

# Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines "Good" and "Bad" Behavior

Severin Engelmann\*

engelmas@in.tum.de

Chair of Cyber Trust

Faculty of Informatics

Technical University of Munich

Mo Chen\*

mo.chen@tum.de

Chair of Cyber Trust

Faculty of Informatics

Technical University of Munich

Felix Fischer<sup>+</sup>

flx.fischer@tum.de

Chair of Cyber Trust

Faculty of Informatics

Technical University of Munich

Ching-Yu Kao

ching-yu.kao@aisec.fraunhofer.de

Fraunhofer Institute for Applied and

Integrated Security

Jens Grossklags

jens.grossklags@in.tum.de

Chair of Cyber Trust

Faculty of Informatics

Technical University of Munich

## ABSTRACT

China's Social Credit System (SCS, 社会信用体系 or shehui xinyong tixi) is expected to become the first digitally-implemented nationwide scoring system with the purpose to rate the behavior of citizens, companies, and other entities. Thereby, in the SCS, "good" behavior can result in material rewards and reputational gain while "bad" behavior can lead to exclusion from material resources and reputational loss. Crucially, for the implementation of the SCS, society must be able to distinguish between behaviors that result in reward and those that lead to sanction. In this paper, we conduct the first transparency analysis of two central administrative information platforms of the SCS to understand how the SCS currently defines "good" and "bad" behavior. We analyze 194,829 behavioral records and 942 reports on citizens' behaviors published on the official Beijing SCS website and the national SCS platform "Credit China", respectively. By applying a mixed-method approach, we demonstrate that there is a considerable asymmetry between information provided by the so-called Redlist (information on "good" behavior) and the Blacklist (information on "bad" behavior). At the current stage of the SCS implementation, the majority of explanations on blacklisted behaviors includes a detailed description of the causal relation between inadequate behavior and its sanction. On the other hand, explanations on redlisted behavior, which comprise positive norms fostering value internalization and integration, are less transparent. Finally, this first SCS transparency analysis suggests that socio-technical systems applying a scoring mechanism might use different degrees of transparency to achieve particular behavioral engineering goals.

We thank the anonymous reviewers for their comments. We appreciate the support from the German Institute for Trust and Safety on the Internet (DIVSI).

\* Severin Engelmann and Mo Chen equally contributed to this work.

<sup>+</sup> Felix Fischer is also affiliated with Projects by IF, London.

FAT\* '19, January 29–31, 2019, Atlanta, GA, USA

© 2019 Association for Computing Machinery.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in FAT\* '19: *Conference on Fairness, Accountability, and Transparency (FAT\* '19)*, January 29–31, 2019, Atlanta, GA, USA, <https://doi.org/10.1145/3287560.3287585>.

## CCS CONCEPTS

• **Social and professional topics** → *Government technology policy*; • **Information systems** → *Decision support systems*; • **Security and privacy** → *Social aspects of security and privacy*; • **Applied computing** → *Anthropology*;

## KEYWORDS

Social Credit System, Socio-Technical Systems, Transparency, Behavioral Engineering.

### ACM Reference Format:

Severin Engelmann\*, Mo Chen\*, Felix Fischer<sup>+</sup>, Ching-Yu Kao, and Jens Grossklags. 2019. Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines "Good" and "Bad" Behavior. In *FAT\* '19: Conference on Fairness, Accountability, and Transparency (FAT\* '19)*, January 29–31, 2019, Atlanta, GA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3287560.3287585>

## 1 INTRODUCTION

Moral thinking and action necessarily depend on informational resources. When an individual asks: "What is the right thing to do?", he or she essentially relies on information that renders a conclusion morally justifiable. In philosophy and anthropology, descriptive morality refers to how groups or societies negotiate codes of conduct (or norms) that are morally acceptable or unacceptable [8, 36]. As a consequence, an individual's moral accountability tends to be proportional to his or her knowledge of good and bad moral behavior underlining the epistemic character of morality [7]. In 2014, the Chinese government issued a plan for a nationwide digital scoring system known as the Chinese Social Credit System (SCS) classifying behavior into morally "praise-" and "blameworthy" [29]. Thereby, all legal entities including companies and public institutions (among others) receive an 18-digit ID called the Unified Social Credit Code,<sup>1</sup> which corresponds to the 18-digit ID card number for Chinese citizens. Presumably, based on these IDs, the SCS will collect and evaluate behavioral data and may assign scores that result in material benefits and reputational praise or material exclusion and reputational loss. Or, in the words of the Chinese government,

<sup>1</sup>[http://www.gov.cn/zhengce/content/2015-06/17/content\\_9858.html](http://www.gov.cn/zhengce/content/2015-06/17/content_9858.html), last accessed on November 19, 2018.

the goal of the SCS is to “allow the trustworthy to roam everywhere under heaven while making it hard for the discredited to take a single step” [21, 29].

But how can citizens, companies, and social institutions know what behaviors are “good” and “bad” in the SCS? Put differently, how transparent is the current SCS in providing information on “good” and “bad” behaviors? Answering this question requires a conceptualization of transparency. Here, we rely on the definition proposed by Turilli and Floridi, which conceptualizes transparency as “the choice of which information is to be made accessible to some agents by an information provider”[30]. First, this definition distinguishes between an information provider, which makes information accessible, in this context the Chinese government, and agents or entities that depend on this information for their decision-making. Secondly, this definition recognizes that information transparency is an “ethically impairing or enabling factor when the information disclosed has an impact on ethical principles”[30]. Both of these components are highly relevant for the SCS since participants are dependent on the information provided to make decisions that can lead to reward or punishment.

Recently, the Chinese government has started issuing behavioral information on several platforms (see Section 2 for more information). In this empirical study, we review a subset of this behavioral information released on two central SCS platforms: the official SCS national website “Credit China” and its equivalent municipal outlet “Credit China (Beijing)”.

On the former site, we collect and analyze 156 news reports about “good” behaviors (we refer to as “positive” cases), and 789 equivalent reports about “bad” behaviors (“negative” cases). In these “negative” portraits, individuals are commonly stereotyped as so-called “Lao-lai (老赖)” – the epitome of a financially dishonest individual in China. Since all stories we collected are news reports about real-life events portraying a morally “good” or “bad” individual, they all include descriptive norms highlighting “desirable” and “undesirable” characteristics of individuals in Chinese society today.

Next, on “Credit China (Beijing)”, we retrieve a large number of records of “good” and “bad” behavior from the so-called Redlist and Blacklist. Thus, our approach is as follows: first, we collect and statistically analyze close to 200,000 Blacklist and Redlist records from “beijing.gov.cn/credittbj”, the SCS’s information platform for China’s capital, Beijing. Hence, based on machine learning topic modeling and manual text coding, we identify the common semantic patterns of close to 1000 reports on “good” and “bad” behavior published on the national SCS platform “www.creditchina.gov.cn”.

We show several informational asymmetries that characterize the current degree of transparency of the governmental SCS’s information platforms. Finally, we discuss how degrees of transparency could correspond to different incentive strategies of socio-technical systems that rate legal entities in society.

Our paper has the following structure. In Section 2, we discuss the development of China’s SCS and review related work. In Section 3, we present our data acquisition and data analysis approach. We conduct our analysis in Sections 4 and 5. We discuss our results and offer concluding remarks in Section 6.

## 2 BACKGROUND

The implementation of the SCS rests on at least three main factors: First, lack of honesty and trust<sup>2</sup> in Chinese society has become a serious issue manifested in the numerous news reports about food poisonings, chemical spills, financial and telecommunications fraud, and academic dishonesty over the past two decades [13, 22]. It is estimated that Chinese enterprises suffer from a loss of 600 billion RMB (around 92 billion USD) per year due to dishonest activities<sup>3</sup>. According to a survey conducted by Ipsos Public Affairs [14], “moral decline” was regarded as the most serious issue in China in 2017. 47% of Chinese respondents ranked it as one of the top 3 greatest concerns, while the same issue was only mentioned by 15% of total respondents worldwide.

Secondly, China’s SCS is expected to boost the domestic economy. The Chinese government hopes that the SCS will give millions of Chinese citizens without a financial history access to credit and investment opportunities in the domestic market. China has the largest unbanked population in the world (in absolute numbers), with more than 225 million citizens having no bank account [5]. So far, only 320 million Chinese citizens have a credit record<sup>4</sup>. However, the sustainability of China’s economic growth partially depends on an increase in domestic spending. Through the SCS, citizens could apply for loans based on trustworthiness scores without having to prove their financial creditworthiness.

Finally, in Chinese society, the concept of personal identity is largely determined by Confucian principles [6, 32]. Accordingly, personhood is supposed to extend from the private to the public sphere thereby somewhat losing its private and public boundaries. In other words, normative expectations on individuals hardly account for the distinction between a private and a public sphere. The division between a private and a public persona is often conceived as trying to be secretive as privacy is commonly conceived as hiding something shameful [34]. In fact, until recently, privacy was primarily protected under the right of reputation in Chinese civil law [33]. At the same time, the public interest ranks highly in Chinese civil law [3]: “private information protected from disclosure refers to information that is irrelevant to the public interest or to the interests of other persons.” However, while the Chinese concept of privacy is evolving, it is expected to remain distinct from other societies [18]. Overall, the introduction of the SCS is hardly perceived as a privacy-violating system in Chinese society, which is perhaps surprising from a Western perspective [16].

### 2.1 Current state of the SCS

At the current stage, the SCS remains fragmented, being developed at national, provincial, municipal, and ministerial levels with no clear unified structure. In the past years, provinces and cities have developed various prototype models for the SCS [17, 35]. Importantly, the SCS also takes companies, government departments and

<sup>2</sup>The characters “诚信 (chengxin)” literally mean both honesty and trust in Chinese.

<sup>3</sup>This information is included in the “Report on China’s Honesty Building Situation (Zhongguo Chengxin Jianshe Zhuangkuang Baogao)”. The full report is not publicly available, but parts of the report (in Chinese) are accessible through: <http://society.people.com.cn/n1/2016/0523/c1008-28370202.html>, last accessed on November 19, 2018.

<sup>4</sup>See “Inspiration of the US Non-traditional Credit Information Mechanism” available on the platform of “Credit China” at [http://www.creditchina.gov.cn/zhengcefa/gui/tashanzhishi/201712/t20171207\\_98701.html](http://www.creditchina.gov.cn/zhengcefa/gui/tashanzhishi/201712/t20171207_98701.html), last accessed on November 19, 2018.

judicial organizations as its targets [29]. This means that some companies have a special role in the SCS. Since 2015, eight companies were granted permission to run individual credit services with the purpose to implement pilot SCS programs [23]. Individually, none of the eight companies received a licence to continue individual credit services after the two-year trial period ended in 2017. Instead, together with the China Internet Finance Association (run by the People’s Bank of China), they recently have become common shareholders of a company called Baihang Credit, which received the first credit scoring licence in February 2018.

## 2.2 Related Work

We are unaware of *any* research project that conducts a data-driven analysis of the currently observable data practices of key sites of China’s SCS. However, we have identified two empirical research studies that help understand how the SCS is being communicated and discussed by Chinese media [23], and how it is being perceived by Chinese citizens [16].

Ohlberg et al. collected official Chinese news articles and public communications, as well as social media postings on Chinese blogs, forums, and bulletin board services about the SCS for a six-month period in 2017 [23]. The large majority of news articles has a positive focus and highlight the SCS as a “cure-all for social and economic problems”. Criticism is mostly aimed at the slow implementation progress or directed at commercial initiatives in the SCS. Citizens’ social media postings rarely address privacy issues and rather focus on how to game the system to achieve a higher social credit score within commercial SCS applications. Of relevance to the latter point, the implications of gamifying social credit are also being discussed from a non-empirical perspective by other scholars [19, 24].

Kostka [16] conducted an online survey with about 2,200 Chinese citizens that was distributed via different channels including websites and apps. Due to the widespread internet surveillance in China, the validity of such online surveys remains questionable at least to some extent. According to her findings, about 80% of the respondents have a positive perception of governmental and commercial SCS initiatives. Interestingly, older and more educated respondents have a higher approval rating. In contrast, these demographic factors are typically associated with higher privacy concerns in Western societies (see, for example, [1]). Several policy papers address the relationship between the SCS and the danger of mass surveillance (e.g., [20]).

Finally, there is rigorous work on comparing financial credit reporting systems [15], which, however, predates the emergence of the SCS in China and focuses on the financial aspects of credit reporting. Likewise, privacy considerations concerning private entities facilitating credit and background reporting have, for example, been explored by Hoofnagle [12].

## 2.3 Ethical Issues

Our analysis is built on publicly available data from key sites of China’s SCS, which is posted with the intent of public scrutiny. Our paper includes screenshots from the *currently available* implementations. We have blurred any personally identifiable data.

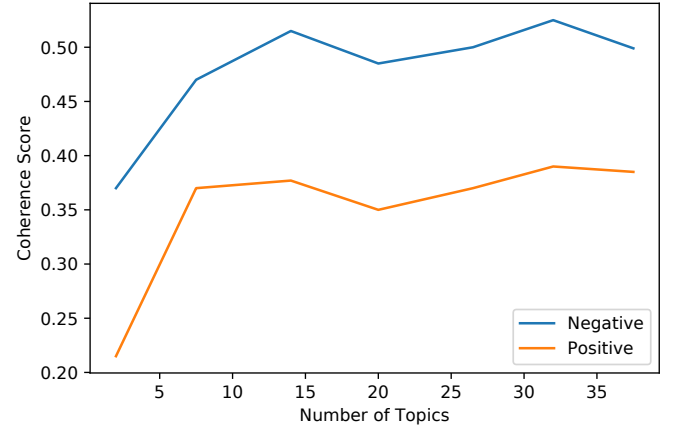


Figure 1: Coherence score  $C_v$  for topic models of negative and positive case studies using different topic counts.

## 3 METHODS

We used computer-assisted content analysis methods to explore the level of transparency of current behavioral information published on the two previously mentioned SCS websites. First, the column-and-row structured records of both the Blacklist and the Redlist on the SCS’s Beijing platform<sup>5</sup> were crawled and statistically evaluated. Hence, to understand the semantic and structural patterns of both “positive” and “negative” case studies, we crawled news reports on “bad” behavior labeled as “Typical Cases (典型案例)”<sup>6</sup> and on “good” behavior labeled as “Stories of Integrity (诚信人物/故事)” under the section of “Integrity Culture (诚信文化)”<sup>7</sup> on the national SCS information platform “Credit China”<sup>8</sup>. We then applied statistical topic modeling based on Latent Dirichlet Allocation (LDA) to all available 156 news reports on “good” behavior (“positive” cases) and 789 news reports on “bad” behavior (“negative” cases) on August 12, 2018.

We preprocessed the downloaded documents by applying *jieba*<sup>9</sup> for segmentation and stopword filtering of Chinese text. We used the stopword corpus compiled by the Chinese search engine Baidu<sup>10</sup>. After tokenization of the given text, we applied *tf-idf* to re-weight term counts. As we had no reasonable expectation for the number of topics  $k$  to be detected within the given document corpus, we performed optimal topic number search. Thereby, we created several LDA models for “positive” and “negative” case studies and calculated the topic coherence measures  $C_v$  as proposed in [25]. We started with  $k = 2$  and increased the number of topics until an upper bound of  $k = 40$ . As shown in Figure 1, coherence values of models for both document sets increased until  $k = 15$  before flattening out. Therefore, we investigated the top-30 most salient terms for each of the fifteen topics produced by these models [4]. Thereby, we set  $\delta = 0.6$  within the applied relevance metric [28].

<sup>5</sup><http://www.creditbj.gov.cn/xyData/front/creditService/initial.shtml%20?typeId=4>.

<sup>6</sup><https://www.creditchina.gov.cn/home/dianxinganli1/?navPage=6>.

<sup>7</sup><https://www.creditchina.gov.cn/chengxinwenhua/chengxingushi/>.

<sup>8</sup><https://www.creditchina.gov.cn/>.

<sup>9</sup><https://github.com/fxsjy/jieba>.

<sup>10</sup><http://www.baidu.com/baidu-stopwords>.



**Figure 2: Three lists publishing records of “negative” behavior: from left to right, the first arrow points to Blacklist, the second arrow to Special Attention List, and the third arrow to Administrative Punishment.**

Moreover, we also reviewed the results for  $k = 10$ ,  $k = 20$ , and  $k = 30$  in order to further manually verify the optimal topic number. We found the optimal model with  $k = 10$  for both “positive” and “negative” cases. Finally, we further selected 5 main topics for the “positive” cases and 7 topics for the “negative” cases (see Table 3 in the Supplementary Materials for topics selected for the “positive” cases, and Table 4 in the Supplementary Materials for “negative” case topics).

Based on our topic modeling results, we selected the 4 most related cases (highest predicted probability of belonging to the topic) for each of the topics.<sup>11</sup> We then manually analyzed 20 “positive” cases and 26 “negative” cases<sup>12</sup> in detail. One author first reviewed 5 “positive” and 5 “negative” cases, respectively, and drafted a coding guide, which was then reviewed iteratively by another author, refined, and retested to generate consistent definitions. As a result, we developed two coding schemes for “positive” and “negative” cases (see Table 1 for the coding scheme applied to “positive” cases and see Table 2 for the coding scheme used to analyze “negative” cases). After reliability was established, we examined all 46 cases for structural and thematic commonalities. Each coding sheet contained the information from one “positive” or “negative” case. Once the coding sheets were completed, we grouped and analyzed the information contained in them.

## 4 RESULTS

### 4.1 Blacklists

On the platform of “Credit China (Beijing)”, we found three publicly accessible databases providing information on “bad” behavior, all of which could be queried by search term. Translated from Chinese (see Figure 2), they were termed the following: 1) Blacklist (1,137,546 entries), 2) Special Attention List (9,229,179 entries), and 3) Administrative Punishment (14,885,789 entries).

The Blacklist further contained 16 subcategories for “bad” behavior. For the Blacklist, we crawled two of these subcategories, one containing records of individuals that have been banned from

<sup>11</sup>For “negative” cases, there are only three cases for Topic 6 (measures taken against crime) and Topic 7 (public transport regulation violation), respectively.

<sup>12</sup>There were only 3 cases for 2 out of the 7 topics.

Pattern	Definition	Example
Bio-info	full name	今年70岁的刘某某，为了一句诺言，一辈子踏踏实实做一名“小村大医生”。
	age	古亭村77岁的老人蓝某某为了归还欠银行的一笔500元死账。
	living place	蓝某某出生在遂昌县云峰街道古亭村。
	profession	这位一天两次捡到钱包的“好运人”就是蒙阴一中的英语教师耿某某。
Social class	low	一个清贫的普通农家，父亲、儿子、孙女毫无怨言地赡养一位无任何血缘关系的“外人”。
	middle	陈某某的妻子说，他们家里也就是普通家庭，上有老下有小。
	high	这句话时常在内蒙古明泽集团董事长王某某的心里翻腾着。
Sacrifice for the common interest	material sacrifice	他隔天检查药柜，受潮的药直接销毁，损失的药费自己承担。
	non-material sacrifice	每天为他做三餐，每天打针吃药，就连端屎端尿的活也揽下来。
Rewards	reputational rewards	他被评为全国农村青年创业致富带头人、北京市优秀农村实用人才。
	material reward refusal	钱包的主人一个劲地要给她塞钱。肖某某坚决地拒绝了。
Virtue cascade	trustworthy and honest	为了不让养殖户遭受损失，彭某某把风险留给自己，仍按照回收合同原价收回了养殖户的肉鸭。
	hardworking	虽然有时一天连饭都顾不上吃，还帮助菜农一起装菜卸菜，忙到了深夜还要了解市场信息、掌握蔬菜的价格趋向。
	self-discipline	虽然银行减免并注销了这笔贷款，但放在我私人账户的钱一定要还上。
	helpful	积极参加协会组织的慰问残疾人、资助贫困大学生活动。
	care-taking	他们一家三代几十年如一日地照顾着丁某某老人。
	sense of responsibility	她以当好水资源质量的守望者为己任。

**Table 1: Coding scheme for “positive” cases. All “positive” cases included biographical information of the individual and indicated his or her social class. Other codes described the individual’s sacrifice for the common interest, the rewards obtained, and the further attribution of other virtues (virtue cascade).**

数据来源:	证监会
数据类别:	市场禁入
主体名称:	吴 [ ]
加配证件号码:	110104*****4X
证件类型:	身份证
个人代码:	
处罚处理名称:	证券法律[2008]12号市场禁入决定书(案:[ ]、[ ]、孙 [ ])定
处罚处理日期:	2008/11/27
处罚处理种类:	市场禁入(5年)
处罚对象类型(1组织机构, 2个人)	
真正证件号码:	
信息类型:	
处罚机关:	中国证监会
处罚决定书id:	
处罚处理内容:	<p>证监法律字[2008]12号市场禁入决定书当事人：梁 [ ] 女，1946年出生。北京中兴信託投资有限公司（以下简称中兴信託）法定代表人、億富投资有限公司（以下简称億富投资）法定代表人。住址北京市东城区前门东大街1号2单元207号。吴 [ ] 女，1949年出生。中兴信託北京运营总部经理、億富投资管理，住址北京市朝阳区东四环东路106甲3号楼1804号。孙 [ ] 男，1942年出生。億富投资管理，住址北京市东城区前门东大街1号2单元207号。亿恒康《中华人民共和国证券法》（以下简称《证券法》）的有关规定，本会对上述北京中兴信託运营总部等机构违反证券法律法规进行了立案调查。审理，并依法向当事人告知了实施市场禁入措施的事实、理由及依据及当事人依法享有的权利、陈述意见、申辩资格。经查明，自2000年4月20日起，億富投资在北京中兴信託北京营业部开设并控制“億富公司”、“华隆发展”、“华发机械”、“广发广告”、“利通通讯”以及“恒康”、“华隆”、“利通”、“恒康”等证券账户，让上述人员从原账户中取出资金，买卖股票。上述事实，有相关开户资料、客户对账单、资金划转凭证、情况说明等证据在案证实。证据确实、充分，足以认定。億富投资的上述行为违反了原《证券法》第七十四条“在证券交易所，严禁以个人名义开立账户、买卖证券的规定”，构成了原《证券法》第一百九十一条所述违反本法规定，法人以个人名义设立账户买卖证券”的行为。根据当事人违法行为的客观、性质、情节和社会危害程度，依照《证券市场禁入暂行规定》第二条第七款等相关规定，本会决定立即采取适当人等，自本会作出该处罚之日起三年内不得从事上市公司管理人和从事证券投资活动；[ ]为市场禁入者。</p>

Reasons for being on the Blacklist	Number of records
Obstructing or resisting enforcement.	669
Fails to fulfill debt obligation without any justifiable reason.	2651
Other circumstances for evasion of enforcement.	6486
Violating the property reporting system.	19062
Failing to fulfill debt obligation.	98634

**Figure 4: The top 5 reasons for being on the Blacklist of company debtors.**

at the time of crawling (corresponding field not shown). Entries included a reference to legal regulation and specified the full name of the company (see Figure 5). Note that some companies listed had multiple entries corresponding to multiple breaches. Together with these explanations, we crawled the date of publication on the Blacklist for each entry. We found that on one day in June 2018, 95.6% of all entries (125,747) had been published on the Blacklist for companies (see Figure 6). This probably indicates that these records had already been collected and processed by another entity before being transferred to and published on the Blacklist.

For the Special Attention List, we collected 30,625 entries containing information on companies that had violated business operation regulations. For all records collected, companies had been blacklisted for providing various types of false information to the authorities (see Figure 7).

Finally, our crawler returned 32,719 entries for the Administrative Punishment register that contained information on both



数据来源:	司法
数据类别:	失信被执行人名单
案号:	(2018) 鲁1092执47号
主体名称:	威海久置建材有限公司
企业法人姓名:	
组织机构代码:	558932122
执行法院:	威海经济技术开发区人民法院
地域名称:	山东
执行依据文号:	(2016) 鲁1092民初1085号
作出执行依据单位:	威海经济技术开发区人民法院
法律文书生效书确定的义务:	被执行人给付申请人款项共计3023657.49元
被执行人的履行情况:	全部未履行
失信被执行人具体情况:	有履行能力而拒不履行生效法律文书确定义务
发布日期:	20180525
立案时间:	20180118
已履行部分:	暂无
未履行部分:	暂无
最新更新日期:	2018-06-11

Figure 5: Screenshot of a company's Blacklist entry. Left column, the first arrow points to a field explaining the specific context of the case, the second arrow points to the date of publication of this entry on the Blacklist.

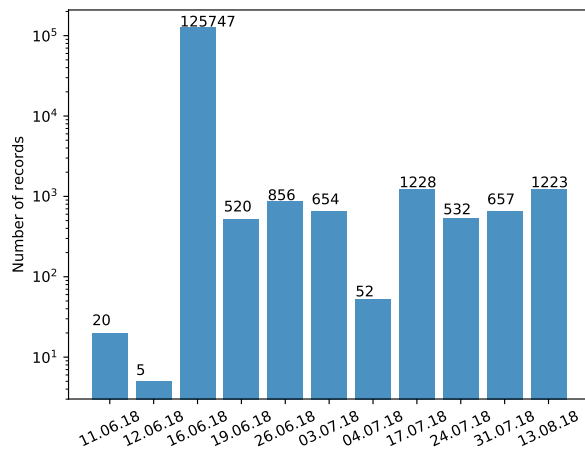


Figure 6: Publication dates of Blacklist entries for company debtors.

individuals and companies (see Figure 8). As Figure 9 shows, the majority of records of the Administrative Punishment register reported traffic rule violations.

Correspondingly, fines were the most widely used measure (see Figure 10). We also found that only company entries of the Administrative Punishment register and the Blacklist consistently featured the Unified Social Credit Code.

On the national SCS information platform “China Credit”, we found another Blacklist issued by the Civil Aviation Administration of China (中国民用航空局)<sup>13</sup>. This list, which is updated every month, publishes information on individuals that are excluded from aircraft travel for a period of one year due to misbehavior on airplanes or airports (data collected on August 10, 2018; see Figure 11). According to the list published in August, 2018, 946 individuals were banned from air travel for one year. Among others, the list provided full name, censored ID number, and explanations why individuals had been punished (see three arrows in the first row of Figure 11). Being banned from air travel resulted from taking illegal objects on airplanes, smoking on airplanes, or boarding airplanes

<sup>13</sup><https://hmd.creditchina.gov.cn/>, last accessed on November 5, 2018.

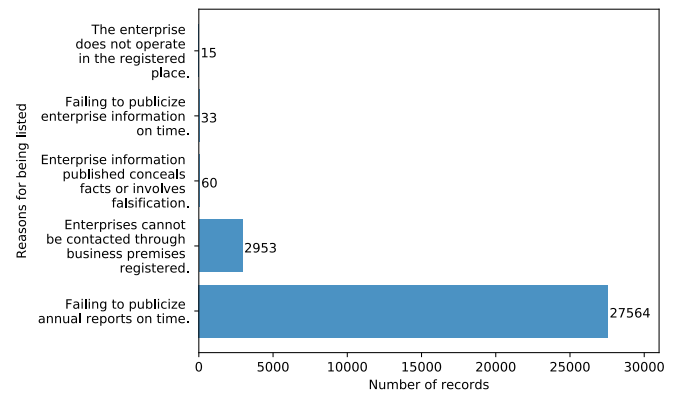


Figure 7: The top 5 reasons for companies to be on the Special Attention List.

主体名称:	赵某
统一社会信用代码:	
处罚类别1:	罚款
处罚事由:	过度疲劳仍继续驾驶的
处罚依据:	暂未入库
处罚名称:	过度疲劳仍继续驾驶的
处罚类别2:	
组织机构代码:	
工商登记码:	
税务登记号:	
法定代表人名称:	
处罚结果:	交通警察总队武清支队对赵文选进行罚款的处罚
处罚期限:	
处罚机构:	交通警察总队武清支队
处罚部门:	
处罚决定日期:	2017/01/03
当前状态:	正常
严重程度:	
地方编码:	120000
创建时间:	2018-07-11 00:00:00

Figure 8: A record of the Administrative Punishment register. The first column, from top to down: the first arrow points to the field "type of punishment" and the second points to the field "reasons for punishment".

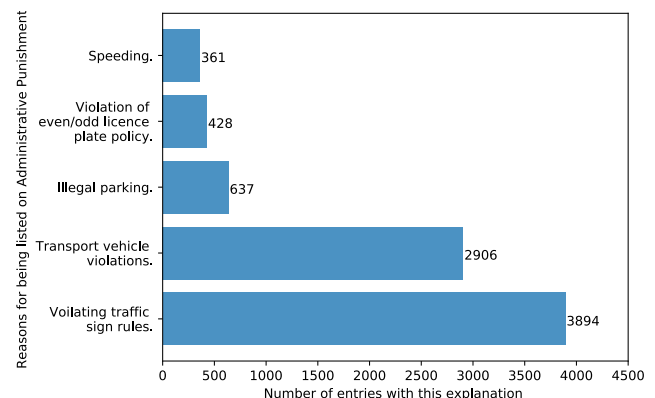


Figure 9: The top 5 reasons why individuals or companies are placed on the Administrative Punishment register.

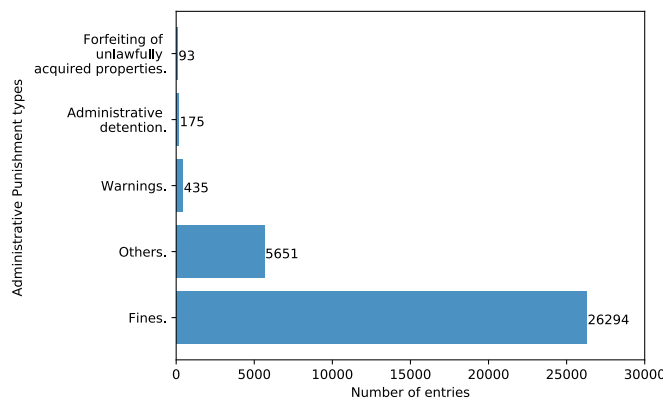


Figure 10: The 5 types of Administrative Punishments.

with a fake passport. The figure also indicates that the list contained names and ID numbers of non-Chinese citizens providing some evidence that foreigners were not excluded from the SCS.

## 4.2 Redlist

We found one type of list documenting information on "good" behavior - the Redlist. It contained a total of 1,206,944 entries distributed across 24 categories (3 categories for redlisted individuals, 21 categories for redlisted companies). The categories for individuals, translated from Chinese, are: 1) Taxi Star (1557 entries), 2) Top Ten Tour Guides (14 entries), and 3) Five-Star Volunteer (603 entries). For all entries, the full name of the person and his or her partially censored ID number were given. The Five-Star Volunteer category displayed the gender of the person as well as the amount of volunteering hours carried out per person. The lowest amount of volunteering hours documented was 1500 (which was probably the necessary threshold to be listed) and the highest was 25,400. None of the entries we collected from the Redlist provided an explanation justifying why such a honorary title had been awarded to that person (see Figure 13). Thus, we cannot report any observations about justifications on "good" behavior from our Beijing Redlist analysis.

Company categories referred either to tax awards (e.g., A Class Taxpayer) or to other honorable statuses such as Harmonious Labor Relations or Excellent Contributor to Developing Chinese Socialism. Just like the Redlist entries of individuals, there were no justifications explaining why a honorable title had been awarded to a company. No Redlist entry contained the Unified Social Credit Code. Generally, Figure 13 shows a single record of an entity that can display several "positive" and "negative" entries. Thus, there is reason to believe that the interface shown in Figure 13 functions as the governmental SCS information template: recording and making transparent information on rewards and/or sanctions to the public.

Importantly, every Blacklist and Redlist record we collected featured a "Disagreement/Correction (异议/纠错)" function (see Figure 12). This function allowed citizens to object to a Blacklist or Redlist decision by providing a statement of up to 2000 Chinese characters (submission required 18-digit ID number).

## 4.3 Coding results for "positive" cases on "good" behavior

News reports on "good" behavior were introduced as "Stories of Integrity (诚信人物/故事)" posted under the section of "Integrity Culture (诚信文化)" on the national SCS information platform "Credit China". All of the 20 "positive" cases selected described how a protagonist sacrificed his or her self-interest (both material and non-material) for the common good. Moreover, all cases centered on "trustworthiness" and "honesty" as key SCS virtues. The stories all followed the same narrative structure: they first provided detailed biographical information of a person (full name, social class, profession, family status), followed by a dilemma: the protagonist could either engage in "dishonest" behavior winning him or her an immediate small reward or get a large future reward by being "honest". Once the person had enacted the "honest" behavior, which happened in all the "positive" reports we analyzed, the narratives ended with a virtue cascade.

Take, for example, cases in which individuals found and returned lost property to an owner. Here, all four cases assigned to the topic "return lost property to owner" ended by further attributing "self-discipline", "helpfulness", "care-taking for others", and a "sense of responsibility" to the protagonist as part of a virtue cascade. Another commonality across the selected cases was that all protagonists were morally "praised" by their social environment. Also, the protagonist was recognized for his or her "good" behavior by official agencies or the media in the form of "honors", "decorations", or a "cute nickname". On the other hand, when a material reward was offered for the "good" behavior, as in all cases with topics "family and community relationship and repayment", "return lost property to owner", and "social entrepreneurship to help people out of poverty", the protagonist refused the material reward at all times.

## 4.4 Coding results for "negative" cases on "bad" behavior

Reports about "bad" behavior were labeled as "Typical Cases (典型案例)" on the homepage of "Credit China" with the sources being both local newspapers and the platform itself. The 26 selected "negative" cases relating to 7 topics all revolved around one common theme, the "Laolai (老赖)": a term specifically referring to individuals and companies refusing to repay debts. These cases were presented in two ways. The 4 cases with the topic "public shaming" were about the courts' actions in solving repayment problems. The remainder of the stories were about specific individuals or companies. All individuals and companies were anonymous in the selected cases. Local courts collaborated with local telecommunication companies in all 4 cases with the topic "public shaming", and the Public Security Bureau played an important enforcement role in all cases with topic "public transport regulation violation". In these reports, both the compulsory actions taken by the court and the threat of being placed on the Blacklist forced the "Laolai" to fulfill the stated obligation. Generally, both "positive" and "negative" case studies we analyzed were homogeneous in structure, framing, and content. This could indicate that they had been deliberately formulated to propagate the SCS's conceptualization of "good" and "bad" behavior.

256	刘	420281****0919691X	随身携带或托运国家法律、法规规定的危险品、违禁品和管制物品的;在随身携带或托运行李中故意藏匿国家规定以外属于民航禁止、限制运输物品的	湖北省公安厅机场公安局直属分局航站区派出所 行政处罚决定书	鄂机公直航【行罚决】字(2018)第49号
257	姜	370103****12182531	随身携带或托运国家法律、法规规定的危险品、违禁品和管制物品的;在随身携带或托运行李中故意藏匿国家规定以外属于民航禁止、限制运输物品的	北京首都国际机场公安分局公安行政处罚决定书	京机公分(法)决字[2018]第296号
258	李	340121****0918915X	随身携带或托运国家法律、法规规定的危险品、违禁品和管制物品的;在随身携带或托运行李中故意藏匿国家规定以外属于民航禁止、限制运输物品的	安徽省公安厅机场公安局合肥机场派出所 行政处罚决定书	直公(机)行罚决字[2018]64号
259	赵	532331****06090635	随身携带或托运国家法律、法规规定的危险品、违禁品和管制物品的;在随身携带或托运行李中故意藏匿国家规定以外属于民航禁止、限制运输物品的	云南省公安厅民用机场公安局直属公安分局 当场处罚决定书	0548462018050702
260	SAURAM	AA7242****	随身携带或托运国家法律、法规规定的危险品、违禁品和管制物品的;在随身携带或托运行李中故意藏匿国家规定以外属于民航禁止、限制运输物品的	深圳市公安局机场分局 行政处罚决定书	深公(机)行罚决字【2018】00280号
261	SHOTAYRNVUR	N11668****	随身携带或托运国家法律、法规规定的危险品、违禁品和管制物品的;在随身携带或托运行李中故意藏匿国家规定以外属于民航禁止、限制运输物品的	民航新疆机场公安局 行政处罚决定书	民航新机公(候派)行罚决字[2018]009号

Figure 11: A screenshot of the Blacklist for individuals that are banned from flying on commercial airplanes. In the first row, from left to right, the first arrow points a field containing the full name of the individual; the second to censored ID number; and the third to explanations why individuals have been punished. Two arrows at the bottom left indicate entries of two foreign passengers.

重庆百乐维克动物药业有限公司

统一社会信用代码: 91500224768892595K

查看时间: 2018-11-20 04:29:21

该企业黑名单记录 10条

生成报告

风险提示: 本网站仅基于已掌握的信息提供查询服务, 查询结果不代表本网站对被查询对象信用状况的评价, 仅供参考, 请注意识别和防范信用

基础信息	行政许可(5)	行政处罚(0)	红名单(0)	重点关注名单(0)	黑名单(10)	其他(0)
数据来源:	高法					
数据类别:	失信被执行人					
案号:	(2018)渝0151执1094号					
主体名称:	重庆百乐维克动物药业有限公司					
企业法人姓名:						
组织机构代码:	91500224768892595K					
执行法院:	重庆市铜梁区人民法院					
地域名称:	重庆					
执行依据文号:	渝铜劳人仲案字【2017】第442号					
作出执行依据单位:	重庆市铜梁区劳动人事争议仲裁委员会					
法律生效文书确定的义务:	被申请人支付申请人工资11961元					
被执行人的履行情况:	全部未履行					
失信被执行人具体情形:	有履行能力而拒不履行生效法律文书确定义务					
发布时间:	20180719					
立案时间:	20180404					
已履行部分:	暂无					
未履行部分:	暂无					
最新更新日期:	2018-07-24					

Figure 12: Example of a company’s Blacklist entry. The black circle on the upright corner indicates the “Disagreement/Correction (异议/纠错)” function.

5 ANALYSIS

The results of our content analysis demonstrate that there are currently multiple informational asymmetries in both datasets.

5.1 Listed companies versus listed individuals

Currently, companies make up the majority of entries on both the Blacklist and Redlist of Beijing’s SCS platform. We found that companies which are involved in the construction of the SCS were also included in the list. For instance, Alibaba (with Zhima Credit) and Tencent (with Tencent Credit) were both granted permission to start individual pilot credit service programs in 2015 and have provided digital data collected from online shopping and social media to the SCS. Both Alibaba and Tencent were listed as A-level

Taxpayers on the Redlist. Since we only crawled the Beijing SCS platform, we cannot make any claims about the transparency of other SCS Blacklist and Redlist websites.

Our analysis of “positive” and “negative” cases demonstrates the opposite: here, the majority of reports on either “good” or “bad” behavior focuses on individuals’ behaviors. For our manually coded sample, only 15.4% of “negative” reports and 30.0% of “positive” reports featured companies. In both “negative” and “positive” cases that featured companies, however, reports centered on the person in charge of the company typically highlighting the CEO’s virtues and vices. In other words, it is not the company as such that is “blamed” or “praised,” but rather the person responsible for the company. Such portraits, therefore, signal that individuals are not shielded by large institutions but can be made responsible for their “good” or “bad” decision-making.

5.2 Justifying punishments versus justifying rewards

All entries of the Blacklist explain why a person or company is currently registered on the Blacklist. Moreover, Blacklist explanations include legal terms and refer to laws and regulations. In other words, Blacklist explanations make transparent the mechanism of punishment by specifying a causal link between behavior and consequence. This is perhaps best illustrated by the Blacklist on individuals excluded from air travel (see Figure 11). The legal threat contained in the entries of the Blacklist could furthermore signal that a specific “dishonest” behavior can be detected and sanctioned.

On the other hand, not a single entry of the Redlist includes a formulated explanation on why a person or company has been awarded a honorary title. We found that fulfilling legal obligations (Class A Taxpayer), performing professional (Taxi Star) or volunteering (Five-Star Volunteer) duties can result in reputational gain in the current SCS. However, the mechanisms or criteria determining when an individual or a company secures a place on the Redlist are not further explained. Taken together, the current SCS makes behaviors leading to punishments more transparent than behaviors



基础信息	行政许可(0)	行政处罚(0)	红名单(1)	重点关注名单(0)	黑名单(0)	其他(0)
数据来源：	市团委					
数据类别：	五星志愿者					
主体名称：	李[REDACTED]					
身份证号：	110229*****8001					
志愿者编号：	110229100087339					
性别：	男					
服务时间(小时)：	6587.0					

**Figure 13: Example of a Redlist entry for an individual with the honorary title Five-Star Volunteer. The record does not justify why the honorary title was awarded.**

resulting in rewards. More generally, our study could not identify publicly available information associating specific behaviors to a scoring or rating mechanism.

### 5.3 Types of punishments versus types of incentives

The most common reason for a company to be placed on any of the “negative” lists is failure to pay back debt (the second most common reason is informational misconduct). Failure to pay back debt is also the most prominent reason given for why protagonists of the “negative” cases are registered on the Blacklist. The Chinese term for “Laolai” appeared 481 times in the 789 “negative” reports we collected. All “negative” stories we manually coded report on the activities of a “Laolai” person (either as an individual or as the legal representative of a company). In terms of punishment, individuals and companies face both the material loss specified in the corresponding legal regulation as well as the consequences of being publicly shamed on the Blacklist. In more than 40% of the narratives on “negative” behavior, an individual is threatened to be placed on the Blacklist leading to the immediate compliance of the individual.

On the other hand, individuals and companies on the Redlist receive moral “approval” and reputational gain. Similarly, “positive” cases report on individuals that gain reputational rewards, while at the same time rejecting material incentives when offered as a consequence of their “role-model” behavior. Still, being listed on the Redlist is not mentioned or even indicated by any individuals as a motivational factor for their behaviors. All stories we analyzed emphasize that a morally “praiseworthy” activity is “praiseworthy” when it is “genuinely” moral rather than instrumental in obtaining a material reward. Furthermore, all “positive” stories feature a virtue cascade: once an individual is described as “genuinely honest” or “trustworthy”, he or she is attributed other “positive” virtues as a consequence.

## 6 DISCUSSION & CONCLUDING REMARKS

In this first study of key websites of the Chinese SCS, our goal was to understand how transparent the SCS currently is in providing information on “good” and “bad” behavior. To this end, we collected

and analyzed 194,829 Blacklist and Redlist entries from the Beijing SCS website “beijing.gov.cn/creditbj” and applied a machine learning topic modeling algorithm to almost 1000 reports on “positive” and “negative” behavior crawled from the national SCS information platform “www.creditchina.gov.cn”. Finally, we manually coded a sample of these texts to understand what kind of specific behavioral information they contain.

The main question arising from our findings, we believe, is whether the degree of the current SCS’s transparency is intentionally engineered or whether it is simply a manifestation of work in progress. Is there a purpose in explicitly describing and publishing the causal link between behavior and sanction while leaving information on getting rewards deliberately vague? First, the asymmetries in information provided between the Redlist and the Blacklist could be motivated economically: while an infinite amount of people can be excluded from valuable material resources, only a finite amount can be given valuable resources (e.g., a first-class train ticket). Detailed instructions on how to win rewards could therefore lead to distribution problems since many individuals could implement them. On the other hand, another explanation for the current informational asymmetries of the SCS might be that already existing records of legal offenses were used to start filling Blacklists. Consequently, these records entail more justifications since they refer to specific legal articles or regulations.

The degree of transparency of the SCS observed in this work could also be motivated by behavioral engineering goals. Let’s imagine for the moment the system were completely inscrutable (i.e., the system did not justify a score increase or decrease and eventually a given punishment or reward, respectively). In this case, individuals would have little possibility to understand when the SCS rewarded and when it sanctioned specific types of behaviors. Moreover, besides being oblivious to the moral code of conduct, individuals would not have the ability to contest the system’s decision-making process (again, to negotiate a norm one must have the necessary epistemic resources to do so). Note that this issue is also debated in the context of the “Right to Explanation” of the European Union’s General Data Protection Regulation [27, 31]. A fully transparent scoring system, on the other hand, would precisely map behaviors to rewards or sanctions. Indeed, in the context of a nationwide

digitally-implemented scoring system, full transparency must account for the mechanism that leads to the distribution of rewards or sanctions. This degree of transparency would offer individuals the possibility to understand the system's decision-making procedures at least to a certain extent. In our analysis of SCS Blacklist and Redlist records, we did not identify an explicit SCS scoring mechanism. We have shown, however, that the SCS already enables citizens to dispute single Blacklist and Redlist records. On the other hand, a fully transparent SCS would possibly create other problems: if the SCS became fully transparent in regard to its scoring mechanisms, complying to a norm would likely become a market transaction. In fact, research on intrinsic and extrinsic motivation suggests that introducing an external reward to a norm-guided behavior turns this behavior into a commodity that can be bought [10, 11]. This phenomenon, termed "crowding-out effect", results in fewer people engaging in this behavior since the consequences of failing to act can simply be compensated by financial means [2, 9, 26]. For example, if one reliably receives monetary compensation for being honest, being honest will no longer be evaluated as a moral behavior for both the actor and the recipient. As this line of research suggests, individuals will likely stop attributing a genuine moral character to individuals with a high score in a fully transparent SCS.

Our analysis provides evidence that the currently implemented SCS possibly attempts to counter such a transformation of moral behavior into market transactions. All of the "positive" case studies unambiguously emphasize that norm conformity is "good" because it is "morally valuable" for both average citizens as well as CEOs. None of the Redlist entries describe a connection between moral behavior and external material reward. Rather, they contain virtue signals and reputational gains by awarding symbolic honorary titles (e.g., Five-Star Volunteer). On another sub-page of the national SCS website, we found the publication of 32 ancient Chinese fables (not shown) also promoting self-concepts comprising virtues of being a morally "good" Chinese citizen. In contrast, our analysis on the corpus of "negative" case studies demonstrates the propagation of a "negative" self-concept ("Laolai") attributable to a specific offense (i.e., intentionally not paying back debt). Taken together, our analysis suggests that degrees of transparency can serve different behavioral engineering goals in the context of a digital scoring system.

## REFERENCES

- [1] Alessandro Acquisti and Jens Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE Security & Privacy* 3, 1 (2005), 26–33.
- [2] Roland Bénabou and Jean Tirole. 2006. Incentives and prosocial behavior. *American Economic Review* 96, 5 (2006), 1652–1678.
- [3] Yongxi Chen and Anne Cheung. 2017. The transparent self under big data profiling: Privacy and Chinese legislation on the Social Credit System. *The Journal of Comparative Law* 12, 2 (2017), 356–378.
- [4] Jason Chuang, Christopher Manning, and Jeffrey Heer. 2012. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 74–77.
- [5] Asli Demircug-Kunt, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess. 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. The World Bank.
- [6] Charles Ess. 2006. Ethical pluralism and global information ethics. *Ethics and Information Technology* 8, 4 (2006), 215–226.
- [7] Luciano Floridi. 2013. *The Ethics of Information*. Oxford University Press.
- [8] Bernard Gert and Joshua Gert. 2017. The definition of morality. In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.
- [9] Uri Gneezy, Stephan Meier, and Pedro Rey-Biel. 2011. When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives* 25, 4 (2011), 191–210.
- [10] Uri Gneezy and Aldo Rustichini. 2000. A fine is a price. *The Journal of Legal Studies* 29, 1 (2000), 1–17.
- [11] Uri Gneezy and Aldo Rustichini. 2000. Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115, 3 (2000), 791–810.
- [12] Chris Hoofnagle. 2003. Big Brother's little helpers: How ChoicePoint and other commercial data brokers collect and package your data for law enforcement. *North Carolina Journal of International Law and Commercial Regulation* 29 (2003), 595–637.
- [13] Yanzhong Huang. 2012. Why is not there a bottom line for food security issue in China (Zhongguo Shipin Anquan Weihe Meiyou Dixian?) (in Chinese). <https://cn.nytimes.com/opinion/20120821/c21huang/>
- [14] Ipsos Public Affairs. 2017. What worries the world? [https://www.ipsos.com/sites/default/files/2017-08/What\\_worries\\_the\\_world-July-2017.pdf](https://www.ipsos.com/sites/default/files/2017-08/What_worries_the_world-July-2017.pdf).
- [15] Nicola Jentzsch. 2006. *The Economics and Regulation of Financial Privacy: An International Comparison of Credit Reporting Systems*. Springer Science & Business Media.
- [16] Genia Kostka. 2018. *China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval*. Technical Report. Free University of Berlin. Available on SSRN: <https://ssrn.com/abstract=3215138>.
- [17] Jianzhou Liu. 2011. Building the Social Credit System: The content, the model, and the trajectory (Shehui Xinyong Tixi Jianshe: Neihan, Moshi yu Lujing Xuanze; in Chinese). *Journal of the Party School of the Central Committee of the C.P.C.* 15, 3 (2011), 50–53.
- [18] Yao-Huai Lü. 2005. Privacy and data privacy issues in contemporary China. *Ethics and Information Technology* 7, 1 (2005), 7–15.
- [19] Andrzej Marczewski. 2017. The Ethics of Gamification. *XRDS* 24, 1 (2017), 56–59.
- [20] Mirjam Meissner and Jost Wübbeke. 2016. IT-backed authoritarianism: Information technology enhances central authority and control capacity under Xi Jinping. *China's Core Executive: Leadership Styles, Structures and Processes under Xi Jinping* (2016), 52–57.
- [21] Simina Mistreanu. 2018. Life inside China's Social Credit laboratory: The party's massive experiment in ranking and monitoring Chinese citizens has already started. <https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/>.
- [22] Nature. 2018. China sets a strong example on how to address scientific fraud. <https://www.nature.com/articles/d41586-018-05417-1>
- [23] Mareike Ohlberg, Shazeda Ahmed, and Bertram Lang. 2018. Central Planning, Local Experiments: The Complex Implementation of China's Social Credit System. <https://www.merics.org/en/microsite/china-monitor/central-planning-local-experiments>.
- [24] Zahy Ramadan. 2018. The gamification of trust: The case of China's "Social Credit". *Marketing Intelligence & Planning* 36, 1 (2018), 93–107.
- [25] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 399–408.
- [26] Richard Ryan and Edward Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25, 1 (2000), 54–67.
- [27] Andrew Selbst and Julia Powles. 2017. Meaningful information and the right to explanation. *International Data Privacy Law* 7, 4 (2017), 233–242.
- [28] Carson Sievert and Kenneth Shirley. 2014. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. 63–70.
- [29] State Council. 2014. Notice of the State Council on Issuing the Outline of the Plan for Building a Social Credit System (2014-2020); (in Chinese). [http://www.gov.cn/zhengce/content/2014-06/27/content\\_8913.htm](http://www.gov.cn/zhengce/content/2014-06/27/content_8913.htm).
- [30] Matteo Turilli and Luciano Floridi. 2009. The ethics of information transparency. *Ethics of Information Technology* (2009), 105–112.
- [31] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi. 2017. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law* 7, 2 (2017), 76–99.
- [32] Pak-Hang Wong. 2012. Dao, harmony and personhood: Towards a Confucian ethics of technology. *Philosophy & Technology* 25, 1 (2012), 67–86.
- [33] Yanfang Wu, Tuenyu Lau, David Atkin, and Carolyn Lin. 2011. A comparative study of online privacy regulations in the US and China. *Telecommunications Policy* 35, 7 (2011), 603–616.
- [34] Jinghong Xu. 2015. Evolving legal frameworks for protecting the right to Internet privacy in China. *China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain* (2015), 242–259.
- [35] Zhiling Zhao and Feng Ding. 2007. Shanghai, Zhejiang, Shenzhen Social Credit System models, problems and revelation (Shanghai, Zhejiang, Shenzhen Shehui Xinyong Tixi Jianshe Moshi, Wenti yu Qishi) (in Chinese). *Wei Shi* 10 (2007), 70–73.
- [36] Jarrett Zigon. 2008. *Morality: An Anthropological Perspective*. Berg.