BIG DATA IN CHINA: BIGGER PROBLEMS, EASIER SOLUTIONS OR BOTH?

Çin'de Büyük Veri: Daha Büyük Sorunlar mı, Daha Kolay Çözümler mi, Her İkisi Birden mi?

Ulaş Başar GEZGİN*
Fatma YILDIRIM**

Abstract

Just like the case in any other power centers of the world, such as U.S., Europe and Russia, big data in China have information-wise polarized the researchers. On the one hand, some have been excited about the opportunities big data will bring to scientific research (e.g. new discoveries already made in medical research); on the other, some others are worried about increased level of privacy violations. Possibly because of the difficulties in self-expression within the political context of China, other more detailed objections to the use of big data by governments and corporations are rarely discussed with regard to China. While Chinese government implements the social credit system which is based on the big data of each citizen collected by any surveillance device –unethical ones included, so far this system has not been discussed extensively within academic circles. However, such a system can be a model for other countries and become a global nightmare rather than a Chinese one only. This social credit system is closely related with the notion of Internet of Things which appears to be benign at first blush, but has the potential for misuse and abuse in the hands of authoritarian governments and greedy corporations. The lines between public and private will be blurred by the advent of Internet of Things which will make any electronic device capable of transferring data to the Big Brother.

In this article, big data discussions, both pros and cons are presented and other themes that are understudied in China are also listed.

Keywords: *Big data, Big data in China, Big brother, Surveillance, Data distortion, and Data manipulation*

* Prof.Dr., Duy Tan University, Danang, Vietnam, ulasbasar@gmail.com
** Doç.Dr., Ankara University, Faculty of Political Sciences, Department of Labor Economics and Industrial Relations, yildirim@politics.ankara.edu.tr

Özet

Dünyanın ABD, Avrupa ve Rusya gibi diğer güç merkezlerinde olduğu gibi, Çin'de büyük veri, araştırmaları bilişim açısından kutuplaştırmış durumda. Bir yandan, kimileri büyük verinin bilimsel araştırmalara getireceği fırsatlara (örneğin tıp araştırmalarında daha şimdiden görülen keşifler) ilişkin olarak heyecana kapılıyor; diğer yandan, başkaları, yükselişte olan özel yaşam ihlali düzeyi nedeniyle rahatsız oluyor. Büyük verinin hükümetler ve şirketler tarafından kullanılmasına yönelik daha ayrıntılı başka karşı çıkışlar, Çin'in siyasal bağlamında kendini ifade etme noktasındaki zorluklardan olacak, Çin'le ilişkili olarak nadiren tartışılıyor. Çin hükümeti, yurttaşların etik olmayanları da içermek üzere her tür gözetim aracıyla toplanan büyük veriye dayalı olan sosyal kredi sistemini uygulamaya dökerken, bu sistem akademik çevrelerde şimdiye dek yeterince kapsamlı bir biçimde tartışılmış değil. Ancak, böyle bir sistem başka ülkeler için de bir model olabilir ve yalnızca Çinli değil küresel ölçekli bir kabusa dönüşebilir. Bu sosyal kredi sistemi, ilk bakışta iyi niyetli görünen, ancak otoriter hükümetler ve gözünü kâr hırsı bürümüş şirketler elinde kötüye kullanım ve suistimale açık olma potansiyeline sahip olan Nesnelerin İnterneti kavramsallaştırmasıyla da yakından ilişkili. Her tür elektronik aracın Büyük Birader'e veri aktarmasını sağlayabilecek Nesnelerin İnterneti'nin gelişiyle kamuya açık olanla özel yaşama ilişkin olan arasındaki sınırlar iyice belirsizleşecek.

Bu makalede büyük veri tartışmaları, olumlu ve olumsuz yönleriyle sunuluyor ve konuyla ilgili olarak Çin'de yeterince çalışılmamış diğer izleklerin bir dökümü sağlanıyor.

Anahtar Sözcükler: *Büyük veri, Çin'de büyük veri, Büyük Birader, Gözetim, Veri çarpıtma ve Veri değişimleme (manipülasyon).*

Introduction

Big data, in the most generic sense, "refers to very large datasets with complex structures, with high volume of data mass, high velocity of data flow and high variety of data types" (Zhang, 2015, p.32). The rise of big data has mostly been met with enthusiasm. China is no exception. Big data enthusiasts in China and elsewhere can be roughly classified into two: Researchers working on natural data (including, for instance, geological research) vs. those on human data (including health areas, e.g. Zhang, 2017; Zhang et al., 2018).

Among the optimists of human big data, we can mention Zhang (2014) who hails the uses of big data for medical research assuming that big data collection methods would provide "reliable and high-quality big data" (p.426) which is rarely the case.[1] Another associated assumption claiming that China will have an

1 On the other hand, the conclusion of the paper looks more cautious than the general tenor of the discussion, but rather than discussing the serious problems presented in Zhang et al (2017), Zhang (2014) argues that the

advantage in big data research as 1.3 billion people would be covered in research studies is not applicable either, as the method of data collection and interpretation is more important than the scale or the population size: With the wrong method, errors can be replicated billions of times.

Rather than a naive optimism, Zhang et al (2017) can be considered as a case of cautious optimism. They agree that medical big data has the potential to contribute to research and scientific advancement in China, but the country is still far from the aspirations despite of government initiatives prioritizing uses of big data such as the China Kidney Disease Network (CK-NET) (see Saran, Steffick, & Bragg-Gresham, 2017). Non-optimal electronic record systems, non-standardized medical terminologies, untraceable medical practices, different degrees of veracity and accurateness of medical big data, and lack of proper legal protections against breaches of privacy in medical big data are the problems listed as the obstacles against the government aspirations in medical big data in China (Zhang et al, 2017).

As also noted by Zhang et al (2017), human big data involves ethical problems associated with massive breach of privacy. Thus, big data enthusiasts have headaches in proposing anti-privacy measures.[2] In China, another solution is found: Rather than reflecting on anti-privacy options, the big brother redefines the notion of privacy in a way to make it disappear. For a Chinese citizen, there can't be anything private that is out of reach of the government. Thus, in the near future, with the rise of Internet of Things in addition to big data, the governments and corporations will be allowed to know private life details of all citizens including how many times a citizen takes a bath in a week and how many times he cleans the house (via the vacuum enabled by internet) which will be used to evaluate whether the citizen is a match for the government job he applied for or not. No transparency, no chances to know about what is reported about you and no way to appeal… These may look like science fiction stories, but in many cases, fictions come true. On the other hand, a cross-cultural study reports that Chinese offer a higher level of acceptance for governmental surveillance compared to U.S. citizens (LaBrie et al, 2018). This can be explained or related to conformity, obedience and social control.

According to Botsman (2017),

main problem about the use of medical big data in China is about the need for "more training courses on data management and statistical analysis" (p.428). Such an argument ignores fundamental problems discussed in Zhang et al (2017).

2 Huang (2018) and Zhao & Dong (2017) provide reasonable accounts of privacy violations and recommend personal anti-privacy measures in China. As their measures are at a personal level, they identify an educational gap in this field. However, in an omnipresent surveillance environment where the lines between state and corporate surveillance on the one hand, and private life and public life on the other are getting more and more blurred, they attribute too much power to each user in case of anti-privacy measures, which is not realistic.

"Instead of trying to enforce stability or conformity with a big stick and a good dose of top-down fear, the [Chinese] government is attempting to make obedience feel like gaming. It is a method of social control dressed up in some points-reward system. It's gamified obedience.

(...)

You could see China's so-called trust plan as Orwell's 1984 meets Pavlov's dogs. Act like a good citizen, be rewarded and be made to think you're having fun" (np.).

There are also fundamental epistemological issues about big data not comprehensively elaborated in China. Error management in big data is a headache. The problem of errors in big data is a more serious problem, because once a mistake is made, it will be replicated endless times, as mentioned previously. There are 5 major sources of errors in big data:

1- Governments distorting the data for PR activities, e.g. faking GDP data, inflation rate, unemployment rate, popularity ratings, even election results in some countries.

2- Corporations not only distorting the already existing data, but manipulating the data even before its production (remember the Facebook-Cambridge Analytica scandal).

3- Technically advanced citizens/users taking anti-surveillance measures (e.g. uses of other IPs or the use of VPN in China).

4- Recording errors (these may be due to the technical problems and/or social and cognitive biases of the recorders).

5-Measurement errors (again these may be due to the technical problems and/or social and cognitive biases of the recorders).

For instance, those who claim that big data leads to discoveries in medical research need to ensure that the data at hand is not contaminated and faked by the interests of pharmaceutical companies, insurance companies and relevant state departments. Pharmaceutical companies by various forms of influence can lead to a much higher number of drug prescription reports for even a simple health issue, whereas insurance policies can distort health data in yet other ways.

In fact, the discussion of social credit system in China predates the rise of big data (cf. Guowei, 2009; Hai-yan, 2010; Huang & Wu, 2002; Hui-yu, 2011; Huogen, 2013; Jun-yue, 2011; Tao & Mengwei, 2014; Wang & Li, 2011; Wei & Jing, 2012; Xi-qin, 2005; Yang-ping, 2007; Yong, 2012; Yu-xia, 2012; Zhao & Feng, 2002). In these early discussions of the system in the context of China, we see that financial implications of this system is overstudied eclipsing the socially detrimental consequences. Such a system was considered to be a missing pillar stone for Chinese economic development according to the authors. Interestingly

enough, some other Chinese researchers were checking similar practices in the global West, especially in the United States and Europe and discussing its feasibility for China (e.g. Mei, 2008; Xiao-fan, 2005).

However, in practice it became a social credit system "with Chinese characteristics", not only serving the banks for credit decisions, but also the government for surveillance and control. These were all before the advent of the conceptualization of big data. China due to its population size had big data all the time in history, but the conceptualization of it was recent. In that sense, the recent big data models provided the excellent technical infrastructure for the implementation of the system serving multiple, simultaneous functions of business, surveillance and control.

Following this initial lopsidedness of early research on Chinese social credit system, most of the research topics on big data, big brother and surveillance with regard to China are either understudied or completely unexplored. For one thing, the problems about big data are more fundamental than what the common discussions on the subject consider, as stated above. Before all, one needs to question the epistemology of big data which inevitably direct us to question the nature of representation and misrepresentation in big data. Thus, before moving to our more specific discussions about big data and surveillance in China, we need to theoretically discuss epistemic qualities of big data with regard to how they are recorded, measured, collected, interpreted, analyzed, serviced, distributed etc.

Data Democracy and Misrepresentations about Self and Others: Data Errors, Biases, Distortions and Manipulations

In this section, we focus on social representations excluding the natural representations. That is because the former brings even more complicated challenges to data science. As presented and discussed in Gezgin (2018), both natural and human big data lack immunity to errors in recording and measurement; and are open to errors, misuse and abuse in interpretation. On the other hand, they differ in certain dimensions: While unauthorized surveillance and ethical problems are applicable for human big data, they are applicable only at a limited sense (e.g. only for animals) for natural big data. However, manipulatability of data at generation phase as discussed in detail below distinguishes human big data from natural big data. That means humans can be manipulated to produce a particular kind of data, while nature can't be tempted (Gezgin, 2018).

In this context, for the case of human big data, the terms biases, errors, distortions, manipulations etc. are misleading and confusing with regard to information. A better term covering them would be misrepresentations. Biases and errors are unintended misrepresentations, while distortions and manipulations are

intentional. Usually, in the case of unintentional misrepresentations of others or the world in general, the infographer (or datagrapher i.e. the recorder) is not aware of his/her own biases and errors. But when they were firmly shown that they are wrong in their representation and they insist in their original misrepresentation, this becomes distortion and manipulation. Error is a problem about recording and/or measurement. For example, because of limited statistical data, we may underestimate the number of a minority group (for example Gypsies). This is an error, and I don't do it deliberately; whereas negatively representing them unintentionally (in other words without reflective thought on the subject) would be my bias. Biases can be either social or cognitive. The recorder may be socially biased against people of particular social groups; while cognitive biases are due to the organization of our mental processes.[3] For instance, a person may read only columnists that agree with his/her views. As a result, these beliefs can be viewed as the reality. E.g. by selectively reading articles against abortion all the time, the person would start to believe that majority of people are anti-abortion.[4]

Social biases are common in search engines, big data and artificial intelligence, as explained by Howard & Borenstein (2018):

"Recently, there has been an upsurge of attention focused on bias and its impact on specialized artificial intelligence (AI) applications. Allegations of racism and sexism have permeated the conversation as stories surface about search engines delivering job postings for well-paying technical jobs to men and not women, or providing arrest mugshots when keywords such as "black teenagers" are entered. Learning algorithms are evolving; they are often created from parsing through large datasets of online information while having truth labels bestowed on them by crowd-sourced masses. These specialized AI algorithms have been liberated from the minds of researchers and startups, and released onto the public. Yet intelligent though they may be, these algorithms maintain some of the same biases that permeate society. They find patterns within datasets that reflect implicit biases and, in so doing, emphasize and reinforce these biases as global truth" (p.1521).

3 Social biases of big data and AI are mentioned in the next paragraph. How about cognitive biases? Are they peculiar to Sapiens? Yes and no. Yes, since a machine can't have cognitive processes. No, because instead of cognitive processes, machines have design processes. They can be biased not only due to the big data that they process which is full of social biases; they can also be biased because of the way they were designed. For instance, in a nationalist-fascist state, it is highly likely that ethnic backgrounds of the citizens will not be recorded, as they are ignored. Another example would concern a theocratic state. In such a state, oppressed religious minorities will not be officially recognized. For example, all will be forced to be recorded as a member of the majority religion. So a recording device will be set to not record such data. This is what we call as 'design bias' and it can be a combination of Sapiens's social and cognitive biases. It can reflect governmental structure and decisions as well as designers' own representational and misrepresentational inclinations and preferences.

4 For examples of social bias see Bissell & Parrott, 2013; Proverbio, La Mastra & Zani, 2016. For cases of cognitive bias see Adame, 2016; Lieder et al., 2018.

Converging with Howard & Borenstein (2018), Loubere & Brehm (2018) state that

"After all, algorithms are not some naturally occurring phenomena, but are the reflections of the people (and societies) that create them. For this reason, the rule of algorithms must not be mistaken as an upgraded, more rational, and hyper-scientific rule of law 2.0" (p.42).

On the other hand, distortions and manipulations are completely different matters: In case of distortion, I am intentionally lying, but denying that I am lying. An obvious example for this situation is in colonial history. To justify occupation, colonialists had continuously lied. Colonial mentality is not a matter of distant past. For instance, declassified files showed that the casus belli of Vietnamese-American War (i.e. the Gulf of Tonkin Incident of 1964) was fake.[5] As stated above, distortions and manipulations are deliberate. They differ in their intervention to the social life. In distortion, the information provided is intentionally false, but it is recorded as truth. In case of manipulation, this distorted information is used to produce a situation to the advantage of the powerful. By distortion, the powerful manipulates the world affairs, and this manipulated situation is recorded as the truth. In distortion, the representation is intentionally faked; while in manipulation, the events based on distorted representations are misrepresented.

While these 2 types (intentional vs. unintentional) and 4 forms (error, bias, distortion and manipulation) of misrepresentation of information are directed towards the others, we also have self-misrepresentations. As elaborately depicted and discussed in social psychology research, a mentally sane person usually engages in self-serving misrepresentations.[6] We are inclined to think that we are good people, as well as the groups that we belong to. That may not be necessarily the case. We can misrepresent ourselves intentionally to form positive impressions, such as on Facebook, or on a date or on a job interview (which is called as selective positive self-presentations).[7] Thus, we have another dimension of misrepresentation which is self vs. others. This self does not necessarily refer to an individual. In many occasions, it refers to the groups that individual representer belongs to such as nationality, citizenship, ethnicity, gender, educational status, income group and many other typical demographic groupings. This self-misrepresentation, just like the misrepresentation of others can be intentional or unintentional. For example, a man can have a flowery shirt in a predominantly conservative society and unintentionally be considered as LGBTI by the members of the society, or he can intentionally wear particular kinds of clothes to look as such.

5 For an official discussion of how American congress and public were deceived and manipulated to wage a war against Vietnam see Paterson, 2008.
This is an article penned by a commander and published on U.S. Naval Institute's Naval History Magazine.

6 For self-serving bias, see Lammers & Burgmer, 2018; Moosa & Ramiyah, 2018; Zhang et al., 2018.

7 See Ellison, Hancock, & Toma, 2012; Gentile et al., 2012; Gonzales & Hancock, 2011; Walther, 2007.

Table 1 summarizes what we discussed so far:

Table 1. 8 Typologies of Misrepresentations in Information		
	Unintentional	Intentional
Other	M1: Error vs. M2: bias	M3: Distortion vs. M4: manipulation
Self	M5: Error vs. M6: bias	M7: Distortion vs. M8: manipulation

After we identified 8 typologies of misrepresentation in information, we can proceed to see the most common examples of misrepresentation on the basis of our model. In the rest of this section, we are mostly referring to gender-related misrepresentations, colonial misrepresentations, cultural misrepresentations which are not necessarily colonial, and overall what we call as asymmetric representations whereby the representer and the represented are categorically divided and socially segregated. For instance, there is no surprise to see that Native Americans who can be interpreted as defenders of their homeland against White occupiers in recent history are represented by non-native infographers as uncivilized folks attacking people without any reason.[8] Thus, we propose infographical studies to identify and debunk various misrepresentations. So let's see what we mean by infography and infographer (as well datagraphy and datagrapher).

The notion of infograph is mostly used in the relevant literature and professional practice to refer to a peculiar kind of graph that provides information.[9] In our case, we re-coin (i.e. coin in another meaning) the notion of infograph as a parallel to the notion of historiography. In premodern understandings of history, this discipline was about narrating historical events as they are. In that sense, history was a realistic sort of story, also evidenced by the common etymological root of the words 'history' and 'story'. In such a conceptualization, history was closer to non-fiction works. However, in modern times, the researchers realized that what constitutes history, historical events and the way historians make an account of history heavily depends on who writes the history, for whom, under what circumstances, with what purpose, with which limitations, blindspots, prejudices etc. In order to study these questions, the new field of historiography was proposed.

For one thing, history has multiple sides. The same event can be interpreted in completely different ways by different sides (e.g. liberation vs. conquest vs. loss vs. fall vs. occupation etc. of a city such as Istanbul/Constantinopolis) or can even be blatanly distorted (eg. Nanking massacre (a major, bloody event, a massive war crime by Japanese colonial forces in Chinese historiography) vs. Nanking incident (a minor, negligible, small scale event in Japanese great war

8 For an alternative account of American history see Zinn, 2015.
9 Eg. Dunleavy, 2005; Li et al., 2018; Otten, Cheng, & Drewnowski, 2015.

heroism narratives).[10] Secondly, we can focus on palace residents' lives to write a typical official history vs. people's lives to write a people's history. Historiography opened the eyes of the historians. We also need such an awakening in information and data science. That is why we propose the new field of infography to study unintentional and intentional misrepresentations of others and self which are directly related to the raison d'être of historiography. In other words, who represents whom on what purpose determines the content of the information and data. Infography will help us to democratize information and data science which have been used, misused and abused by what we call as twin big brothers (i.e. repressive governments and greedy corporations).[11]

As a case example, we can take gender-related misrepresentations: Until recently, most of the representers were male. As a consequence, until the arrival of strong woman rights movements, women were represented in negative terms. Rather than rightly associating the disadvantaged position of women with the patriarchal society, the representers blamed the women and described the social gender inequalities in natural and biological terms as if women can't be, for instance, pilots, presidents, engineers, philosophers etc. As a second example, in many cases, colonial occupiers were the first to bring writing and/or large-scale recording to certain cultures. So from the very beginning, the representer was absolutely superior to the represented. In other cases, we see that the representational resources of cultures with longer written and recorded histories such as India and China were disabled to give way to globally Western models of representations. Although colonial times are over, it is not the case for colonial legacies. There is no realistic way to think about how India and China would look like if colonial occupiers would not interfere with and interrupt their development. The global West's representations and misrepresentations of India and China have had detrimental effects on their self-image.

This representational asymmetry is also visible between governments and citizens on the one hand, and corporations and consumers on the other. This double asymmetry in this form is a serious threat to democracy and people's welfare in general. In a truly democratic welfare state, there should be no digital divide, and representational tools should be freely and equally available to every citizen. This requires elucidation of the infographic circumstances of the representations so that citizens can be aware of potential misrepresentations. This will open the gates to the notion of data democracy where datagraphers will be accountable and transparent.[12]

10 For Nanking massacre narratives see Schwartz, 2012; Wang, 2009; Xu & Spillman, 2010.

11 Before moving on, let us also note that we use infography interchangeably with datagraphy.

12 For further discussions of data democracy see Baack, 2015; Feng & China, 2015; Ruijer & Martinius, 2017.

To conclude this section, information science in general and data science in particular should be supplemented by infography/datagraphy as a new research area to cope with misrepresentations and in order to move towards informational/data democracy. These reflections on epistemology of big data and democracy will help us interpret the situation in China in a more systematic way. They will also be conducive for identification of understudied topics in the relevant areas.

Big Data and Surveillance in China

As Hodson (2015), Loubere & Brehm (2018) and Botsman (2017) correctly point out, neither surveillance nor social credit system are peculiar to China. The companies are watching actual and potential customers, and as Snowden revelations uncover, state surveillance is omnipresent and quotidian in the global West. Thus, pointing our fingers to China under such circumstances and political conjuncture can be considered as a form of double standard and even a case of orientalism. That is why, in the relevant discussions, the fact that China is not the only country surveilling its citizens should be kept in mind. However, through the social credit system with -what we call as- "Chinese characteristics", the level of surveillance becomes elevated, integrated, systematic and extremely invasive. Furthermore, in the global West, there are still a set of legal channels to challenge surveillance, although that is not applicable for Snowden-like secret ones.[13] Additionally, people have higher chances to express their views, challenge the big brothers, organize against them and ultimately resist them in the global West. That is why, Chinese social credit system worths discussions separately. For Loubere & Brehm (2018), social credit is "more than simply a Chinese version of big brother: it is an unprecedented climax of the global financialisation project and a signal of a potential dark digital future dominated by algorithmic rule" (p.38).

Hodson (2015) delineates the Chinese Social Credit System in the following way on 'New Scientist':

"Where you go, what you buy, who you know, how many points are on your driving licence: these are just a few of the details that the Chinese government will track – to give scores to all its citizens. China's Social Credit System (SCS) will come up with these ratings by linking up personal data held by banks, e-commerce sites and social networks. The scores will serve not just to indicate an individual's credit risk, for example, but could be used by potential landlords, employers and even romantic partners to gauge an individual's character."

(...)

"It assigns people a score of up to 950 points based on factors such as how

13 For discussions of the legal conundrums of state and corporate surveillance with regard to citizenship rights and freedoms see Joh, 2016; Miller, 2014.

often they shop and their general credit history. Spending more through Alibaba's payment app, Alipay, or connecting to more friends via Sesame Credit can raise your score. The higher your score, the more privileges it opens up. People scoring above 600 can rent cars from the Chinese companies Car Inc and eHai.com, without putting down a deposit. A score above 650 lets you check out of hotels faster, while more than 700 cuts the paperwork when applying for visas to Singapore" (p.22).

Botsman (2017) presents the system's evaluation criteria:

"So just how are people rated? Individuals on Sesame Credit are measured by a score ranging between 350 and 950 points. Alibaba does not divulge the "complex algorithm" it uses to calculate the number but they do reveal the five factors taken into account. The first is credit history. For example, does the citizen pay their electricity or phone bill on time? Next is fulfilment capacity, which it defines in its guidelines as "a user's ability to fulfil his/her contract obligations". The third factor is personal characteristics, verifying personal information such as someone's mobile phone number and address. But the fourth category, behaviour and preference, is where it gets interesting.

Under this system, something as innocuous as a person's shopping habits become a measure of character. Alibaba admits it judges people by the types of products they buy. "Someone who plays video games for ten hours a day , for example, would be considered an idle person," says Li Yingyun, Sesame's Technology Director. "Someone who frequently buys diapers would be considered as probably a parent, who on balance is more likely to have a sense of responsibility." So the system not only investigates behaviour - it shapes it. It "nudges" citizens away from purchases and behaviours the government does not like.

Friends matter, too. The fifth category is interpersonal relationships. What does their choice of online friends and their interactions say about the person being assessed? Sharing what Sesame Credit refers to as "positive energy" online, nice messages about the government or how well the country's economy is doing, will make your score go up.

(...)

"a person's own score will also be affected by what their online friends say and do, beyond their own contact with them. If someone they are connected to online posts a negative comment, their own score will also be dragged down" (np.).

As it can be inferred from these quotations, in China, credit ratings in theory turned out to be life ratings in practice, and it is a global trend also visible in the global West (Botsman, 2017). In addition to ethical concerns and problems associated with breach of privacy, and accordingly citizenship rights, this system

obviously has the potential to promote obedience and even reporting on others which will eventually destroy social trust and cohesion. This is an ironic situation, as mentioned by Loubere & Brehm (2018): "While social credit can be seen as an outgrowth of our collective impulse to achieve a more trustworthy society, a unified fully-functioning social credit system will ultimately turn the quest for trust through transparency and accountability upside down (...)" (p.39).

It is also predicted that Chinese social credit system will exacerbate the income inequalities and various forms of social inequalities (Loubere & Brehm, 2018), as the high and middle income citizens and those of advantaged backgrounds will be generally considered as more trustworthy under this system, as they have more resources to prove that they are 'model citizens'. Although there will of course be exceptions, not all the citizens are on the same starting line vis-a-vis the surveilling state and corporations, in the conceptualization of getting surveilled as a competition to win favors. In that sense, the social credit system in its current form can ultimately alter the functions of the state. For instance, initially it was hoped that big data could optimize allocative efficiency in finding poverty-stricken students and offering them scholarship and other forms of compensation in China (Cao & Wang, 2016). However, these students will be at disadvantage without the necessary resources in this race of popularity. Thus without benevolent state intervention, poorest of the poor will have no chance to win awards in the system. Let's forget about the system; even at the recording and measurement stage of their data, it is highly likely that the poorest will be at disadvantage. This is especially so in a country where wealth is considered to be the major indicator of success. In case of poverty, Chinese and Western public opinion, due to psychological and sociological factors is inclined to blame the victims, i.e. poor people.[14] It is usually assumed that the world is just, thus if a person is poor, that is because of him/her, not because of capitalism or the unfair political/economic system in general.[15]

From another perspective, we can view Chinese social credit system as a form of financialization of citizenship (Loubere & Brehm, 2018) as well as panopticalization of the relationship between the state and the citizens. With the former process, citizenship gets commodified, while with the latter one, constant feeling of being watched leads to permanent changes in behavior. This commodification process is associated with the notion of 'surveillance economy' which is defined as "a capitalist system in which the drive to encapsulate everything within "big data" is the engine driving economic growth and profitability" (Hirst, 2014, p.20).

14 For research on blaming the victim in case of poverty see Buraschi, Bustillos, & Huici, 2018; Chagnon, 2017; Godfrey, & Wolf, 2016.

15 This is called as 'the just world hypothesis'. For research on this subject, see Begue, & Bastounis, 2003; Furnham, 2003; Goodman & Carr, 2017.

As to the second process, panopticalization, the Hawthorne effect, rather than an exception, becomes the norm; which can be generally defined as "a change in behaviour as a response to observation and assessment" (Sedgwick & Greenwood, 2015, p.1).[16] Citizens will modify their behaviors due to the expectations and fear that they are being surveilled everywhere and all the time. That means state surveillance will no longer involve recording and collecting citizen data only, it will also lead to manipulation of citizen data, in one sense resembling the Facebook-Cambridge Analytica scandal.[17]

When the behavior changes become permanent, stable and long-term, we will even be able to talk about personality change. A number of studies especially Sennett (2001) shows us how capitalism forge and mold employee's personalities.[18] The same can be expected under Chinese surveillance system, but at a far larger scale. This reminds us the Heisenberg principle of uncertainty from quantum physics which means the mere act of observation changes the data observed.[19] However, the analogy is not completely applicable: In case of quantum physics, because of this problem, we can't collect the data, whereas in Chinese surveillance system, it will be possible to record the data albeit that their authenticity and genuineness will be problematic.

Although, the Chinese social credit system is predominantly a matter of China and Chinese, its success can direct other states to imitate or adopt this model for their countries (Loubere & Brehm, 2018). Belt and Road initiative (also known as the New Silk Road) has the potential to accelerate spread of Chinese surveillance models and social credit systems as Chinese IT companies are among the top investors on the designated itineraries.[20]

As we had discussed extensively in previous sections, error management is Achilles' heel of the social credit system even in the best case scenario without all these ethical and political problems. The 8 forms of misrepresentations as previously explained in Table 1 referring to errors, biases, distortions and mani-

16 The Hawthorne effect was proposed in industrial psychology field. However, recently it is extensively researched with regard to health areas which helps us understand how a more generic surveillance would influence human behavior other than work settings which are more limited in scope. For more information about the Hawthorne effect at health contexts, see Berthelot, Nizard, & Maugars, 2018; Chen et al., 2015; Rosenberg et al., 2018.

17 For more details and scholarly discussion on Facebook-Cambridge Analytica scandal see Berghel, 2018; Isaak & Hanna, 2018; Laterza, 2018.

18 For some other relevant works about the effects of capitalism on personality see Fernando, 2018; Fitzi & Mele, 2017; O'Connor & Crome, 2016.

19 For Heisenberg principle see D'Ariano, 2003. D'Ariano (2003) alternatively calls it as information-disturbance trade-off which looks like even more inspirational for social sciences. For an early elobaration of the applicability of the principle in social sciences see Sposito, 1969. Later on Martin (1981) discusses the same principle with regard to measurement of crimes.

20 For instance, Reuters reports that Chinese IT companies are helping Venezuela to install a monitoring system to track Venezuelan citizens (Berwick, 2018).

pulations about self and others require serious attention, otherwise the realism, reliability and authenticity of the big data and social credits will be a moot issue.

Conclusion

In this article, we pointed out a number of problems never or almost never considered in Chinese big data discussions. Epistemologically speaking, we tried to show that Chinese big data, surveillance and the social credit system are not as sound as the recorders wished for. As mentioned earlier, without a proper model of error management, Chinese big data will have the potential to replicate hidden and obvious misrepresentations of self and others such as errors, biases, distortions and manipulations innumerable times. Thus, in addition to ethical, social, political and axiological concerns over Chinese surveillance system, and even prior to such discussions, epistemology is problematic. The system in its current form can't guarantee the desired and expected epistemic qualities for the collected and to-be-collected data

References

Adame, B. J. (2016). Training in the mitigation of anchoring bias: A test of the consider-the-opposite strategy. Learning and Motivation, 53, 36-48.

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. Big Data & Society, 2(2), 2053951715594634.

Begue, L., & Bastounis, M. (2003). Two spheres of belief in justice: Extensive support for the bidimensional model of belief in a just world. Journal of Personality, 71(3), 435-463.

Berghel, H. (2018). Malice Domestic: The Cambridge Analytica Dystopia. Computer, 5, 84-89.

Berthelot, J. M., Nizard, J., & Maugars, Y. (2018). The Negative Hawthorne Effect: Explaining Pain Overexpression. Joint Bone Spine.

doi: 10.1016/j.jbspin.2018.10.003

Berwick, A. (2018). How ZTE helps Venezuela create China-style social control. Reuters, 14.11.2018. https://www.reuters.com/investigates/special-report/venezuela-zte/

Bissell, K., & Parrott, S. (2013). Prejudice: The role of the media in the development of social bias. Journalism & Communication Monographs, 15(4), 219-270.

Botsman, R. (2017). Big data meets Big Brother as China moves to rate its citizens. Wired UK, 21.10.2017. https://www.wired.co.uk/article/chinese-go-

vernment-social-credit-score-privacy-invasion

Buraschi, D., Bustillos, A., & Huici, C. (2018). Attitudes toward Immigrants, Beliefs about Causes of Poverty and Effects of Perspective-Taking. The Spanish journal of psychology, 21.

Cao, X., & Wang, Y. (2016). An Approach to Granting Subsidies to College Students in China Using Big Data. Journal of Education and Practice, 7(26), 1-4.

Chagnon, N. J. (2017). Racialized Culpability: Victim Blaming and State Violence. In Race, Ethnicity and Law (pp. 199-219). Emerald Publishing Limited.

Chen, L. F., Vander Weg, M. W., Hofmann, D. A., & Reisinger, H. S. (2015). The Hawthorne effect in infection prevention and epidemiology. infection control & hospital epidemiology, 36(12), 1444-1450.

D'Ariano, G. M. (2003). On the Heisenberg principle, namely on the information-disturbance trade-off in a quantum measurement. Fortschritte der Physik: Progress of Physics, 51(4-5), 318-330.

Dunleavy, D. (2015). Data visualization and infographics. Visual Communication Quarterly, 22(1), 68-68.

Ellison, N. B., Hancock, J. T., & Toma, C. L. (2012). Profile as promise: A framework for conceptualizing veracity in online dating self-presentations. new media & society, 14(1), 45-62.

Feng, G., & China, O. D. (2015). Open data: concepts, status and opportunities. Big Data Research, (2), 2.

Fernando, W. D. A. (2018). Exploring character in the new capitalism: a study of mid-level academics' in a British research-intensive university. Studies in Higher Education, 43(6), 1045-1057.

Fitzi, G., & Mele, V. (2017). The corrosion of character: Work and personality in the modern age. Journal of Classical Sociology, 17(2), 143-155.

Furnham, A. (2003). Belief in a just world: Research progress over the past decade. Personality and individual differences, 34(5), 795-817.

Gentile, B., Twenge, J. M., Freeman, E. C., & Campbell, W. K. (2012). The effect of social networking websites on positive self-views: An experimental investigation. Computers in Human Behavior, 28(5), 1929-1933.

Gezgin, U. B. (2018). An invitation to critical social science of big data: from critical theory and critical research to omniresistance. AI & SOCIETY, 1-9.

doi: 10.1007/s00146-018-0868-y

Godfrey, E. B., & Wolf, S. (2016). Developing critical consciousness or justifying the system? A qualitative analysis of attributions for poverty and wealth among low-income racial/ethnic minority and immigrant women. Cultural

Diversity and Ethnic Minority Psychology, 22(1), 93-103.

Gonzales, A. L., & Hancock, J. T. (2011). Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. Cyberpsychology, behavior, and social networking, 14(1-2), 79-83.

Goodman, S., & Carr, P. (2017). The just world hypothesis as an argumentative resource in debates about unemployment benefits. Journal of Community & Applied Social Psychology, 27(4), 312-323.

Guowei, W. (2009). A Tentative Discussion on Establishing Social Credit System in China. Journal of Shanxi Radio & TV University, (6), 34.

ttp://en.cnki.com.cn/Article_en/CJFDTOTAL-SXGB200906034.htm

Hai-yan, W. A. N. G. (2010). A Legal & Economic Analysis on Social Credit Crisis in China. Credit Reference, (6), 5.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-HNJG201006005.htm

Hirst, M. (2014). "Big Data" Is Creating a Surveillance Economy. Issues magazine, 109, 19-21.

Hodson, H. (2015). Big brother is rating you (if you're Chinese). New Scientist, 228(3043), 22–23. doi:10.1016/s0262-4079(15)31412-3

Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. Science and engineering ethics, 24(5), 1521-1536.

Huang, J. X., & Wu, Y. (2002). Some problems in establishment of social credit system in China [J]. Journal of Harbin Institute of Technology (Social Sciences Edition), 3.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-HRBG200203014.htm

Huang, X. (2018, August). Research on the Legal Protection of Personal Information in China in the Environment of Big Data. In 3rd International Conference on Judicial, Administrative and Humanitarian Problems of State Structures and Economic Subjects (JAHP 2018). Atlantis Press.

https://download.atlantis-press.com/article/25902184.pdf

Huogen, W. (2013). The Development Trend and Its System Construction of Social Credit in China [J]. Chongqing Social Sciences, 7.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-CQSK201307008.htm

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. Computer, 51(8), 56-59.

Joh, E. E. (2016). The new surveillance discretion: Automated suspicion, big data, and policing. Harvard Law & Policy Review, 10, 15-42.

Jun-yue, L. I. N. (2011). Social Credit System: China′ s Model for Efficiently

Establishing Credit Reference System [J]. Credit Reference, 2.
http://en.cnki.com.cn/Article_en/CJFDTOTAL-HNJG201102002.htm

LaBrie, R. C., Steinke, G. H., Li, X., & Cazier, J. A. (2018). Big data analytics sentiment: US-China reaction to data collection by business and government. Technological Forecasting and Social Change, 130, 45-55.

Lammers, J., & Burgmer, P. (2018). Power increases the self-serving bias in the attribution of collective successes and failures. European Journal of Social Psychology.

Laterza, V. (2018). Cambridge Analytica, independent research and the national interest. Anthropology Today, 34(3), 1-2.

Li, N., Brossard, D., Scheufele, D. A., Wilson, P. H., & Rose, K. M. (2018). Communicating data: interactive infographics, scientific data and credibility. Journal of Science Communication, 17(2), A06.

Lieder, F., Griffiths, T. L., Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. Psychonomic bulletin & review, 25(1), 322-349.

Loubere, N., & Brehm, S. (2018). The Global Age of Algorithm: Social Credit and the Financialisation of Governance in China. Made in China: A Quarterly on Chinese Labour, Civil Society, and Rights, 3(1), 38-43.

Martin, E. (1981). A twist on the heisenberg principle: Or, how crime affects its measurement. Social Indicators Research, 9(2), 197-223.

Mei, H. (2008). Establishing Efficient Social Credit System in China from European Experience of Social Credit System/ Etablissement du Systeme de Credit Social Efficace en Chine d'apres les Experiences du Systeme de Credit Social Europeen. Canadian Social Science, 4(1), 1-3.

Miller, K. (2014). Total surveillance, big data, and predictive crime technology: Privacy's perfect storm. Journal of Technology of Law & Policy, 19, 105-146.

Moosa, I. A., & Ramiah, V. (2017). Overconfidence and Self-Serving Bias. In The Financial Consequences of Behavioural Biases (pp. 45-69). Palgrave Macmillan, Cham.

O'Connor, P., & Crome, K. (2016). Learning together: Foucault, Sennett and the crisis of the co-operative character. Journal of Co-operative Studies, 49(2), 30-42.

Otten, J. J., Cheng, K., & Drewnowski, A. (2015). Infographics and public policy: using data visualization to convey complex information. Health Affairs, 34(11), 1901-1907.

Paterson, P. (2008). The Truth About Tonkin. Naval History Magazine, 22(1), np.

https://www.usni.org/magazines/naval-history-magazine/2008/february/truth-about-tonkin

Proverbio, A. M., La Mastra, F., & Zani, A. (2016). How negative social bias affects memory for faces: An electrical neuroimaging study. PloS One, 11(9), e0162671.

Rosenberg, M., Pettifor, A., Twine, R., Hughes, J. P., Gomez-Olive, F. X., Wagner, R. G., Sulaimon, A., Tollman, S., Selin, A., MacPhail, C. & Kahn, K. (2018). Evidence for sample selection effect and Hawthorne effect in behavioural HIV prevention trial among young women in a rural South African community. BMJ Open, 8(1), e019167.

doi: 10.1136/bmjopen-2017-019167

Ruijer, E. H., & Martinius, E. (2017). Researching the democratic impact of open government data: A systematic literature review. Information Polity, 22(4), 233-250.

Saran, R., Steffick, D., & Bragg-Gresham, J. (2017). The China Kidney Disease Network (CK-NET):"Big Data—Big Dreams". American Journal of Kidney Diseases, 69(6), 713-716.

Schwartz B (2012) Rethinking conflict and collective memory. The case of Nanking. In J. Alexander, R. Jacobs and P. Smith (eds.). The Oxford Handbook of Cultural Sociology. New York: Oxford University Press, pp. 529–563.

Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. Bmj, 351, h4672.

Sennett R (2001) The Corrosion of Character: The Personal Consequences of Work in the New

Capitalism. New York: Routledge.

Sposito, G. (1969). IV. Does a generalized Heisenberg principle operate in the social sciences?. Inquiry: An Interdisciplinary Journal of Philosophy, 12(1-4), 356-361.

Tao, Y., & Mengwei, Y. (2014). The internal operating mechanism of the construction of social credit system in contemporary China. Theory and Modernization, 1(1).

http://en.cnki.com.cn/Article_en/CJFDTotal-LYXD201401001.htm

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. Computers in Human Behavior, 23(5), 2538-2557.

Wang, X. J., & Li, S. T. (2011). Thoughts on Construction of Social Credit System in Liaoning Province, China. Credit Reference, (1), 19.

http://en.cnki.com.cn/Article_en/CJFDTotal-HNJG201101019.htm

Wang, Z. (2009). Old wounds, new narratives: Joint history textbook writing and peacebuilding in East Asia. History & Memory, 21(1), 101-126.

Wei, Z., & Jing, C. (2012). The Collaborative Governance: The Selected Mode of Social Credit System in China [J]. Social Sciences in Nanjing, 11.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-NJSH201211017.htm

Xiao-fan, Z. (2005). Construction of Social Credit Reporting System: International Comparison and Domestic Study——on Integration of Personal Credit Reporting and Firm Credit Reporting of China [J]. The Theory and Practice of Finance and Economics, 2.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-CLSJ200502021.htm

Xi-qin, W. (2005). Social Credit Structure and Improvement of Social Credit Environment in China. Journal of Wuhan University of Science and Technology, (2), 12.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-WHKS200502012.htm

Xin, J. (2007). Establishing Efficient Social Credit System in China from American Experience of Social Credit System/Etablissement d'un Système Efficace de Crédit Social en Chine d'après les Expériences Américaines. Canadian Social Science, 3(6), 64.

Xu, X., & Spillman, L. (2010). Political centres, progressive narratives and cultural trauma: Coming to terms with the Nanjing Massacre in China, 1937–1979. In Northeast Asia's Difficult Past (pp. 101-128). Palgrave Macmillan, London.

Yang-ping, H. (2007). " Three-layer Theories": The Path Choice of Social Credit System Construction in China. Journal of Jimei University (Philosophy and Social Sciences), (3), 10.

http://en.cnki.com.cn/Article_en/CJFDTOTAL-JMDZ200703010.htm

Yong, H. (2012). Countermeasures for Improving Social Credit Reference System of Individual Credit in China. Credit Reference, (5), 16.

http://en.cnki.com.cn/Article_en/CJFDTotal-HNJG201205016.htm

Yu-xia, X. (2012). Thoughts about Strengthening Construction of Social Credit System in China. Credit Reference, (2), 13.

http://en.cnki.com.cn/Article_en/CJFDTotal-HNJG201202013.htm

Zhang, L., Wang, H., Li, Q., Zhao, M. H., & Zhan, Q. M. (2018). Big data and medical research in China. bmj, 360, j5910.

Zhang, S. (2015). Big Data for Healthcare in China: a Review of the State-of-art. Cambridge Journal of China Studies, 10(1), 32-40.

Zhang, Y., Pan, Z., Li, K., & Guo, Y. (2018). Self-Serving Bias in Memories.

Experimental psychology.

Zhang, Z. (2014). Big data and clinical research: focusing on the area of critical care medicine in mainland China. Quantitative Imaging in Medicine and Surgery, 4(5), 426-429.

Zhang, Z. (2017). Release of the national healthcare big data in China: a historic leap in clinical research. AME Medical Journal, 2(19).

Zhao, H., & Dong, H. (2017). Research on personal privacy protection of China in the era of big data. Open Journal of Social Sciences, 5, 139-145.

Zhao, W. L., & Feng, X. F. (2002). Current Situation and Improvement of Social Credit in the Transition Period of China. Journal of Xi'an Jiaotong University (Social Sciences Edition), (2), 22. http://en.cnki.com.cn/Article_en/CJFDTOTAL-XAJD200202022.htm

Zinn, H. (2015). A people's history of the United States: 1492-present. Routledge.