

Social Credit Systems

governing in the age of technology

By

Thomas de Dooij
5689953

Utrecht University Department of Philosophy

Faculty of Humanities

First supervisor: Martin Blaakman
Second supervisor: Jo van Caeter

Date of submission on the 19th of June, 2020

Abstract

The social credit system is a governing device which has been criticized for being oppressive and dystopian by nature. In this paper, it is argued that this criticism is unfair towards social credit systems as a phenomenon. The goal of this paper is to answer the question: “Are social credit systems compatible with libertarian paternalism?”. This is done by first determining what the shared values are of social credit systems. Social credit systems are centralized reputation systems wherein every social encounter can be regarded as a transaction. Social credit systems are compatible with libertarian paternalism to an extent. Binding a reputation to a user, and making this reputation public, is a perfect example of Nudge Theory in practice as users will alter their behavior based on the reputation of other users. However, in the decision-making process of creating a social credit system there could be bound consequences to the system by the governing party, diverting social credit systems from libertarian paternalism. Social credit system can also fall prey to autonomy gaps. However, not all variants of social credit systems will suffer from autonomy gaps equally, as not every system presupposes the same capabilities of the users. Lastly, as the transparency and normativity of social credit systems are different from system to system, it is unfair to compare social credit systems, as a phenomenon, to dystopian models. The opposite holds true: social credit systems may provide a libertarian paternalistic answer to governing in the age of technology.

Contents

Abstract.....	2
Introduction	4
Chapter 1: To build a Social Credit System	6
1.1 Information and Reputation System.....	6
1.2 Three Variants of the Social Credit System.....	7
1.2.1 The Broad System	7
1.2.2 The Small System	8
1.2.3 The Cumulative System.....	8
1.3 The normative problem of virtue ethics	9
1.4 Transparency.....	10
1.5 Conclusion.....	11
Chapter 2: Libertarian Paternalism, Autonomy Gaps and Social Credit Systems.....	12
2.1 Libertarian Paternalism and the nudge theory.....	12
2.2 Autonomy Gaps	14
2.3 Libertarian Paternalism and the Social Credit System	16
2.4 Autonomy Gaps and the Social Credit System.....	17
Conclusion.....	19
Bibliography	20

Introduction

After the failure of the cultural revolution in the '70s, the Chinese government had to rethink and reshape social cohesion amongst its people.¹ Prominent Chinese scholars soon worked in cooperation with the government to create a system-based structure to achieve this goal. The plans worked out during this time can be seen as the first blueprints of what is called the *Social Credit System*. The Chinese government aims to enhance the trust in every aspect of the societal being by monitoring, processing and grading virtually every aspect of behaviors which could indicate a certain aspect of a by the Chinese government predefined conception of the virtue *Trust*.² Now, almost 50 years after the first plans were drafted, the system is almost ready for full implementation.

The system comes at a time when a new competitor has entered the arena of political philosophy: *Libertarian Paternalism*. Championed by Sunstein and Thaler, the coiners of this term, libertarian paternalism upholds the idea that paternalistic governance is compatible with libertarian values, A position that was thought to be impossible in the libertarian dogma at that time.³ The arguments brought up by Sunstein and Thaler concerned the misconceptions of paternalism, most notably the partial rejection of the absolute rationality of man. Their dissatisfaction with these misconceptions resulted in the introduction of a variant of paternalistic governance which is in coherence with the values of libertarianism as it is based on the concept of *Nudging*. Nudging is an often subtle way of “pushing people in the right direction” for their own good, without taking away nor reducing freedom of choice. Libertarian paternalism did not enter the academic world unanswered, and since its introduction several critics have put up offenses on the theory. One critic for instance—Joel Anderson—suggests that it is not always a lack of rationality on the part of humans which contributes to the makings of irrational choices, but rather a fundamental lack of attention on the part of the policymakers concerning the development of individual autonomy. The capacities and capabilities of individuals which are implied in governance are therefor not always compatible with the actual capacities and capabilities of these individuals simply because it is not realistic for the individuals whom the governance concerns to have all these qualities.

The social credit system is built on a policy which *could* be considered surprisingly liberalistic paternalistic, as it almost completely build on the concept of *Nudging*—though admittedly not every part of the policy is without the taking away of choices.⁴ A philosophical question can be raised here: is the social credit system in accordance with libertarian paternalistic values? The answer to this question, for the Chinese system, would be quickly answered by “no” as it does restrict users with a low reputation score.⁵ However, it is not hard to imagine that the question could have been a lot more difficult to answer if only a few aspects of the Chinese system were different, without changing the basic structure of the system. If such a system would be implemented in the Netherlands for instance, would every aspect still

¹ Creemers, Rogier, 2018, *China's Social Credit System: An Evolving Practice of Control*, Leiden: Social Science Research Network, 22 may.

² Ibid. See also Central Government of China, 2014, "Notice of the State Council on Printing and Distributing the Outline of the Construction of the Social Credit System (2014-2020)."

³ Sunstein, Cass R., and Richard H. Thaler, 2003, "Libertarian Paternalism Is Not an Oxymoron", *The University of Chicago Law Review* 1159-1202.

⁴ Creemers, Rogier, 2018.

⁵ Ibid.

be the same as the Chinese system? A second philosophical question flows from this premise: what are the core values of the social credit system? what is the *idealtyp*e social credit system?⁶ Can other system be created, or are there even existing systems that can be considered social credit systems?

The goal of this paper is to answer the question "Are social credit systems compatible with libertarian paternalism?" To achieve this, the phenomenon 'social credit system' will be conceptually analyzed considering libertarian paternalism and Autonomy Gaps. The paper is divided into two chapters. The first chapter concerns the question 'what is a social credit system'? In this chapter, it will be concluded that the social credit system is a group of systems which can be grouped and classified as *centralized reputation systems*. Three distinct variants of the social credit system will then be discussed: broad systems; small systems; cumulative systems. Special attention will then be given to the answer of the social credit system to virtue ethics, and the level of transparency of social credit system. In the second chapter, the social credit system will be analyzed and classified according to libertarian paternalism. Finally, the critique of autonomy gaps will be applied to social credit systems. The conclusion will be as follows: the Small system and the Cumulative System have the potential to adhere to libertarian paternalism and leave little to no autonomy gaps, whereas the broad system will too adhere to libertarian paternalism but will be more vulnerable to autonomy gaps.

⁶ Kim, Sung Ho, 2017, *Max Weber*, Website, Edited by Edward N. Zalta, Prod, The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=weber>.

Chapter 1: To build a Social Credit System

What is a 'social credit system'? The name 'social credit system' refers to the national reputation system as developed by the Chinese Communist Party.⁷ This system aims to improve the virtue of trust amongst the Chinese citizens by assigning a reputation to the users of the systems, wherein this reputation expresses a quantified character or character trait, in this case 'trust'. In the process of making the Chinese social credit system decision on the rules, laws and layout of the system had to be made. The final product might therefore look different than it could have been if the outcome of said decisions had been different. In this chapter there will be tried to try and define what the shared values of such systems are, and what other variants of what could be called 'social credit systems' are possible.

1.1 Information and Reputation System

First and foremost, the social credit system is an *Information System*. Information systems are defined as follows: "Information system, an integrated set of components for collecting, storing, and processing data and for providing information, knowledge, and digital products. Therefore the main working of the system are the gathering, the storing, and the evaluation of information."⁸ What is the information, knowledge or digital product which a social credit system provides? It provides a *reputation*. The social credit system is therefore a certain variant of information system, namely a *Reputation System*. Jøsang defines a reputation system as follow: "In the case of reputation systems, the basic idea is to let parties rate each other, for example after the completion of a transaction, and use the aggregated ratings about a given party to derive its reputations score."⁹ As social credit system provide a platform for rating, they can be considered a variant of reputations systems called *Centralized Reputation Systems*. How do these systems

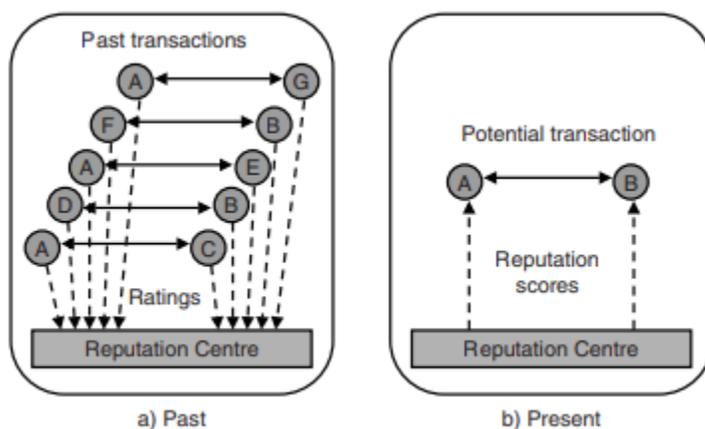


Figure 1. visualization of a centralized reputation systems (Jøsang 2007)

work? In short, there is a centralized reputation center which tracks and scores the ratings of past transactions between two parties. In the case of social credit systems, virtually any actor can be described as a party if it is able to act in the transaction. Each actor within these systems therefore builds up a reputation based on their past transaction with other actors within the system. This

⁷ Central Government of China, 2014.

⁸ Zwass, Vladimir, 2017, *Information System*, Website, Encyclopædia Britannica, inc., 28 December.

⁹ Jøsang, A, 2007, "Trust and Reputation Systems", *Foundations of Security Analysis and Design IV*.

reputation reflects a certain trait or characteristic, or even an entire character of the party. The defining trait of the social credit system is the way 'transactions' are interpreted, as these could be any social encounter as long as it is defined within the laws and rules of the system whereas 'traditional' reputation systems are usually only concerned with digital behavior. In the case of social credit systems, a transaction could be an encounter on the street, a play on stage, a work relation, a school relation, a basic review of a service, even the interaction with the environment, all next to digital behavior. The range of options of what could be defined as a 'transaction' is wide, and the borders of definitions are vague. It is almost guaranteed to pass the definition of transaction if it is stated clearly in the rules and laws of the system itself. Lastly there must be remarked that the reputation built up in social credit systems can be any characteristic or character. The Chinese system defines the reputation as 'trust', but the reputation could also be defined as 'eco-friendliness', 'integrity', 'honesty', 'sustainability', 'generosity', 'kindness' etc. Possible variants of social credit systems could have a host of possible, and not to overstate less controversial applications than the Chinese variant.

1.2 Three Variants of the Social Credit System

From the description given above at least three variants of social credit systems can be deduced which will be named here as follows: the *broad* system, the *small* system, and the *cumulative* system.

1.2.1 The Broad System

The broad system is most recognizable, as many of the Chinese variants can be classified as such.¹⁰ The basic idea of this variant is to take as much data as can be possibly gathered and deduce one single score from this data. This score should then be an indicator of a very broad characteristic of the party which it is applied to, possibly even the entire character. This variant is characterized by the use of intense data gathering, massive data storage and a reliance on big data analysis to deduce the final score. Setting up a broad system will require a lot of time, effort, finances, and expertise. Some examples of already existing broad systems are the local initiatives of the Chinese government and Zhima Credit (also known as Sesame Credit).¹¹ In future workings there is of course the Chinese national social credit system, which will be a cooperation of multiple governmental and non-governmental parties. Though Social Media platforms like Facebook, Google, Twitter et al. are not social credit systems in and of themselves, the first two mentioned have the potential to become, or to design broad systems based on their already existing infrastructure.

¹⁰ Something to take into consideration is that in China at the time of writing only regional initiatives have been put into practice. A national initiative is on its way and should be launched in 2020 according to the Chinese plans for the system as per Central Government of China, 2014.

¹¹ Credit Sesame, Inc, n.d., <https://www.creditsesame.com/>.

1.2.2 The Small System

Small systems are, as the name suggest, not as impressive as the broad systems in terms of its undertaking. The basic idea of these variants is to indicate only a small characteristic of the users. Though the use of intensive data gathering, data storage and big data analysis can be part of small systems, it is far from a requirement. The already existing systems demonstrate this point. One such systems can be found in Airbnb.¹² Airbnb works on the premise of mutual trust between hosts and guests. After the transaction is completed, and the guests have left the property, the host is asked to describe certain characteristics about the guest. By making these reviews about the character of the guests public, Airbnb has effectively introduced a small social credit system to their platform. Because the data storage, gathering and evaluation is limited, this is an example of a small system. Other already existing examples would be the aforementioned Facebook, specifically their marketplace.¹³ There is a system in place for users and buyers to rate each other's trust using stars as a means of quantification. This system too would be labeled as a small system, as the scope of the project is very limited, especially in comparison to Zhima Credit and the Chinese initiatives. There are more examples of already existing systems like these, like for instance Metacritic and Uber.

1.2.3 The Cumulative System

The third variant, the cumulative system, Is a co-operation of multiple small systems. Though the Chinese system too is a co-operation of multiple platforms, the main difference between the broad systems and what would be called a cumulative system is that the broad systems are a coordinated effort by just one party, while the systems that make up the cumulative variant all are working completely independent from each other. Zhima credit is a great example to demonstrate this difference, as it is only one company that runs the entire system from the gathering to the analysis. The Chinese variants lay in the grey area, as there are multiple non-governmental agencies working together with governmental agencies to realize the projects. The reason why these systems, and the eventual nation wide one, would be qualified as broad rather than cumulative is due to their close co-ordination with the Chinese government. The ideal cumulative system would be a cooperation of multiple agencies or platforms that work together to achieve a single system without any coordination from a central command, i.e. a decentralized organization. There still could be a centralized element as all the data could be gathered on one platform. however, all the elements of the information system (data gathering, storage and analysis) must be completely decentralized if it is to be considered a cumulative system.

¹² Airbnb, <https://www.airbnb.com/>

¹³ Facebook, *How do ratings work on Marketplace?*, https://www.facebook.com/help/915385548593204?helpref=about_content.

1.3 The normative problem of virtue ethics

Reputation systems, and the social credit system in particular, aim to quantify the character, or a character trait, of the users of the system. In ethical theory and philosophy there is a long tradition of the study of the character of the individual: *Virtue Ethics*. Social credit systems can be seen as an undertaking to catapult virtue ethics into the age of electronics.

To assert that the Chinese social credit system aims to modernize virtue ethics, one must look no further than its supposed goals. These are set as follows: *“With the support of credit information compliance application and credit service system, the establishment of the concept of integrity culture and the promotion of the traditional virtue of integrity are the internal requirements, and the incentive and punishment mechanism of trustworthiness and breach of trust are the reward and punishment mechanisms.”*¹⁴ As social credit system aim to quantify the character, or a character trait of its user, it is susceptible to at least some of the traditional criticism on virtue ethics.¹⁵ One such criticisms on Virtue ethics is the question concerning normativity. Virtue ethics usually do not enlighten the individual on what to do, rather directing them towards a simple question or its antithesis: “What would a virtuous person do in the given situation?” and “what would a vile person do in this situation?” respectively. Or, as is often rephrased: “Do what is virtuous” and “do not do what is vile”.¹⁶ The Chinese social credit system goes about this problem by defining virtues and vices within the system itself. Therefore the question to be asked shifts from the ideal moral actor or ideal action to the system itself in the form of “what would the system have me do?”. Although it is a practical way to circumvent the problem of normative ethics, this does set the system up for criticism as one could question whether the rules of particular systems actually constitute what it means to be virtuous. As the Chinese government make up the rules of the system, it by proxy the ethical values of the Chinese society. Examples like the aforementioned Airbnb, Uber and the Facebook marketplace however show that this criticism is not applicable to all social credit systems. Systems like the ones mentioned rely on a democratic determination of what it means to have a good reputation, as it is the users themselves that judge each other. Therefore the question for these kinds of systems is not “what would the system have me do”, rather, “what would the users of the system have me do?”. This is supported by the notion that the consequences bound to the reputation in systems like these are almost fully in the hands of the users of the system, not the governing actor.

The problem of normativity is something that will inevitably haunt social credit systems. However, the answer social credit systems give concerning normativity is not determined a-priori as the system do not inherently have an answer. As demonstrated, ethical judgement does not have to be carried out by the administrators of the system.

¹⁴ Translated using Google Translate from Central Government of China, 2014. Original text: 以信用信息合规应用和信用服务体系为支撑，以树立诚信文化理念、弘扬诚信传统美德为内在要求，以守信激励和失信约束为奖惩机制，目的是提高全社会的诚信意识和信用水平。

¹⁵ For a full list of objections (not all are discussed here), see Hursthouse, Rosalind, and Glen Pettigrove, 2018, "Virtue Ethics", *The Stanford Encyclopedia of Philosophy*.

¹⁶ Hursthouse, Rosalind, and Glen Pettigrove, 2018, "Virtue Ethics", *The Stanford Encyclopedia of Philosophy*.

1.4 Transparency

Another aspect to consider is *Transparency*. The system been compared to the dystopian surveillance system in 1984 by George Orwell or the panopticon of Foucault as the impending Chinese national social credit system will incorporate a system known as 'Skynet', which is essentially a mass surveillance system.¹⁷ While the incorporation of Skynet makes this comparison possible towards the Chinese social credit system, it is unfair towards social credit systems in general to make a similar comparison.

Applied to the social credit system, transparency has two dimensions: *Peer-to-Peer* and what I will name here *Administrator-to-Peer* for ease of use. The first dimension is straightforward: How much

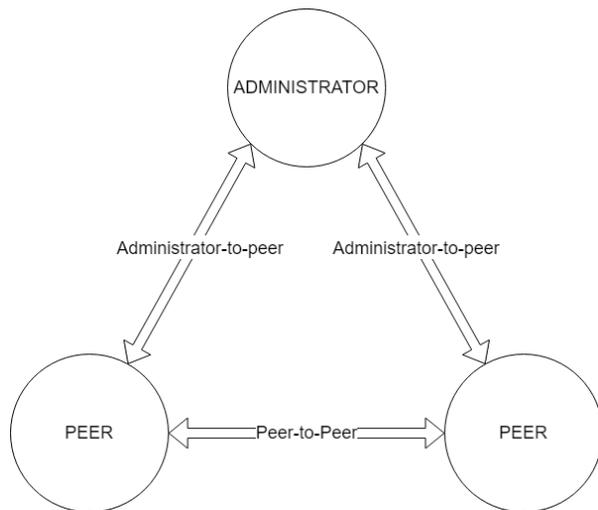


Figure 2: Visualization of the transparency dimensions

information will be available from one user to another? Will only the deduced 'end score' be available for view, or will the make up of this score be available? If the latter is the case, will their entire history be available, or only a part? If the system has multi-layered integration (for instance Corporations and Clients), will group X be made available information of group Y but not vice versa? Or will group Y be made available more information than group X while the information is still a two-way street? For social credit system one thing is essential to make a difference between social credit systems and *Surveillance Systems*: In the economics of information exchange and transparency there must

be a certain degree of peer-to-peer information transparency. The purpose of these systems is to assess the supposed behavior of the other party, so a certain degree of information about the other party must be made available. The second dimension, administrator-to-peer, is a bit more complicated. How much information about the system itself will be made available to the users? Take the matter of gathering information for instance. The Chinese system aims to integrate CCTV surveillance and facial recognition in their system using Skynet.¹⁸ This very active, and for the user constantly visible form of observation could have some serious impacts on both the welfare and behavior of their users.¹⁹ Furthermore: will users themselves be asked to score their peers? Will there be "informers" integrated in the information gathering—users that actively track the information of their peers? Transparency from administrator to peer also accounts for the rules. Will the rules of the system be made available? Will users know what will influence their score/reputation? Will the workings of the system be made available to the public? The answer to all these questions is neither yes nor no for social credit systems in general, as the right transparency profile can differ from system to system.

¹⁷ Shen, Xinmei, 2018, "'Skynet', China's massive video surveillance network", *Southern China Morning Post*.

¹⁸ Ibid.

¹⁹ Roessler, Beate, and Dorota Mokrosinska, 2013, *Privacy and Social Interaction*, Amsterdam: Sage Publications.

To summarize: there are many questions to be considered when considering transparency while building a social credit system, and the answer to all those questions are not set in stone. Each individual system can have its own unique transparency profile. While transparency is something to take into consideration when building a system, by far all elements of this dimension—like mass surveillance from a ‘big brother’-esque observer using CCTV and facial recognition—are essential to all social credit systems.

1.5 Conclusion

To build a social credit system one must make a certain variant of *information systems*, namely, a *centralized reputation system*. The core commitment of such a system would be to quantify, to a certain degree, a character or character trait of individuals in the form of a reputation. What exactly that reputation is, is dependent on the goal of the system. The problem of normativity is something that will inevitably haunt social credit systems. However, the answer social credit systems give concerning normativity is not determined a-priori as the system do not inherently have an answer. Lastly, the transparency profile of each social credit system can be unique, so while the known Chinese variant can be compared to dystopian systems, this comparison is far from fair when applied to social credit systems as a phenomenon.

Chapter 2: Libertarian Paternalism, Autonomy Gaps and Social Credit Systems

In this section there will be given an in-detail account on libertarian paternalism using the definition given by Sunstein and Thaler. Afterwards, special attention will be drawn to a criticism on libertarian paternalism given by Joel Anderson called *Autonomy Gaps*. The social credit system will then be placed within the context of libertarian paternalism, followed by autonomy gaps. It will be concluded that social credit systems by itself are libertarian paternalistic, and that implementing such a system without binding any consequences to reputation will be a libertarian paternalistic form of governance. However, the binding of consequences by the governing party could divert it from a libertarian paternalism cause. There will also be concluded that the small and cumulative systems are not likely to be susceptible to autonomy gaps, while the broad systems are susceptible to autonomy gaps.

2.1 Libertarian Paternalism and the nudge theory

Libertarian paternalism seems like a contradictory terminology if one is familiar with libertarianism and paternalism. Yet Cass Sunstein and Richard Thaler, the coiners of the term, have successfully argued throughout multiple papers that libertarian paternalism is not an oxymoron.

Libertarianism is an ideology that supposes complete freedom of action of individuals, devoid of any interference from any governing actor.²⁰ Sunstein and Thaler put up an offensive on this ideology by discrediting an assumption which is often associated with libertarianism: *“The false assumption is that almost all people, almost all of the time, make choices that are in their best interest or at the very least are better, by their own lights, than the choices that would be made by third parties.”*²¹ This critique stems from what is often seen as the antithesis of libertarianism, namely, paternalism. Standing in direct contrast to libertarianism, paternalism does presume that governing third parties sometimes do make better choices than the individual. A paternalist could therefore find justification in interfering in the decision-making process of choices if it is within the interest of either the individual, the collective or the ‘greater good’ depending on the different variants of paternalism. Libertarian paternalism, at first glance, seems like a *contradictio in terminis*.

In *Libertarian Paternalism is not an Oxymoron*, Sunstein and Thaler bring forth several misconceptions after the initial first false assumption as given above. They begin by stating that in many situations there simply are no alternatives to paternalism, as by making the action as a planner you will always inevitably influence the decision making. For example, by setting up a cafeteria you *must* arrange your products, there is no alternative. And your arrangement of those products will influence the

²⁰ Sunstein, Cass R., and Richard H. Thaler, 2003.

²¹ *ibid*

decision making of the customer. A conclusion to be drawn from this premise is that the preferences of customers are not set as their preferences clearly differ when the only changed factor in the decision making is the presentation of the products: *"If the arrangement of the alternatives has a significant effect on the selections the customers make, then their true "preferences" do not formally exist."*²² In another paper they define this as *bounded rationality*, as the customer can be described as rational only within a set situation, therefore their rationality is bound to the context of the decision making architecture.²³

The second misconception they bring up concerns coercion. Paternalism is often understood to be fundamentally coercive, yet it need not be. In the cafeteria example as given above there is no coercion for example. Yet the decision making of the customer is influenced by a governing actor, as the design made by the creators of the cafeteria will, in general, influence the choice of the customer. Another example could be the use of CCTV. Just by making the presence known of an observer in the form of a camera the behavior and choices of the people being observed can be influenced.²⁴ It is therefore unhelpful to consider whether or not to be paternalistic as almost all decision making will inevitably be paternalistic. Instead, Sunstein and Thaler suggest we should design programs using a utilitarian welfare analysis where there is more focus on the cost and benefits of outcomes rather than relying on estimates of willingness to pay as was the economic dogma at the time. Furthermore, more consideration should be placed on the psychology of decision making.

In practice, libertarian paternalism has brought up the so-called *Nudge Theory*—also championed by Sunstein and Thaler. A nudge is described as *"any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid"*²⁵ The earlier given example of the cafeteria could be turned into a nudge by arranging the products as such that, for instance, the healthy product will be bought before the unhealthy products. There is already widespread research and practice in consumer research of the nudge theory. In planograms made for supermarkets it is well known that eye-level, the front of stores and the end of aisles are preferred locations in supermarkets as products placed here will be bought more frequently.²⁶ Other examples are the alternative labeling of calorie counts on American food products, as American consumers are more likely to read from left to right as opposed to right from left.²⁷ This in turn has led to research supporting the conclusion that the lateral offering of products influences the consumer to buy the product offered on the left rather than the right.²⁸ Marteau, Hollands and Fletcher categorize what can be called nudges into two categories: those that affect the environment of the agent and

²² Ibid, p.1164.

²³ Sunstein, Cass R., Christine Jolls, and Richard H. Thaler, 1998, "A Behavioral Approach to Law and Economics", *Chicago Unbound* 1471-1550.

²⁴ Roessler, Beate, and Dorota Mokrosinska, 2013.

²⁵ Thaler, Richard H., and Cass R. Sunstein, 2008, *Nudge: Improving Decision about Health, Wealth, and Happiness*, Yale University Press.

²⁶ Lewis, Graham, 2006, System and method for automatic placement of products within shelving areas using a planogram with two-dimensional sequencing, US Patent 11/429,328.

²⁷ Dallas, Steven K., and J. Liu Peggy, 2018, "Don't Count Calorie Labeling Out: Calorie Counts on the Left Side of Menu Items Lead to Lower Calorie Food Choices", *Journal of Consumer Psychology*.

²⁸ Romero, Marisabel, 2016, "Healthy-Left, Unhealthy-Right: Can Displaying Healthy Items to the Left (versus Right) of Unhealthy Items Nudge Healthier Choices?", *Journal of Consumer Research*.

those that affect the automatic associative processes.²⁹ Making it slightly more inconvenient by taking the elevator, increasing the effort it takes to order unhealthy foods in a cafeteria, reducing the proximity and density of retail outlets for unhealthy products, increasing the number of 'healthy' options rather than unhealthy options and taking into consideration product design are example of nudges that would fall into the first category. Priming and altering psychological associations (or creating new ones) would be considered in the second example. Example are the consumption of alcoholic beverages in media and corresponding advertisement and using positive or negative terms & images in association with certain products.

To summarize, libertarian paternalism is the theory that paternalistic decision making can be done without restricting the freedom of choice of the individual. This could be done in a variety of ways which would all be considered non-coercive. In practice, libertarian Paternalism relies on the *Nudge Theory*, wherein a nudge is defined as 'any de facto influence on choice that does not restrict the freedom of choice of the consumer and is easy and cheap to avoid.'

2.2 Autonomy Gaps

There are multiple critiques on libertarian paternalism. In this paper however only one of these critiques will be given special attention.³⁰ The theory of Autonomy gaps as given by Joel Anderson builds on earlier works in cooperation with Alex Honneth. Anderson states that there is a fundamental mismatch between institutionalized expectations and the ability of individuals to meet those expectations, and provides a convincing critique on, amongst other theories, libertarian paternalism.³¹

Anderson's critique stems from a critique on liberalism. One of liberalism's core commitments is protecting the autonomy of the individual, and liberal social justice presupposed the commitment to protect the vulnerable.³² To achieve full autonomy there must be set socially supportive conditions. This is a straightforward assumption: for a child to mature, they must undertake a journey they cannot possibly undertake on their own, even if they want to. At every step of development, they will inevitably stand in relationship with other individuals. To confuse 'liberalism' with 'individualism' is therefore a false equivocation. In this relationship with others, our autonomy is vulnerable to disruption. Only the mutual recognition of self-trust, self-respect and self-esteem can safeguard the full autonomy

²⁹ Marteau, Theresa M., Gareth J. Hollands, and Paul C. Flechter, 2012, "Changing Human Behavior to Prevent Disease: The Importance of Targeting Automatic Processes", *Sciencemag* 1492-1495.

³⁰ For further reading and more critiques, see Anderson, Joel, 2009, "Nudge: Improving Decisions about Health, Wealth, and Happiness, Richard H. Thaler and Cass R. Sunstein. Yale University Press, 2008. x + 293 pages. [Paperback edition, Penguin, 2009, 320 pages.]", *Economics and Philosophy* 369-406 and Dworkin, Gerald, 2020, "Paternalism", *The Stanford Encyclopedia of Philosophy*.

³¹ Anderson, Joel, 2009, "Autonomielücken als soziale Pathologie. Ideologiekritik jenseits des Paternalismus", *Sozialphilosophy und Kritik*, 433-453.

³² Anderson, Joel, 2017, "Vulnerability, Autonomy Gaps and Social Exclusion", In *Vulnerability, Autonomy, and Applied Ethics*, 49-68, New York: Routledge.

in a society.³³ Yet not every individual finds itself in a situation wherein their full autonomy can be safeguarded as such. This leaves them vulnerable to ‘fall behind’ in society, thus inequality follows.

Herein lies the critique on libertarian paternalism. In the concept bounded rationality, it is presumed that mankind is rational in decision making. However, according to libertarian paternalism this rationality is bound to the context of the architecture of that decision making. The critique as given above provides an alternative: It is simply unrealistic to expect individuals to make choices within their best interest because they are not all on the same level of autonomy. This is the fundamental basis of the concept of *Autonomy Gaps*. Anderson defines autonomy gaps as follows: “I use the term “autonomy gap” to label this mismatch between institutionalized expectations and individuals’ ability to meet those expectations.”³⁴ So when implementing a governing decision, like the social credit system, there will be expectation in that decision about the capabilities of the individual which are not realistic to presuppose. Anderson gives a number of examples.³⁵ Lottery winners are expected to handle the sudden acquisition of huge amounts of currency, yet it has been shown time after time that most individuals simply do not possess the capabilities to be responsible with this new found wealth. This is the reason why some lotteries assign counselors, either financial, psychological or both, to new winners to increase their capabilities to handle this new responsibility. Closely related to this are cases of lump-sum payments, where individuals receive sums of money and the individual can not cope with this responsibility. Another example is Judicial Elections, where voters are asked to decide on referendums or vote on huge lists of candidates for relatively minor positions. It is simply unrealistic to expect voters to make a well thought out decision on those elections as it is unrealistic for every citizen to be up to date on every topic and to do research on every one of those candidates. Another example is the privatization of pension savings for Swedish citizens in 2000. Two-thirds of the citizens opted for their own portfolio and choose from a list of 456 funds rather than the default, yet research has concluded that in the end this decision was worse off than the default option. It was simply unrealistic for the Swedish government to presume that every citizen could reasonably and effectively make such a decision, as not every one of those citizens can be informed on every aspect of that decision.

In conclusion, governance based on libertarian paternalism and the nudge theory might be an attractive form of governance, yet it can fail its goal or make matter worse on implementation if it leaves open autonomy gaps, as it presupposes capabilities of the users that the users do not possess.

³³ Honneth, Axel, n.d., *The Struggle for Recognition*, Translated by Joel Anderson, Cambridge, Massachusetts: The MIT Press.

³⁴ Anderson, Joel, 2017, p.49.

³⁵ The examples given in this chapter are taken from Anderson, Joel, 2017 and Anderson, Joel, 2019.

2.3 Libertarian Paternalism and the Social Credit System

The social credit system can be seen as a perfect example of nudge-theory. It is what Yeung describes as a *Hypernudge*: Big Data as a mode of regulation by design.³⁶ Yet a question remains as it stands in contrast with reality: are social credit systems inherently libertarian paternalistic? To be libertarian paternalistic it must be non-coercive and rule out no choices for the targeted user, and that is precisely one of the main critiques that can be made on the Chinese variants as the Joint Punishment System (another system like Skynet that is to be incorporated in the Chinese nation social credit system) dictates that users with a low credit score will be banned from a host of public services.³⁷ In this section it will be argued that coercion and restricting choice making are not inherent to any of the variants of the social credit systems.

The broad systems will inevitably have the most impact on the lives of its users, as they have the most far reaching scope of data gathering, analysis and storage. This is clearly demonstrated in the bounded consequences of the Chinese variants, as having a low credit score can result in, amongst other things, exclusion of public services, party membership, military service and much more.³⁸ Maintaining a high credit score on the other hand may result in a fast track to promotion, discounts on certain products and a fast track to better housing.³⁹ These positive reinforcers are in line with libertarian paternalism, the aforementioned negative reinforcers are not. By banning individuals with low scores from public transportation for instance, the freedom of choice of those individuals is obviously impeded. However, the use of governing reinforcers is not inherently part of a social credit system as it needs not to be used to be effective. The answer as to why it is not needed is plain simple: the reputation itself *is* a nudge. Just by making available a score about a character or characteristic which can be improved or deteriorated, decision will be influenced, be it where the individual is the subject or the object. That is the whole idea behind reputation systems in general, you make available a reputation of a party so one knows if the other party, for instance, can be trusted or not.⁴⁰ It does not need to be normative, only descriptive, as the users themselves will alter their behavior according to the scores of others, therefore making the score itself a nudge. This is clearly demonstrated in systems like the one found in Airbnb and Uber. There must be an acknowledgement here: the research about how a user is affected by its own reputation is still limited as reputation systems are a relatively new phenomenon. It is known however that the implementation of reputation systems in general will most definitely affect the choices of users in relationship with others and tend to be very effective at doing so.⁴¹

³⁶ Yeung, Karen, 2017, "'Hypernudge': Big Data as a mode of regulation by design", *Information, Communication & Society* 118-136.

³⁷ Creemer, Rogier, 2018.

³⁸ Ibid.

³⁹ Ibid.

⁴⁰ Jøssang, A. 2007.

⁴¹ Ibid.

The broad systems stand in sharp contrast to the small and cumulative systems, as the broad systems usually are too big of an undertaking to 'just' have a descriptive purpose. Why waste huge amounts of funds and capital to a system meant for governing, if it is not going to be used to actively govern? Yet, even broad systems do not need to be coercive or put any restrictions on the users. It is in the governing decisions made by the administrators of the social credit system in practice, and not the system itself, wherein coercion and restricting the choices of individuals lies. The Chinese variants for instance do not need to implement restrictions: It is a governing decision made by the administrators of the Chinese system to implement those restrictions in the system. All the Chinese variants, and the eventual national Chinese system, would still be labeled social credit systems if they left those restrictions out.

To summarize: the systems itself are libertarian paternalistic, it is the administrators/governors which can introduce coercion and restriction into the rules of the system, making some of the variants of social credit systems non-libertarian paternalistic.

2.4 Autonomy Gaps and the Social Credit System

As social credit systems are choice-architecture influencing platforms, they might be vulnerable to autonomy gaps. To be vulnerable to autonomy gaps, the system must presume expectations on the part of individuals when it is unrealistic for individuals to have those capabilities. In this part there will be argued that all variants of the social credit system will be vulnerable to autonomy gaps to a certain degree. However, the broad systems will be especially vulnerable to autonomy gaps.

Consider the following: Let there be a small global social credit system on the Eco-friendliness of corporations. A question could be raised on its implementation: Is it fair to use the same standards of evaluation for a company like Apple or Google and a small family business in a developing country? Could it not be said that large corporations with a lot of financial assets have an easier time adhering to presupposed standards of Eco-friendliness in such systems? It should not be controversial to state that a lot of companies, even in non-development countries, simply will not have the resources available to be as eco-friendly as the next one. This could be a direct example of an autonomy gap: there is a fundamental mismatch between the presupposed capabilities of the individual companies and the institutionalized expectations in the governing decision making.

In the example given, the value given to the score is determined by the rules of the system as done by the system administrators. It therefore seems that the expectation of the users is not predetermined by social credit systems itself, but by the usage of those systems as stated in chapter 2.3. Yet, even though the value of the score, and the real world consequences bound to this score, are determined by the rules of the system as given by the administrators, there is the inherent expectation that the users will adhere to those rules in some form or another. It must either have that character or character trait, or not. I.e., it must comply with the system. There can be raised doubts on the capabilities of the users to adhere to this requirement of social credit systems. This doubt can be

interpreted in different ways. It can be taken at face value: Do the users have the technological resources available to keep up with the system? Do the administrators themselves have the capabilities to effectively maintain and moderate such a system effectively? Not every part of the world is on the same level of technological development. This doubt can also be interpreted in a more abstract sense: Can users keep up with the requirements of the system? For instance, studies done on the implementation of electronic devices to monitor performance can not only change the behavior of the ones observed, it can also cause psychological stress, decreased satisfaction and commitment to the organization.⁴² Or what if the system requires the timely handing in of certain forms and paper? The system might require users to rate a set number of other users on a timely basis. It might require the user to log into the system on a timely basis. All these things can put extra responsibilities and psychological stress on the individual, and not everyone is on the same level of developed autonomy to cope with this added pressure. So here there must be made a difference between the broad, small, and cumulative systems. Broad systems will, in general, suffer far more from autonomy gaps than small and cumulative systems, as the requirement of intense data gathering, storage and analysis puts more stress on the users than the two alternatives. That is not to say that the small and cumulative systems are entirely devoid of the premise of 'keeping up with the system': they just require a lot less from the users as they require relatively little data to be able to function.

⁴² Tomczak, David L., Lauren A. Lanzo, and Herman Aguinis, 2018, "Evidence-based recommendations for employee performance monitoring", *Business Horizons* 251-259.

Conclusion

The aim of this paper is to answer the question “Are social credit systems compatible with libertarian paternalism?”. The answer to this question is yes, to an extent. Social credit systems can be used to help govern a large body of the population without being coercive, nor taking away freedom of choice. It is in the process of binding consequences to a social credit system wherein a social credit system could divert from the core values of libertarian paternalism. Social credit systems can suffer from autonomy gaps if too much pressure and expectation is put on the users. However, not every variant of social credit system will suffer from autonomy gaps in the same degree as one system can differ significantly from another system in a lot of different ways. This can be seen in the broad systems versus the cumulative and small systems. The former can leave open a lot of autonomy gaps due to its presupposed scale and intensity, while the latter have the potential to leave open no autonomy gaps as they require relatively little data and intensity to be able to function.

Though social credit systems can offer an enticing and modern form of governing which fully embraces technology, there still is a lot of uncertainty and room for future research. What are the psychological consequences of binding a reputation to a person? Will it promote equality, or will it promote inequality? In how far will the behavior of a person change when that person has a bound and public reputation? Though research on this subject in context with reputation systems is promising, it is still very limited. In how far should we incorporate such systems into our lives? There are some serious ethical concerns too: are some areas of our social lives too private? Not a lot of people would have trouble building such a system for the purpose of promoting eco-friendliness, sustainability, or political integrity, but could the same be said for promoting sexual performance? What about intelligence, attractiveness, wealth, social standing, or racism/sexism? Though there are still a lot of uncertainties and dilemmas, it is in my belief that social credit systems will become an invaluable asset to governing in the modern, technologic era.

Bibliography

Airbnb. n.d. <https://www.airbnb.com/>.

Anderson, Joel. 2009. "Nudge: Improving Decisions about Health, Wealth, and Happiness, Richard H. Thaler and Cass R. Sunstein. Yale University Press, 2008. x + 293 pages. [Paperback edition, Penguin, 2009, 320 pages]." *Economics and Philosophy* 369-406.

—. 2009. "Autonomielücken als soziale Pathologie. Ideologiekritik jenseits des Paternalismus." *Sozialphilosophy und Kritik.*, 433-453.

Anderson, Joel. 2017. "Vulnerability, Autonomy Gaps and Social Exclusion." In *Vulnerability, Autonomy, and Applied Ethics*, 49-68. New York: Routledge.

Anderson, Joel, and Axel Honneth. 2005. "Autonomy, Vulnerability, Recognition, and Justice." In *Autonomy and the Challenges to Liberalism*, 127-149. Cambridge University Press.

Central Government of China. 2014. "Notice of the State Council on Printing and Distributing the Outline of the Construction of the Social Credit System (2014-2020)." http://www.gov.cn/zhengce/content/2014-06/27/content_8913.htm.

Credit Sesamy, Inc. n.d. <https://www.creditsesame.com/>.

Creemers, Rogier. 2018. *China's Social Credit System: An Evolving Practice of Control*. Leiden: Social Science Research Network, May 22. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175792.

Dallas, Steven K., and J. Liu Peggy. 2018. "Don't Count Calorie Labeling Out: Calorie Counts on the Left Side of Menu Items Lead to Lower Calorie Food Choices." *Journal of Consumer Psychology*.

Devereaux, Abigail. 2018. "The nudge wars: A modern socialist calculation debate." *The Review of Austrian Economics*.

Dworkin, Gerald. 2020. "Paternalism." *The Stanford Encyclopedia of Philosophy*.

Facebook. n.d. *How do ratings work on Marketplace?* https://www.facebook.com/help/915385548593204?helpref=about_content.

Foucault, Michel. n.d. *Discipline & Punish: The Birth of the Prison*. Translated by A. Sheridan. Vintage Books.

Grimmelikhuijsen, Stephan. 2009. "Do Transparent Government Agencies Strengthen Trust?" *Information Polity: The International Journal of Government & Democracy in the Information Age* 173-186.

Honneth, Axel. n.d. *The Struggle for Recognition*. Translated by Joel Anderson. Cambridge, Massachusetts: The MIT Press.

- huifeng, he. 2019. *China's social credit system shows its teeth, banning millions from taking flights, trains.* South China Morning Post, 18 Februari. <https://www.scmp.com/economy/china-economy/article/2186606/chinas-social-credit-system-shows-its-teeth-banning-millions>.
- Hursthouse, Rosalind, and Glen Pettigrove. 2018. "Virtue Ethics." *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/ethics-virtue/#ObjVirtEthi>.
- Jøysang, A. 2007. "Trust and Reputation Systems." *Foundations of Security Analysis and Design IV*.
- Kim, Sung Ho. 2017. *Max Weber*. Website. Edited by Edward N. Zalta. Prod. The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=weber>.
- Lewis, Graham. 2006. System and method for automatic placement of products within shelving areas using a planogram with two-dimensional sequencing. US Patent 11/429,328.
- Marteau, Theresa M., Gareth J. Hollands, and Paul C. Flechter. 2012. "Changing Human Behavior to Prevent Disease: The Importance of Targeting Automatic Processes." *Sciencemag* 1492-1495.
- Orwell, George. 1977. *1984*. New York: Signet Classic.
- Roessler, Beate, and Dorota Mokrosinska. 2013. *Privacy and Social Interaction*. Amsterdam: Sage Publications.
- Romero, Marisabel. 2016. "Healthy-Left, Unhealthy-Right: Can Displaying Healthy Items to the Left (versus Right) of Unhealthy Items Nudge Healthier Choices?" *Journal of Consumer Research*.
- Schafer-landau, Russ. 2013. *Ethical Theory An Anthology*. 2nd. Blackwell Publishers.
- Shen, Xinmei. 2018. "'Skynet', China's massive video surveillance network." *Southern China Morning Post*.
- Shiv, Baba, and Alexander Fedorikhin. 1999. "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making." *Journal of Consumer Research*.
- Snyder, M., E. D. Tanke, and E. Berscheid. 1977. "Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes." *Journal of Personality and Social Psychology* 656-666.
- Sunstein, C., Lucia Reisch, and Micha Kaiser. 2018. "Trusting Nudges? Lessons from an international survey." *Journal of European Public Policy* 1-27.
- Sunstein, Cass R., and Lucia A. Reisch. 2018. "Behavioral Economics and Public Opinion." *Intereconomics* 5-7.
- Sunstein, Cass R., and Richard H. Thaler. 2003. "Libertarian Paternalism Is Not an Oxymoron." *The University of Chicago Law Review* 1159-1202.

- Sunstein, Cass R., Christine Jolls, and Richard H. Thaler. 1998. "A Behavioral Approach to Law and Economics." *Chicago Unbound* 1471-1550.
- Talisse, Robert B. 2016. *Engaging Political Philosophy*. New York: Routledge.
- Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decision about Health, Wealth, and Happiness*. Yale University Press.
- Tomczak, David L., Lauren A. Lanzo, and Herman Aguinis. 2018. "Evidence-based recommendations for employee performance monitoring." *Business Horizons* 251-259.
- Yeung, Karen. 2017. "'Hypernudge': Big Data as a mode of regulation by design." *Information, Communication & Society* 118-136.
- Zwass, Vladimir. 2017. *Information System*. Website. Encyclopædia Britannica, inc., 28 December. <https://www.britannica.com/topic/information-system>.