and V. Marchman (1993) From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, **48**(1), pp. 21–69. K. Plunkett and V. Marchman (1996) Learning from a connectionist model of the acquisition of the English past tense. Cognition, 61(3), 299–308.
17. D. Boyd and K. Crawford (2012) Six provocations for Big Date in a decade. In: Paper presented at Oxford Internet Institute's Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011.
18. N. Ellison, R. Heino and J. Gibbs (2006) Managing impressions online. Quoted in L. Manovich (2011) *Trending: the promises and challenges of Big Data. Debates in the Digital Humanities* (edited by M.K. Gold) (Minneapolis: The University of Minnesota Press) [online].
19. D. Byrne (2015) A great curator beats any big company's algorithm. *New Statesman*, 1 June.
20. D. Ribes and S. Jackson (2013) Data bite man: The work of sustaining a long-term study. Quoted in R. Kitchen (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, April–June, pp. 1–12.
21. B. Ravazi, P. Eshani, M. Kamrava, L. Pekelis, V. Nangia, C. Chafe and J. Parvizi. (2015) The brain stethoscope: a device that turns brain activity into sound. *Epilepsy and Behaviour*, **46**, pp. 53–54.
22. C. Spence (2017) *Gastrophysics: the New Science of Eating* (London: Penguin Books).
23. See C. Spence, M. Auvray and B. Smith (2014) How not to confuse taste and flavour in perception and its modalities. In: D. Stokes, M. Matthen, and S. Briggs (Eds), *Perception and its Modalities* (Oxford: Oxford University Press).
24. O. Mouritsen and K. Styrbæk (2017) *Mouthfeel: How Texture Makes Taste (Arts and Traditions of the Table: Perspectives on Culinary History)* (New York: Columbia University Press).
25. Y. Wang, X. Ma, Q. Luo and H. Qu (2016) Data edibilization: representing data with food. Proceedings of the 2016 CHI Conference 2016, San Jose, CA – dl.acm.org
26. S.N. Garfinkel, A. Seth, A. Barett, K. Susuki and H. Critchley (2015) Knowing your own heart. *Biological Psychology*, **104**, January, pp. 65–74.
27. N. Kandesamy, S. Garfinkel, L. Page, B. Hardy, H. Critchley, M. Gurnell and J. Coates (2016) Interoceptive accuracy predicts survival on a London trading floor. *Scientific Reports*, **6**, article number: 32986.
28. For the duality of our sense of smell see P. Rozen (1982) Taste–smell confusions and the duality of the olfactory sense. *Perception & Psychophysics*, **31**, pp. 397–401. For recent work on smell disorders, see T. Hummel, B. Landis and K. Huttenbrink (2011) Smell and taste disorders, current topics in otorhinolaryngology. *Head and Neck Surgery*, **10**, pp. 1–15.

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press

About the Author

Barry C. Smith is a professor of philosophy and directs the Institute of Philosophy at the University of London's School of Advanced Study. He is the founding director of the Centre for the Study of the Senses, which pioneers collaborative research between philosophers, psychologists and neuroscientists. From 2012 to 2019 Barry was the Arts and Humanities Research Council's Leadership Fellow for the Science in Culture Theme. His current research is on the multisensory nature of perceptual experience, focusing on taste, smell and flavour. He has written theoretical and experimental papers, publishing in *Nature*, *Food Quality and Preference*, *Chemical Senses* and *Flavour*. He collaborates with chefs and artists and consults widely for the food and drinks industry. In 2007, he edited *Questions of Taste: The Philosophy of Wine* (Oxford University Press); in 2008, he edited *The Oxford Handbook of Philosophy of Language* (with Ernest Lepore) and, in 1998, *Knowing Our Own Minds* (Oxford University Press, with Crispin Wright and Cynthia Macdonald). Barry is a frequent broadcaster on BBC Radio with appearances on BBC 1 and 2. In 2010 he wrote and presented a four-part series for the BBC World Service called *The Mysteries of the Brain* and in 2017 wrote and presented a 10-part series for BBC Radio 4 called *The Uncommon Senses*.

https://doi.org/10.1017/S1062798719000048 Published online by Cambridge University Press