Give Us a Little Social Credit: To Design or to Discover Personal Ratings in the Era of Big Data

Abigail Devereaux  
George Mason University, Economics  
abigail.devereaux@gmail.com

Linan Peng  
George Mason University, Economics  
lpeng2@gmu.edu

Abstract

In 2014, China announced the institution of a social credit system by 2020. Social credit ratings of the type being developed by China go beyond existing financial credit ratings in an attempt to project less-tangible personal characteristics like trustworthiness, criminal tendencies, and group loyalty onto a single scale. The advent of Big Data—characterized by a large and increasing volume of personal data and tools like machine learning to detect patterns and generate predictions based on that data—strongly indicates that various kinds of social credit ratings will become a reality in the near future. Supposing that the emergence of Big Data-enabled personal ratings is both a cost-saving adaptation to and general improvement upon traditional forms of signaling trustworthiness, we use both traditional modeling techniques and evidence-based argument to determine whether “optimal” social credit should develop publicly, privately, or in a polycentric fashion.

Keywords: social credit, personal ratings, big data, trustworthiness, public goods, institutions

JEL Codes: D02, E02, O33, O35, P21, P50
1. Introduction

In 2014, as part of its Twelfth Five-Year Plan, the Chinese Communist Party (CCP) issued a document titled, “Planning Outline for the Construction of a Social Credit System.” “The social credit system is an important part of the socialist market economic system and social governance system,” the document begins, tying the fate of Chinese communism-with-capitalistic-elements to the social credit system right out of the gate. The CCP’s stated motivation for developing a publicly designed and administered social credit system is to “raise the awareness of integrity and the level of trustworthiness of Chinese society.”

Social credit ratings are examples of social institutions meant to help coordinate the ends of individuals in the system. Institutions can be constructed from the top-down, can emerge from no conscious plan on the part of those whose actions help form them, or can have both emergent and constructed qualities (Wagner 2012, 2016). China’s SCS is intended to realize centralized personal ratings on a single scale of social desirability and undesirability, where CCP officials define which traits are socially desirable or undesirable and what rewards and punishments accrue to individuals based on their rating (CCP 2014). Low enough social credit scores earn the bearer a place on a “blacklist,” a public database of low-score individuals maintained by the Supreme People’s Court of China (SPC 2018). Several provinces have devised ways to publicly shame blacklisted individuals, including pictures and personal information on giant screens in parks and squares with voice-overs proclaiming: “the person you are calling is a dishonest debtor” (Zeng 2018).

1“社会信用体系是社会主义市场经济体制和社会治理体制的重要组成部分.” (CCP 2014)
2 “目的是提高全社会的诚信意识和信用水平.” (CCP 2014)
China’s SCS would not exist but for Big Data. Most people have a visceral sense for what is meant by Big Data: for every individual every day, individuals, private companies and federal entities collect a huge volume of personal data in various databases. Examples of data types include GPS coordinates, satellite imagery, online gaming behavior, how long employees use which apps on their work computers, search results on Google and Baidu, products viewed on Amazon and Alibaba, customer loyalty card usage, Paypal and Alipay transactions, “liked” posts on Facebook and Twitter, comments made on Reddit and news sites, steps taken and hours spent sitting or exercising, emails and SMS sent and received, and cellphone call metadata.

If one’s personal Big Data set were to be somehow centrally coordinated and processable, it would represent a relatively effortless digital trace of what was once too difficult to trace or was legally untraceable. Science fiction writers have long warned their readers about a future where public or private entities track and rate citizens for possessing socially desirable or detrimental traits. In 2016, the personal ratings app Peeple went live, and was almost universally panned by the media and its users. Despite the attitude that rating people on a single scale is “horrible, cruel, and disgusting” as one Peeple user wrote in his iTunes app store review (see footnote 2), privacy laws in the West have not kept up with Big Data advances. In 2016, the European Union approved the General Data Protection Regulation (GDPR), granting a ‘right of explanation’ of all decisions made algorithmically or using automated decision-making processes. However, some have noted ambiguity in the regulation that obviates the feasibility

3 The most obvious example is George Orwell’s 1984. The latest example is the widely viewed “Nosedive” episode (3.1, 2016) of Charlie Brooker’s science-fiction series Black Mirror.
4 As of 03/15/2018, the Peeple app has a 1.7/5 rating on the iTunes app store. The top 1-star review pans the app on principle, calling the idea of rating people “horrible, cruel, and disgusting” (user JackTripodi, posted on Dec. 4, 2016).
of defining or protecting such a right (Wachter et al 2017).

Elsewhere, where privacy is less protected, centralized Big Data poses even more of a threat to the populace. In China, Big Data has become a new and powerful tool for public governance by virtue of its Social Credit System.

Despite the dystopic possibilities of centralized single-scale personal ratings whose parameters are determined by a political elite, credit ratings and other kinds of personal ratings could serve useful informational purposes. Financial credit ratings are hegemonic in the US and the EU. The sharing economy relies heavily on personal ratings on both sides of a transaction (Hawkins 2017). For example, Yelp ratings for services venture deeply into the personal realm; Uber allows customers to rate drivers, and drivers to rate users. Matching apps, like those that match potential romantic and professional partners, use proprietary algorithms to determine how well two people match utilizing salient user-provided information on many aspects considered representative of important relationship variables (Carr 2016; TaskRabbit 2018). The process of rating an individual based on the categories deemed important by some interested party can become extremely costly; see, for instance, Bryan Caplan’s new book that employs signaling theory to assess why especially higher education is largely a ratings mechanism and not at its core about the acquisition of knowledge (Caplan 2018).

This paper assumes that people will be incentivized to use Big Data to make signaling certain attributes less costly, and therefore, we should expect personal ratings to emerge from private sources as well as public sources. Our main claim is that an ecology of personal ratings emanating from polycentric (private and public) sources is epistemologically superior to a hegemony of centralized single-scale ratings, even though the provision of centralized ratings is
presumably less costly to users than navigating an ever-changing ecology of fragmented, competing, and perhaps conflicting ratings. The process of ratings emergence is informationally different depending on from where ratings emanate, much in the same way that a decentralized price discovery process has different informational qualities than a centralized price setting process (Hayek 1945). We use China as a case study to model the informational qualities of ratings in the framework of exploration versus exploitation, and consider which of the two mechanisms plays how much of a role in decentralized versus centralized ratings processes.

We borrow the concepts of exploration and exploitation from entrepreneurial management and ecological literature to differentiate our analysis from traditional equilibrium-type analyses. Traditional equilibrium analyses, which may be adequate in a simple, static system, may be insufficient to determine which types of ratings are more trust-enhancing in a complex evolving system where individual choices are entangled with a prevailing and rapidly changing institutional framework.

Exploitation is exemplified by neoclassical search intended to dissipate the advantages of a current institutional framework by, say, finding more complete contractual arrangements to reduce the uncertainty surrounding the expected behaviors in any given interaction or exchange (Barzel 1997; Cheung 1983). Exploitation is amenable to design. Exploration involves evolving the system into entirely new spaces of economic opportunities and available actions by means of an endogenous novelty-generation mechanism, say, by instantiating a small set of entrepreneurs who are incentivized by the institutional environment to bring in new ideas that disrupt stable product spaces and, therefore, expand the set of potentially trust-enhancing
solutions (Kirzner 1996). Exploration is a discovery process (Hayek 1978).

This paper is structured as follows: Section 1 is the introduction. Section 2 presents centralized social ratings by using China’s SCS as a case study. Section 3 investigates and categorizes how public and private personal and social ratings systems are actualized in both in China and the US, and models the co-production of a static, neoclassical personal ratings system to explain the justifications of designing a centralized ratings system in an exploitative frame. Section 4 considers public, private, and public-private personal ratings systems in an explorative frame, focusing primarily on the epistemological qualities of each kind of system. Section 5 concludes.

2. A Study of Centralized Social Ratings: China’s SCS

Credit, or xinyong (信用) in Chinese, has much the same meaning as reputation does in English. China’s social credit system began as a stated desire by the CCP (2014) to improve trustworthiness in financial transactions by instituting some kind of standardized financial credit ratings. As noted by Rogier Creemers (2017), Chinese planning takes the form of a “champagne pyramid” where officials identify some “dominant contradiction” at the top of the pyramid that impedes social progress, then identify the subproblems and sub-subproblems derivable from the dominant contradiction that must be resolved if the blockading tension is to be removed so that society can progress to its next stage of development⁵. Official resolution of dominant contradictions requires institutional design, the political form of which is expert rule,

⁵ For example, see discussions of Dobson and Kashyap (2006) on the banking system reform and Coase and Wang (2012) on the Great Leap Outward in the late 1970s
with all its attendant pitfalls (Koppl 2018). The dominant contradiction in the SCS planning document is a lack of trust and trustworthiness in financial transactions, exchange, personal interactions, judicial actions, and in government (CCP 2014). While trustworthiness affects the development of socially coordinative institutions as noted in Ostrom (2010), trustworthiness can also be created and sustained by certain institutions like “seals of approval,” membership, social cues, even accents (Botsman 2017b). China’s SCS is a type of seal of approval, whose parameters are defined and whose operation is designed by experts affiliated with or approved by the Chinese Communist Party.

2.1 Proposed Measures of the SCS

There are three basic requirements for the workability of China’s SCS, according to its own statement: 1) a credit system covering the entire country and its supporting database, 2) a “trustworthiness incentive mechanism,” and 3) an “untrustworthiness disciplinary mechanism” (CCP 2014). The language validating the expedience and necessity of an SCS is full of planning terminology, arguing that an SCS is, essentially, more coordinative of the plans of individuals and businesses in the Chinese “market socialist” system. The document stipulates that efforts will be made to improve the government “image” of honesty and transparency through standards the government sets for itself. The disciplinary mechanism was designed to be harsh, with one former party official stating that the social credit system is crafted so that “discredited people become bankrupt” (Liu 2018).

Businesses are a special focus of the report. The SCS is intended as a compliance mechanism, a way to track compliance with safety and quality standards, blacklist companies
that do not comply, and use the blacklist to target market bans and force withdrawals of some businesses. Banks will be brought into the credit rating system as a way of reducing “fraud.” Taxation will become part of the system in an apparent effort to maximize tax revenue by using the social credit system’s databases to better track property ownership and income. Tax violators will go on a blacklist. Pricing will come under even stricter control than it already is, the goal being to “standardize market price order.” The document urges the use of social credit scores as determinants of who contracts with whom in any given business deal, effectively shutting people out of commercial life if their credit scores are too low (likely, as part of the “untrustworthiness disciplinary mechanism”).

Despite the many stipulations about citizen behavior and blacklisting, there is no system in place for preventing abuse by officials, ensuring data quality and score accuracy, allowing citizens to dispute scores, or any protections for citizen privacy. As Chen and Cheung (2017) put it, “[i]ndividuals risk being reduced to transparent selves before the state in this uneven battle” (357).

Furthermore, it is unclear the Chinese suffer such a paucity of trust services to justify the creation of a single-scaled centralized social credit system. 21st century Chinese business practices demonstrate elaborate trust-gaining practices, where social gatherings with new employees and clients are essential to establishing professional trust (Chua 2012). Though Chinese had relatively low credit usage as of 2014 (36.3% of people aged 15+ compared to 51.4% in the United States), they had high levels of formal bank account ownership (79%) (Fungáčová and Weill 2014). Furthermore, Chinese credit usage as of 2014 was a good deal higher than China’s wealthy protectorate Hong Kong (25.8%), and higher even than long-
developed France (28.3%) (WorldBank Global Findex 2014).

Chinese citizens may be credit-shy because of the state’s heavy-handedness in banking. Chinese banks are nationalized; credit is issued from and controlled by the Chinese government. Chinese banks can force family-members and spouses to settle unpaid debts (He 2017). In some circumstances, like for a “malicious overdraft,” the CCP considers unpaid debts as criminal offenses. The debtor must then prove overdraft was non-malicious (Liu 2011). Even if a Chinese citizen avoids a visit from the real police after defaulting, they may get a visit from debt collectors. Debt collectors in China have been known to kidnap people and hold them until their debts are repaid (Feng 2017).

2.2 Technical Structure of the SCS

So far in the pilot program, China has given eight companies the opportunity to come up with a credit scoring system. The biggest two contenders are WeChat developer Tencent’s partner, China Rapid Finance, and Sesame (Zhima) Credit, created by Ant Financial Services Group, an affiliate of the behemoth online retailer Alibaba. Scores are linked to a Chinese citizen’s official ID (shenfenzheng). Credit scores administered through Sesame Credit range between 350 and 950 points, and users earn or lose points based on things they do (Botsman 2017a). “Citizens can earn bonus points up to the value of 200 by performing ‘good deeds’, such as engaging in charity work or separating and recycling rubbish. In Suzhou city, for example, one can earn six points for donating blood” (Zeng 2018). These ‘good deeds,’ whose goodness is determined by CCP officials, come with their own rewards. Another good deed? Sharing a link from a state-sponsored news agency praising the CCP (Osborne 2018). Citizens
with high social credit scores are rewarded with free gym facilities, high-speed broadband access, cheaper public transport, foreign travel visas, shorter wait times in hospitals, better insurance premiums, higher quality schooling for their children, access to better restaurants, and public approbation (Zeng 2018; Margolis 2017).

‘Bad deeds’ include a wide range of behavior, some forms of which are not explicitly known by most citizens, most who find the social rating system opaque (Chen and Cheung 2017). Beyond the obvious—holding too much unpaid debt—behavior that hurts one’s credit score includes “...not showing up to a restaurant without having cancelled the reservation, cheating in online games, leaving false product reviews, and jaywalking” (Zeng 2018); sharing a news article about China’s recent stock market collapse (Osborne 2018); not walking your dog enough, refusing to carry out military service, spending too much time playing video games or posting on social media, spreading “fake news,” and being a bad driver (Ma 2018); reading the wrong books, not paying utility bills on time, having the wrong friends, saying negative things about one’s friends or economy or government, having friends or family that say negative things about their friends or economy or government, and of course, mentioning Tiananmen Square (Botsman 2017a).

According to the second SCS description document put out by the CCP (2016), a bad credit score can restrict you from buying plane or train tickets, especially for traveling abroad; sending your children to “high-fee” schools; eating in certain restaurants and staying in certain hotels; purchasing insurance; purchasing or renovating property; obtaining financial credit; obtaining social security and other benefits; establishing a social organization; obtaining government contracts; becoming a Party member; joining the military; managing a SOE or
acquiring other management or “responsible” positions; taking jobs in the food sector, the drugs sector, or in construction; obtaining larger rewards in court as a legal representative; using public parks, fishing areas, beaches or forests.

If one’s score is low enough, or one’s behavior is deemed bad enough by the CCP, she will be placed on a publicly accessible blacklist (SPC 2018). Sometimes, all it takes is a single offensive action to find oneself blacklisted. Dissent can earn a citizen a blacklisting, as happened in the case of the journalist Liu Hu who has vocally criticized the Chinese government. In 2017, Hu attempted to buy a plane ticket and discovered he’d been deemed “not qualified.” He later discovered that he could not buy property, take out a loan, or travel on the posher Chinese trains (Vanderklippe 2018). In 2016, lawyer Li Xiaolin was blacklisted for failing to carry out a court order in 2015 (Wang 2017). In 2017, tech mogul Jia Yueting, the founder of LeEco, was placed on the blacklist after failing to follow a court order to turn over 477 million yuan ($72 million) (Tham & Jourdan 2017). As of February 2018, over 9.59 million people have been placed on the blacklist since it was created in 2013 (Xinhua 2018). As of May 2018, blacklisted individuals had been barred from over 11 million flights and 4 million high-speed train trips (Chan 2018). The final section of the CCP 2016 document exhorts officials to give the media “full rein” to “create public opinion pressure” against blacklisted individuals.

Other social credit systems are being developed and tested in China. Alibaba and Tencent have a large amount of data associated with consumer behavior on their sites, which they’re using to create “user credit files.” Sesame Credit uses the payment app Alipay, with more than 500 million users (Huang 2017), to monitor what links you share and comments you make using social media accounts, and what you purchase online. In addition to the massive
amount of data it gets from Alibaba and Alipay, Sesame Credit has developed partnerships to acquire data from China’s mega ride-sharing service Didi Chuxing, and China’s mega online dating service, Baihe (Botsman 2017a). The CCP has tasked Tencent with turning WeChat into an electronic ID that would become a kind of electronic social security card to access social services like pensions, medical insurance, unemployment insurance, and special leave benefits like maternity and work-related injury (Chen 2018). WeChat is able to achieve this by coercing users to verify their identities via users’ official ID.

2.3 The SCS Rollout

Chinese police have begun to wear facial recognition augmented reality (AR) sunglasses to track citizens in the train stations of Zhengzhou. Pilot programs initiated at the beginning of 2018 using facial recognition glasses to detect wanted criminals resulted in a spike in local arrests, including high-profile arrests of catching wanted criminal who had alluded capture for years. These arrests have raised concerns about human rights and lack of oversight in a country where government officials have license to exploit user data at will (Vincent 2018b).

The number of CCTV cameras in China is exploding with advanced built-in features like gait recognition (Vincent 2018a). Furthermore, artificial intelligence (AI) is being built into new cameras, allowing them to analyze footage automatically and send alerts to police in the vicinity when suspicious activity is detected (Vincent 2018b). Facial recognition can now sufficiently recognize drivers and pedestrians who break traffic laws, resulting in deductions

---

6 Using features of commercial transaction, such as food delivery, purchasing movie tickets, and paying utility bills, on WeChat requires users to verify their identities as well.
from their social credit scores and other penalties. Reportedly, drivers in roll-out cities have changed their driving behavior as a result, with many more drivers in the roll-out city of Rongcheng stopping at crosswalks (Mistreanu 2018).

In March 2018, the CCP announced in two additional documents that they would restrict train and air travel for Chinese citizens based on their social credit scores, which include red-flagged offenses like having been discovered smoking on the train in the past or having tried to board the train with an expired ticket (Reuters 2018).

3. Social and Personal Ratings Emanating from Public and Private Sources

Personal ratings are intended to represent individual characteristics and behaviors deemed amenable to the formation and maintenance of trustworthiness in social interactions, including but not limited to commercial transactions like buying and selling goods, financial transactions like obtaining loans, and social transactions like helping family and neighbors. As such, trust and trustworthiness are central considerations to any system of personal ratings, as noted in China’s rollout document for its Social Credit System. Personal ratings are intended to be trust-enhancing; the more general the ratings, the more generally they are intended to enhance trustworthiness.

Trust is present in “[v]irtually every commercial transaction” (Arrow 1982) and can radically change the outcome of any transaction (Ostrom 2010). Trust modeling in the neoclassical framework has often taken the form of suggesting ways to ameliorate incomplete, asymmetric information, or costly information (Akerlof 1970). Often, anticipatory regulation is suggested to rectify problems of information (see discussion in Klein 1997: 3).
Before suggesting an intervention is needed in a market prone to asymmetric information, we need to understand how trust emerges. Communication between individuals to reduce uncertainty and improve information symmetry is central to the development of trust. Developing a network or forum for communication often requires common rules and norms of the space, like identifying particular establishments with membership, regional, or ritualistic restrictions. See, for example: the creation of gossip networks (Merry 1984; Greif 1989), chatting (Tullock 2005), going out for cocktails with potential business partners to build guanxi (Botsman 2017b), congregating in particular establishments with membership restrictions, as in the London cafes that hatched modern stock trading (Stringham 2016), and the creation of tracking mechanisms like used by Japan’s financial clearinghouses (Ryser 1997). Such networks can utilize information more efficiently than established leadership hierarchies (Banerjee et al 2014).

Klein and Shearmur (1997) argue that when societies become more anonymous, we need some other way of determining the trustworthiness of those with whom we wish to interact, what they call ‘seals of approval’ (Klein 1997: 4). Thus, society is a “a flowing patchwork of reputational communities” (ibid: 5) where people seek seals of approval and try to obtain them for themselves. Businessmen in early 20th century rural North Carolina, for instance, carried proof of membership in the Baptist church as evidence of their trustworthiness (Weber 1948). The mechanism of trust generation is best stated by Klein and Shearmur (1997: 6-7): “Trust services often begin in community practice, embedded in social norms, but evolve by a process of expansion, amalgamation, and standardization into highly concentrated operations.”
Trust services become more expensive the larger the society, the more anonymous the transaction, and the harder it is to enforce the purchase contract (Shearmur and Klein 1997: 37). Trustworthiness is a network good, whose value increases the more people who engage in its co-production. More recently, distributed and decentralized online interactions, notorious for informational asymmetry due to the relative anonymity of online presence, have given rise to many new kinds of seals of approval like product and provider star-ratings, seller feedback, and mutual provider/user ratings in the sharing economy. Distributed and decentralized ledger technology, exemplified by the blockchain and related applications like smart contracts, is greatly reducing the cost of enforcing contracts in large and anonymous markets (Hardy & Norgaard 2015).

Consider a typical utility function for a positive-externality network good, for some user $i$:

$$
U(i) = a_i + f_i(n_{-i}) - \frac{c}{m_i}
$$

(1)

where $a_i$ is the subjective benefit of the network good to user $i$, $n_{-i}$ is the (expected) number of other users of the good, $f_i$ is the (expected) value network participation grants to user $i$, $\frac{c}{m_i}$ is the per-use cost of utilizing the network, and $m_i$ is the number of times $i$ utilizes the network. Typically, the nature of the network and the good determines $f_i$; it is not a subjective value, even though the participation threshold $n_{-i}$ depends on subjective valuation encoded in the $a_i$ term of the utility function.

Under a static analysis and a certain set of assumptions, economists compare the
solution to the maximization of aggregate utility to the aggregation of the individual utility maximizations. If the latter is smaller than the former, the good is deemed "underprovided" and, as there exists no endogenous mechanism within this abstract system to rectify the situation. A good example of apparent underprovision of a network good the provision of effort in work teams, when output and individual payoffs experience nonlinear positive externalities (Jackson and Zenou 2015). We will employ a simple co-production model below to investigate whether personal ratings systems appear to underprovide trust-enhancing services, in the static frame.

Generally, however, such simple, static models are not appropriate outside of a static frame, where entrepreneurial discovery continually improves what types of ratings are available, their quality, and cost and method of delivery. Large volumes of continuously updating data coupled with rapidly improving technologies for harnessing and making sense of that data will undoubtedly continue to provide new means for both the centralized and decentralized generation of seals of approval. Additionally, moving out of the static frame requires acknowledging various public, private, and semi-public institutional levels where co-production can take place. Therefore, we utilize a polycentric framework for our analysis that uses goal-setting and problem-solving as the basis for economic and social action (Ostrom 2005, 2010; Simon 1996), and takes epistemological issues seriously when comparing centralized versus decentralized modes of innovation (Hayek 1945; Koppl 2018).

Since talking about institutions of trust means talking about both individual action and institutional creation and enablement, we employ the concepts of exploration and exploitation from organizational adaptation research (March 1991; Gupta et al 2006). As March (1991: 71)
indirectly notes, the two forces in combination are analogous to creative destruction (Schumpeter 1934).

Exploitation tends to dissipate the advantages of existing solutions. That is, if the set of all possible choices were to remain unchanged, the process of exploitation likely leads to an equilibrium, cycling between equilibria, or some other stable social pattern. Neoclassical utility maximization and search under probabilistic learning is, in our rubric, a type of exploitation of the prevailing systemic infrastructure and economic opportunities. The structure of the economic system under question and the total number and nature of all economic opportunities available to the agent in the system does not change under exploitation.

Exploration is the process of searching for novel solutions to old or new problems, and tends to disrupt prevailing equilibria. Exploration includes the introduction of new physical elements and technologies into the space of all economic opportunities, thus expanding the set of all possible choices available not just to an individual, but to everyone in the system. The structure of the economic system under question and the total number and nature of all economic opportunities available to the agent in the system changes under exploration. Exploration’s expansion of the space of all economic opportunities robs individuals of the ability to theoretically maximize expected utility (Koppl et al 2015), particularly in goods markets that interact strongly with rapidly advancing technologies.

3.1 The Emergence and Construction of Trust

---

7 The reader should not confuse exploration with probabilistic search defined neoclassically over a space of existing economic opportunities. Exploration is not ontologically equivalent to Bayesian learning, the latter which would be considered a part of exploitation.
Especially as it is rooted in information asymmetries and free rider problems, trustworthiness has been frequently categorized as a public good (Anomaly 2017). In traditional economic analysis, trustworthiness is a solution to a commitment problem whereby the gains of deception in the short run are balanced against the gains of long-term cooperation. This calculus dictates the extent to which privately provided trust is effective. As it generally never pays to be honest in a society characterized by a paucity of trustworthiness, such a state of affairs is self-reinforcing, and thereby (according to a static equilibrium analysis) virtually impossible to alter without public intervention.

Personal ratings represent another way to mitigate the costs of trustworthiness in large and anonymous societies. Personal ratings are a type of ‘seal of approval’ indicative of personal characteristics considered coordinative of the ends of the users of any given personal ratings system. Personal ratings are as ubiquitous as product ratings in the internet age, though what we call personal ratings are usually highly localized to the platform on which users interact and are not massive generalizations across an individual’s corpus of interactions.

Personal ratings on most social media platforms consist of how many followers or “likes” someone has, how often their content has been linked or reblogged or retweeted, the difference between how often their comments have been upvoted or downvoted, whether their reviews have been rated “helpful” by other users, and other measures of honesty, quality, and influence. Most existing ratings systems focus on single aspects of a user considered most salient to other users of the platform. Multi-aspect ratings attempt to generalize single-aspect ratings, or to provide for many or customizable aspects. Multi-aspect ratings may or may not be fed into an algorithm to generate some kind of score, what we call single-scale ratings. Giant
databases like Experion specialize in acquiring information about people along as many axes possible, and are often used by legal entities, employers, and the government to score a person’s credit or criminal history\(^8\).

There are three types of personal ratings: **single-scale**, where salient bits of an individual are reduced to a single number; **single-aspect**, which rate particular aspects of a person; **multi-aspect**, which rate people along several independent axes. There are three delivery methods: **crowd-sourced**, where anonymous strangers determine the rating; **algorithmic**, where the app developers create a black-box algorithm that determines the rating; and **mutual**, where two (or three) people known to each other rate each other from their interaction (typically seen in the sharing economy).

1. **Single-scale ratings** attempt to reduce the salient bits of an individual to a single number (Peeple\(^9\), Sesame Credit). These apps have raised the most ethical concerns among users (see Reagle (2015) on Peeple and similar apps, and Margolis (2017) on Sesame Credit). These apps can be crowd-sourced (Peeple) and algorithmic (Sesame Credit), though the television show *Black Mirror* foreshadowed how single-scale ratings could conceivably be mutual.

2. **Single-aspect ratings** attempt to rate particular aspects of a person like their professional capabilities (RateMyProfessors\(^10\), Uber, AirBnB) attractiveness

---

\(^8\) Note especially that we do not include scorable characteristics like IQ, EQ, MBTI, or automated decision-making algorithms associated with parole decisions in what we consider to be personal ratings, as they are not generated by virtue of some ongoing social process.


\(^10\) [http://www.ratemyprofessors.com](http://www.ratemyprofessors.com)
(Spontana\textsuperscript{11}), intelligence (Best IQ Test\textsuperscript{12}), and MBTI-type personality (PersonalityMatch\textsuperscript{13}). These apps can be crowd-sourced (Spontana), algorithmic (Best IQ Test), and mutual (Uber).

3. **Multi-aspect ratings** allow people to rate each other (and themselves) on many axes, most notably a feature of dating apps (OkCupid, Tindr, eHarmony) and employee/employer-rating apps (Dots\textsuperscript{14}, Pluggd, TaskRabbit). These apps can be crowd-sourced (Dots), algorithmic (OkCupid) and mutual (Pluggd, TaskRabbit).

We summarize and provide examples of the three types of ratings in Table 1 (including scorable characteristics). We note in Table 1 the locus of governance associated with the personal rating system; unless marked with a single asterisk (*), all personal ratings are private.

<table>
<thead>
<tr>
<th></th>
<th>Single-aspect</th>
<th>Multi-aspect</th>
<th>Single-scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crowd-sourced</strong></td>
<td>Instagram;</td>
<td>Facebook; Dots; Reddit</td>
<td>Peeple</td>
</tr>
<tr>
<td></td>
<td>RateMyProfessors; Spontana</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Algorithmic</strong></td>
<td>IQ**; MBTI**</td>
<td>eHarmony; OkCupid; Tindr; background checks*; automated decision-making**</td>
<td>Sesame Credit*; insurance premiums*</td>
</tr>
<tr>
<td><strong>Mutual</strong></td>
<td>Uber; Lyft</td>
<td>Pluggd; TaskRabbit; AirBnB</td>
<td>Black Mirror’s Nosedive</td>
</tr>
</tbody>
</table>

\textsuperscript{11} https://spontanaapp.com  
\textsuperscript{12} https://itunes.apple.com/us/app/best-iq-test-pro/id1047228171?mt=8  
\textsuperscript{13} https://play.google.com/store/apps/details?id=com.personalityperfect.app&hl=en  
\textsuperscript{14} Childers 2017.
* these have elements of public scoring, depending on the system under consideration;  
** scorables characteristics, but not social ratings

**TABLE 1**: Examples of personal ratings systems, by type of personal rating and delivery mechanism.

Consider the quality of a personal rating as a product. Personal ratings should be coordinative of the ends of users, which means that ratings are more useful if they advance a user’s subjective ends. Consider single-aspect crowd-sourced and mutual ratings, which rate users on axes relevant to the transaction and are likely knowable by the users. Uber asks riders to rate drivers as drivers and vice-versa. Uber users don’t care or know whether or not their driver can lasso a bull, but they care and can say something about the way a driver drives.

Now, consider algorithmic ratings. Algorithms are designed by programmers or can be inferred from training data using machine-learning methods. Algorithms can fail in several ways: they can mischaracterize the effects they hope to predict, that is, their training sets can be based on bad or incomplete models, and the algorithms themselves can be poor predictors. Examples of algorithmic bias and failure exist in abundance (Brehm & Loubere 2018; Albanesius 2015; Levin 2017; Moscaritolo 2015; Moody 2017).

The various delivery types and governance loci of personal ratings illustrates the tradeoffs between the quality of ratings and the cost of generating ratings. Does a pattern-matching algorithm intended to detect faces in a crowd actually do so with reasonable accuracy? Does a matching algorithm intended to match romantic partners actually increase the number of successful relationships compared to brute-force search? Like for tangible
products, algorithms compete with each other to do a better job satisfying some user’s subjective ends and can suffer a loss of quantity and quality absent a robust competitive environment. Since actual market transactions generate crowd-sourced and mutual ratings, they should be more accurate (Page 2007) and less prone to expert and epistemological bias (Koppl 2018). However, there can be economies of scale in personal ratings, especially considering that ratings are network goods. The construction of ratings may favor more centralized or public aggregation methods if the requisite technological infrastructure to aggregate, filter, track, and publish ratings isn’t advanced enough to make small-scale ratings feasible. We summarize the tradeoffs between the accuracy and cost of personal ratings in Table 2.

<table>
<thead>
<tr>
<th>Delivery Method</th>
<th>Accuracy (high volume)</th>
<th>Cost (high volume)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crowd-sourced/mutual</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Algorithmic</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

TABLE 2: The tradeoffs between the accuracy and cost of delivering personal ratings given a high volume of market transactions, by type of delivery method

3.2 Is an algorithmic, single-scale rating system emanating from a public source a proper or useful way to symbolize and build trustworthiness?

To answer the question of whether single-scale social credit algorithms are appropriate, we must ask first whether polycentric personal ratings are theoretically underprovided. Let’s
set up a simple model to address this issue. We rely on our basic utility function for network goods with positive externalities (1), altered to represent co-produced personal ratings.

We are interested in comparing two different kinds of networks. The first is a centralized social credit system, where all trust-enhanced interactions are subject to a single scale. This is like China’s social credit system. The subjective benefit to the user is, therefore, subject to noise. The first is a specialized system that, along the axis of interest, provides a greater subjective benefit to the user. To formulate the intuition of the model, we make a few assumptions. Our first assumption, following the epistemological argument made above, is that the subjective benefit of a specialized ratings system is greater than the subjective benefit of a single-scale system.

A user’s utility for the ratings system increases with the number of potential applications of that ratings system, and its epistemological quality. But the applications of the ratings system go up with the expected number of members of the ratings system, if we assume that users wish to undergo trustworthy exchanges using the ratings system. The quality of the ratings also increases with the number of users if we are considering mutual and crowd-sourced ratings. In a sense, the users of a private mutual and crowd-sourced ratings system co-produce the "product," which is the personal rating of any given user.

Denote a particular specialized rating system as $\beta$, and usage of that ratings system by a representative individual as $m^\beta$. Frequency of usage $m^\beta$ constitutes an increase in ratings quality, $s$. This quality exhibits decreasing marginal gains, as each additional rating contributes less power to the averaging algorithm used to determine the overall rating. Therefore, we could model the quality $s$ of ratings system $\beta$ as a function of overall usage like so:
where $\gamma$ is a constant.

Overall usage is a function of network participation: $m^\beta = \sum_{i=1}^{n} \varphi_i \ast i$, where we suppose that each user $i$ uses the system $\varphi_i$ times. For the sake of simplicity, assume that $\varphi_i = \varphi$, and therefore $m^\beta = \varphi \ast n$, where $n$ is the total number of users of network $\beta$. Then,

$$s(m^\beta) = \gamma \sqrt{m^\beta}$$

(2)

Suppose the cost of participation depends also on usage, and that the marginal cost of usage is 1. So, $c = \varphi$ for each user. Suppose the subjective benefit of using the system is the same for every user, and denote the subjective benefit for user $i$ by $\alpha_i = \alpha$. Then the utility $U(i)$ for a representative agent $i$ is the benefit minus the cost, or

$$U(i) = \alpha \gamma \sqrt{\varphi \ast n} - \varphi$$

(4)

To determine optimal usage, we optimize $U(i)$ with respect to $\varphi$. The optimal frequency of usage $\varphi^*$ is

$$\varphi^* = n \left(\frac{\alpha}{2}\gamma\right)^2$$

(5)
and, therefore, we combine (5) and (4) and determine that the maximized utility per user $i$ is:

$$U(i)^* = n \left( \frac{\alpha \gamma}{2} \right)^2$$  \tag{6}$$

Summing the individualized maximized utilities, we get that

$$U_{specialized}(i)^* = \sum_{i=1}^{n} U(i)^* = n \left( \frac{\alpha \gamma}{2} \right)^2 = \left( \frac{\alpha \gamma n}{2} \right)^2$$  \tag{7}$$

Solving the planner’s problem for this system means maximizing the sum of utilities over the number of individuals in the network with respect to their individual usage:

$$\max_n \left\{ \sum_{i=1}^{n} \alpha \gamma \sqrt{\varphi * n - \varphi} = n(\alpha \gamma \sqrt{\varphi * n - \varphi}) \right\}$$  \tag{8}$$

Solving (8) in the usual way, we find that the overall maximized utility is the same as in the specialized case:

$$U_{centralized}(i)^* = \left( \frac{\alpha \gamma n}{2} \right)^2 = U_{specialized}(i)^*$$  \tag{9}$$

This isn’t very surprising, as there’s nothing to differentiate the two systems. So, regardless of personal ratings being a network good, there doesn’t seem to be a theoretical underprovision by the market in our simple ratings co-production model in idealized
exploitative and static conditions.

One might then ask whether there are economies of scale in the production of ratings. Might public provision of ratings decrease per-use costs for individuals to the point where centralized ratings yield apparent gains to overall utility relative to specialized ratings? Let’s investigate this case.

The setup when costs and benefits differ between different ratings systems is similar to what we did above. The utility functions faced by individuals in each system are:

\[
U_{specialized}(i) = \alpha_1 \gamma \sqrt{\varphi \cdot n} - \chi_1 \varphi \tag{9}
\]

\[
U_{centralized}(i) = \alpha_2 \gamma \sqrt{\varphi \cdot n} - \chi_2 \varphi \tag{10}
\]

where \(\alpha_1\) is the \(i\)-th individual’s subjective benefit from being a member of the specialized ratings system, \(\alpha_2\) is the \(i\)-th individual’s subjective benefit from being a member of the centralized ratings system, \(\chi_1\) is the cost-per-use of utilizing the specialized ratings system, and \(\chi_2\) is the cost-per-use of utilizing the centralized ratings system. \(\gamma\) is the network efficiency, which we assume is the same for both systems, \(n\) is the number of people in the system, and \(\varphi\) is how many times an individual chooses to utilize the ratings system.

We make two key assumptions in our forthcoming analysis. The first, as per our discussion of the relative epistemological efficiency of using specialized versus centralized ratings systems, is that the subjective benefit of utilizing a specialized ratings system is at least as great as the subjective benefit of utilizing a centralized ratings system. So, \(\alpha_1 \geq \alpha_2\). Our second assumption is that there are economies in scale in centralized ratings, such that the per-
use cost of utilizing the specialized system is at least as high as utilizing the centralized system.

So, \( \chi_1 \geq \chi_2 \). The question we want to ask is whether there are combinations of

\((\alpha_1, \chi_1), (\alpha_2, \chi_2)\) such that a centralized ratings system yields higher overall welfare than

specialized systems.

First, let’s solve the individual’s maximization problem for (9).

\[
\max_{\varphi} \left\{ \alpha_1 \sqrt{\varphi \ast n} - \chi_1 \varphi \right\} \to \varphi^* = n\left(\frac{\alpha_1 \gamma}{2 \chi_1}\right)^2
\]  

(11)

Substituting \( \varphi^* \) back into (9) and summing over the utilities, we get the overall utility for the specialized system:

\[
\sum u_{\text{specialized}(i)}^* = \left(\frac{\alpha_1 \gamma n}{2 \chi_1}\right)^2
\]

(12)

Next, we solve the central planner’s problem for (10). The central planner maximizes overall utility with respect to \( n \), as she cannot affect \( \varphi \).

\[
\max_{\varphi} \left\{ \sum_{i=1}^{n} \alpha_2 \sqrt{\varphi \ast n} - \chi_2 \varphi \right\} \to \varphi^* = n\left(\frac{\alpha_2 \gamma}{2 \chi_2}\right)^2
\]

Substituting \( \varphi^* \) back into (10), we get the overall utility for the specialized system:
\[ \sum U_{centralized}(i)^* = \left( \frac{a_2 y n}{2 x_2} \right)^2 \]  

(13)

How \( \sum U_{centralized}(i)^* \) compares to \( \sum U_{specialized}(i)^* \) very apparently depends on the relative magnitudes of the parameters. That is, if we assume \( \alpha_1 > \alpha_2 \) such that \( \alpha_1 = \sigma \alpha_2 \) where \( \sigma \) is a positive scalar, and \( x_1 > x_2 \) such that \( x_1 = \tau x_2 \) where \( \tau \) is a positive scalar, we care about the relative magnitudes of the scalars, or the ratio \( \rho = \sigma / \tau \). The cases for different values of \( \rho \) are as follows:

\[
\begin{align*}
\sum U_{centralized}(i)^* &= \sum U_{specialized}(i)^* & \text{when } \rho = 1 \\
\sum U_{centralized}(i)^* &< \sum U_{specialized}(i)^* & \text{when } \rho > 1 \\
\sum U_{centralized}(i)^* &> \sum U_{specialized}(i)^* & \text{when } \rho < 1 
\end{align*}
\]

(14)

So, for the static neoclassical case, what we’ve called exploitation, the benefit from utilizing a specialized ratings system needs to be at least as great as the costs-savings of using a single-scale centralized ratings system for the specialized system to claim Pareto optimality relative to the centralized system. As one does in static analyses, we have ignored all potential feedbacks between a single-scale centralized system and the scope of human action. These feedbacks aren’t negligible, as evidenced by the example of China’s social credit system. A single-scale centralized system subjects all types of behaviors and characteristics to one rating scale, and worse, provides no scope for escaping abuse if the ratings system turns out to discriminate against a significant minority of users. In order to adequately treat competition effects and feedbacks, however, we must leave the exploitative frame.
4. The Epistemological Considerations of Exploration

Suppose we find ourselves in the situation where a centralized personal ratings system is apparently trust-enhancing on the system level and therefore welfare-enhancing compared to a private specialized ratings system. An apparent centralized solution in the neoclassical (exploitative) frame does not necessarily imply a solution in the open-ended evolutionary (explorative) frame (Devereaux & Wagner 2018). Polycentric systems allow for public, private, and public-private solutions to apparent trust provision problems. Examples of successful polycentric provision of public goods include polycentric policing services (Ostrom 2010, Stringham 2016), lighthouses and lightships (Candela & Geloso 2018, Foldvary & Klein 2003), common-pool resource provision (Ostrom 2010), and even stock market regulation (Stringham 2016).

The provision of trust-enhancing personal ratings depends on economies of scale in our static model, as optimal cost-per-use faced by the individual is an increasing function of the number of users in the network, as described by Equations (12) and (13). In a static, exploitative frame, public provision is justified when \( \rho = \sigma / \tau < 1 \), where \( \sigma \) is the additional benefit of and \( \tau \) is the additional cost of membership in a specialized system. Is this still true in a dynamic, explorative frame?

Trustworthiness in large, anonymous networks is theoretically enormously costly to produce. Yet, seals of approval like product ratings and new mechanisms have emerged to mitigate the costs of creating and enforcing contracts. In assuming that trust-enhancing ratings are best provided and administered centrally and on a single scale, China’s social credit system subverts the Kleinian format of trust services beginning in community practice.
Skipping the community level has grave epistemological implications for the future of trust-enhancement welfare gains more generally. A centralized social credit system tends to reflect the values of the designers and not the community those services were designed to assist. Single-aspect ratings greatly compress information about an individual’s characteristics relevant for any given ratings-based interaction; that is, the ratio of “noise” to “signal” is high relative to specialized ratings systems. In the last section, we reduced our world to a single specialized system versus a single centralized system. Portraying the problem instead as an ecology of multiple specialized systems specializing in different and sometimes overlapping characteristics versus a single centralized ratings system is more realistic.

The compression of information in the form of the reduction of dimensionality can destroy both the salience and representativeness of personal ratings. Simply, while your butcher might not care about whether you’ve ever been caught smoking on the train, the compression of all your characteristics into a single number would force him to care. Less compression of all information, less loss of salient information, more representativeness of personal ratings to particular transactions, more enhancement of trustworthiness and therefore coordinativeness. That is, an ecology of aspect-focused personal ratings is epistemologically superior to a single-scale ratings system.

An explorative frame is a frame in which entrepreneurs continuously disrupt systemic ratings hegemonies with potentially more coordinative ratings systems. The Deep Web has solved the problem of extreme social extendedness and anonymity in a purely private way by relying on tools like blockchain-ledger smart contracts and dark e-commerce sites that both vet and punish defectors on either side of the sales contract (Norgaard and Hardy 2015). In
essence, the Deep Web solved the trustworthiness problem presented by extreme information asymmetry by making risk symmetric. Risk symmetry provided by blockchain technology is the Deep Web’s ‘seal of approval.’

An explorative frame is a frame in which competitors aren’t pre-reconciled but in active competition to realize some radically uncertain outcome. Centralized single-aspect ratings have no endogenous mechanism for improving the ratings system through learning. Contrast that with private, specialized ratings systems: if a user accumulates a bad rating as an Uber driver or rider, they can presumably start over with Lyft. Similarly, if Lyft’s ratings system is less trust-enhancing than Uber’s, presumably Lyft can tweak its algorithms to stay competitive.

5. Conclusion

We have demonstrated in the preceding sections that a centralized single-aspect ratings system is sometimes justified in a static exploitative frame, and generally unjustified in an explorative frame. As social systems are characterized by both exploitation and exploration, centralized single-aspect ratings seem poised to do more harm than good in the systems in which they are implemented, compared to an alternative state in which private ratings systems are allowed to emerge and compete.

Fragmenting ratings along axes relevant to users and closer to the knowledge of users (like whether your Uber driver drove well) improves the epistemological content of that information. Algorithmic delivery has the potential to further drain ratings of salience and representativeness as it is the developer’s expert knowledge and not local knowledge that drives model outcomes. Therefore, centralized algorithmic single-scale ratings, for which there
is no alternative as in China’s SCS, are discoordinative compared to current alternatives.

Single-scale ratings inevitably discriminate between personal qualities, as the qualities needed to successfully combat informational asymmetries depend on the interaction. The halo effect, which is the false inference that if a person is good at A they must be good at B or C or D, encapsulates how automatically correlating between attributes makes discrimination based on certain traits seem rational (Rosenzweig 2007). The halo effect turns certain traits, like physical attractiveness, into more generally applicable seals of approval, while ignoring other more salient and representative traits that are perhaps costlier to determine. Projecting personal qualities onto a single scale requires judging some subjective interests and activities as inferior to other subjective interests and activities. Playing video games becomes inferior to having a family. Individuals are nudged, in a carrot-and-stick fashion, away from some activities and towards others.

While personal ratings systems will invariably assist coordination in the areas in which they are useful, it is unreasonable to expect the subjective usefulness of ratings to be universal and for all possible ratings to emerge in one particular institutional ecology. “Smart” technologies, like data-driven technologies that use algorithms to improve user experiences and outcomes, are emerging in a piecemeal fashion (Chakravorti and Chaturvedi 2017). Furthermore, the emergence of blockchain technologies may make be quite disruptive to many current social and political hegemonies (Cowen 2018) and may require different institutional ecologies to be most effective.

And what of China’s SCS? In the short run, China’s social credit system will likely be effective due to the authoritarian nature of the Chinese system. Some citizens—often those
with high credit scores—do not object to the social credit system, saying that they have seen the behavior of their fellow citizens respond positively during the implementation of the social credit system in their area (Ma 2018).

However, authoritarianism is both the key to the unitary effectiveness of a centralized ratings system, and the reason why a centralized ratings system is likely to result in—and arguably has already resulted in\(^\text{15}\)—a raft of human rights abuses. In a system with a mix of public and private services, citizens can ostensibly appeal to a public entity when they have been abused by a private entity. China’s SCS is designed to punish citizens for deviating from a set of behaviors CCP officials have determined to be what constitute a good citizen. Deviations and non-conformity merit punishment, often severe in comparison to the so-called crime. Citizens have no recourse if they believe the SCS is abusive or ineffective.

China has recovered in many ways from the catastrophe of Maoist economics and the Cultural Revolution. But measures like the SCS threaten to take China back to a time where China was more focused on meeting the goals of its communist leadership than advancing the well-being of its people. The Chinese government is weaponizing Big Data against its citizenry. This may be the moment Chinese communism proves it cannot stably persist in a manner that protects its citizenry from the abuses historically associated with other communist experiments, and with its own Maoism.

The Chinese Communist Party’s stated reasons for its social credit system sound good on paper, but upon immediate reflection do not hold. The stipulations of the SCS proposal seem

\(^{15}\) See especially the cases of the lawyer Li Xiaolin and journalist Liu Hu (Wang 2017), tech mogul Jia Yueting (Tham & Jourdan 2017), and the over 9 million other Chinese blacklisted from taking planes or trains or taking out loans (Ma 2018).
aimed at dangling the freedom and wealth of market participation in front of Chinese citizens like a carrot, turning an arena that has improved lives and the trajectory of communities into a lever of control. As Deng Xiaoping proved, it is the market that makes China’s special brand of socialism possible. China’s communist leadership walks a knife’s-edge between using the power of the market to advance the political goals of the CCP, and sacrificing the elements of Deng’s socialism that have immensely improved the lives of virtually all Chinese citizens.

References


Hawkins, A. J. (Apr. 26, 2017). “Uber just made it way easier to see your rating in the


