Power, Paternalism and Libertarianism

Libertarian Paternalism – More than a Nudge

Martin Jacobson
Magister level thesis
Fall 2017
Supervisor: Andrew Reisner
1. Introduction ................................................................. 3
2. Libertarianism ............................................................. 4
3. Paternalism ................................................................. 10
4. Heuristics, Biases & Nudging ......................................... 14
5. Analysis ........................................................................ 18
   5.1 Formulating notions in terms of power ....................... 19
   5.2 Are there cases of libertarian paternalism? ................. 22
   5.3 Additional consequences ......................................... 24
6. Summary and Conclusions ............................................. 26
7. References ................................................................. 28
8. Appendix ...................................................................... 30
I. Introduction

It is commonly believed that paternalism is at odds with libertarianism (Boaz 1997, 211). Recent literature has suggested that there are forms of paternalism which are acceptable to libertarians: namely “nudging”, sometimes even referred to as “libertarian paternalism” (Thaler & Sunstein 2008, 4-6). The purpose of this paper is twofold. First, to investigate the taxonomical question of how libertarianism, paternalism and nudging relate to each other. Secondly, to investigate whether, and if so when, paternalism is compatible with libertarianism. I argue that any action which is not coercive is compatible with libertarianism. Thus, any non-coercive paternalist action is compatible with libertarianism. I also argue that there are several paternalist action types, such as nudges, informing and incentivizing, which are not coercive, and thus are compatible with libertarianism.

As a first step in my analysis I formulate libertarianism, paternalism, libertarian paternalism and nudging in a common terminology in terms of methods of, and motives behind, power. I argue that the methods and motives of power are orthogonal, meaning that no method necessitates any particular motive, and vice versa. According to my taxonomy, whether or not an exercise of power is paternalist depends on the motive, while whether or not it is a nudge depends on the method. Libertarianism is the view that one method, namely coercion, may only legitimately be used for one motive, namely defense. Any action which is not coercive is thus compatible with libertarianism. Libertarian paternalism consequentially includes all non-coercive power exercises which are paternalistically motivated.

As a second step in my analysis I discuss whether there are any actions which satisfy the criteria of libertarian paternalism. I argue that nudging, informing, and incentivizing are all methods of power which can be paternalist without being coercive. They are consequentially cases of libertarian paternalism. I therefore draw the conclusion that paternalism is compatible with libertarianism, and that it is compatible iff the method used to exercise the paternalist action is not coercive.

I conclude my analysis by briefly mentioning three interesting implications of my analysis which fall outside the main aim of the paper. First, it indicates that all asymmetrical paternalism is compatible with libertarianism. All libertarian paternalism is however not asymmetrical. Secondly, it follows from my analysis that weak paternalism and strong paternalism are equally compatible with libertarianism. It also indicates that “usage” and paternalism are equally compatible with libertarianism.
The paper is divided into six parts, of which this introduction is the first. In part II I describe more substantially the philosophy of libertarianism. In section III I describe paternalism. In section IV I explain nudging and the surrounding theory of heuristics and biases. Section V provides the central analysis of the paper. In section VI I summarize the paper and its conclusions.

Before I conclude this introduction I wish to make two brief points about the terminology and scope of this paper. When using the word “paternalism” I do not restrict myself to paternalist actions performed by a state. Rather, I include any action which can be considered paternalist regardless of who performs it. In technical terms I discuss “broad” and not only “narrow” paternalism (Dworkin 2017). Similarly, I only discuss whether paternalist actions themselves are compatible with libertarianism, regardless of who performs them. Consequentially the discussion of how libertarianism relates to states and state action falls outside the scope of this paper.

Secondly, I only discuss how libertarianism and paternalism relate regarding mentally sound adult human beings. There may be additional considerations regarding both libertarianism and paternalism in cases involving non-human animals, human children or cognitively impaired human beings, i.e. in cases where there is a strong asymmetry between the paternalist and her target. However, these possible additional reasons fall outside of the scope of this paper.

II. Libertarianism

In this section I provide an overview of the most central commitments of libertarianism. I take libertarianism to be a ‘broad church’ including a multitude of different views. The smallest common denominator, in virtue of which all of these views are libertarian, is that they accept the non-aggression principle, according to which coercion may only be used in defense.¹ As the aim of this paper is to analyze whether paternalism is compatible with libertarianism as a general view, I focus my discussion of libertarianism to this smallest common denominator of all libertarian views. The bulk of this section is consequentially devoted to analyzing the concept of coercion. I argue that \( P \) coerces \( Q \) iff \( P \) uses violence or the threat of violence to

¹ Some would claim that self-ownership (Vallentyne & van der Vossen 2014), or natural rights (Nozick 1974, ix), is the ‘true foundation’ of libertarianism. I have argued elsewhere (Jacobson 2016, 17-8) that these three are co-extensional, but that the non-aggression principle is preferable to the others since it is more explanatory than the others.
exercise power over Q. P is violent towards Q iff P 1) causes Q physical harm or pain, 2) physically restrains or enforces Q’s actions, or 3) destroys, intrudes on or otherwise deprives Q of Q’s legitimate property. I conclude the section by briefly discussing how libertarianism might develop beyond its minimal central commitments.

Much of the contemporary debate concerning coercion relies on Robert Nozick’s essay “Coercion”. Nozick argues that P coerces Q iff P threatens to bring about some to Q disagreeable consequence C unless Q performs action A, Q performs A, and part of Q’s reason for doing A is to avoid that P brings about C (Nozick 1969, 441–5). The notion of coercion which Nozick offers is thus based on the notion of a threat. Interestingly, the structure of a threat (P threatens Q iff P claims that P will bring about the bad consequences C iff Q does (not) A) perfectly mirrors that of an offer (P makes an offer to Q iff P claims that P will bring about the good consequences C iff Q does (not) A). Whether one’s statement is an offer or a threat therefore depends on whether the consequences one claims to bring about are good or bad (Anderson 2017). In what sense should these consequences be considered good or bad?

In order to settle this question, we can use some baseline compared to which a state of affairs could be considered good or bad. This baseline could either be moralized, i.e. compared to what we believe ought to be the case, or non-moralized i.e. compared to what actually is the case. Since libertarians argue that justice should be understood in terms of coercion it might be problematic to rely on a moralized baseline, as it risks making the analysis circular. Furthermore, I wish to make an inclusive account of libertarianism which is independent of any prior moral commitments. I therefore do not discuss any moralized baselines.

The baseline must consequentially be non-moralized. Examples of non-moralized baselines include the state of affairs which is normal, or the state of affairs which is expected. However, used as baselines these face difficulties raised by Nozick’s “slave case”: Consider a slave owner who regularly beats his slave. One day the slave owner proposes to spare the slave from her regular beating if the slave does (not) A (Nozick 1969, 450). According to the non-moralized baseline this statement is an offer, and not a threat. If the slave owner were to beat the slave, the slave owner would not diverge from normality or expectations. The slave owner would thus, according to these baselines, not be acting coercively. This, however, seems obviously incorrect. Therefore, these baselines must be rejected.

Another possibility would be to rely on the preferences of the subject as our baseline (Gorr 1986, 388). The slave obviously does not want to be beaten. The beating is therefore worse
than, and inconsistent with, her preferences. However, this baseline seems too inclusive. Consider an employer telling her employee that if she shirks from doing her job she is going to get fired. Being fired is contrary to the preferences of the employee. However, the employer’s demands are simply part of their contract, and should thus not be considered coercive. This analysis is therefore too inclusive and must also be rejected. (Anderson 2017)

Another issue with this case is that the employer seems to have benefitted her employee, all things considered. The employee would prefer a scenario in which she is first hired but later laid off, to a scenario in which she was not hired in the first place. Such a counterfactual scenario, in which the two parties would never have met or interacted at all, could also be used as a baseline. In such a counterfactual world the slave could not have been beaten by the slave owner, thus making the world where she is being beaten worse than the baseline. However, this analysis runs into problems when one actor both benefits and harms another. Consider P who gives Q a gift, but later steals it back. Q has not moved from the baseline of how she would have been if she had not gotten the gift in the first place. We would however say that P has wronged Q or made Q worse off in some sense. In a more extreme case, a doctor who has saved a person’s life is not allowed to arbitrarily kill the person later. I therefore conclude that this baseline should also be rejected.

The problems stated above have led some theorists to deny that there is any major difference between an offer and a threat, thus arguing that there are coercive offers. Some have suggested that P’s suggestion to Q is coercive iff P is actively preventing Q from getting a pre-suggestion bargaining position which Q would strongly prefer (Zimmerman 1981, 133). Assume that Q is marooned on an uninhabited island. P has destroyed all vessels which Q could have used to get away from this island, but offers to transport Q from the island in exchange for an exorbitant amount of money. P’s suggestion is an offer rather than a threat, as the suggestion provides Q with an opportunity which Q would not otherwise have had. But by destroying Q’s vessels, P has prevented Q from getting a bargaining position which Q would have strongly preferred. The offer which P make to Q is thus coercive. However, it seems that the coercive element in this situation lies in P’s destruction of Q’s vessels and not in the subsequent offer P makes. If P were to destroy Q’s vessels without making any offer she would still be acting coercively, making the utterance of a suggestion redundant to the issue of whether the situation is coercive. (Anderson 2017)
Let us instead consider a case where $P$ prevents $Q$ from improving her bargaining position in a way which is not obviously coercive. Imagine that $P$ knows that $Q$ has a gambling addiction. $P$ uses this to lure $Q$ into making very unfavorable bets. Eventually $Q$ ends up owing $P$ a large sum of money. $P$ suggests writing off the debt if $Q$ performs some disagreeable task which $P$ wants done. According to the analysis this would be coercive. I do however rather believe that this case should be described as exploitative, and while exploitation may be disagreeable, it does not follow that it is coercive. I therefore believe that we ought to reject this account of coercion.

Some writers have attempted to analyze coercion in terms of the effect which the suggestion has on the recipient’s will. Some suggestions do not only affect the recipient’s choices, but also her ability to make a choice at all. One could argue that the fear which a threat instILLS in the recipient can make her unable to act upon her rationally considered desires, thereby making the suggestion coercive (Frankfurt 1988, 78). A similar analysis is that a proposal could be considered coercive if it creates a sufficient pressure on the recipients will to make the choice unfree (Feinberg 1986, 198). A problem with these recipient-based accounts is that they also seem to include offers that are too good to refuse. If $P$ makes $Q$ an offer $Q$ considers too good to be true this might make $Q$ too ecstatic to use her rationally considered desires, or cause pressure on $Q$’s will to a sufficient degree to make it unfree. We would however not agree that these too-good-to-be-true proposals are coercive. (Anderson 2017)

These analyses of coercion share the assumption that coercion essentially is some sort of pressure on the will of the coerced person. There is however a problem with this assumption. Assume that the children $P$ and $Q$ are playing in a sandbox. However, $P$ wants the sandbox for herself. $P$ could threaten $Q$ and say that unless $Q$ leaves the sandbox she will get violent or call for her protective parent. $P$ could however also simply push $Q$ until $Q$ fall out of the sandbox. As $Q$’s falling out of the sandbox is enforced, it does not in any way rely on, or create any pressure on $Q$’s will. It would nevertheless be coercive. Similarly, there are threats which pressure the will of their targets which are not coercive. If $P$ claims that $P$ will not trade with $Q$ under certain circumstances $P$ is threatening $Q$ with a boycott. We would however not say that $P$ is coercing $Q$. At least not in the sense of coercion which most libertarians are interested in.

These kinds of arguments have convinced some philosophers that the essential feature of coercion is not the effects it has on the will of the target, but rather that it forces its target to
behave in some way (Andersson 2010, 6-8). In this text I adopt such an enforcement view of coercion. Spelled out in more detail, I assume that \( P \) coerces \( Q \) iff \( P \) uses violence or the threat of violence against \( Q \). \( P \) is violent towards \( Q \) iff 1) causes \( Q \) physical harm or pain, 2) physically restrains or enforces \( Q \)’s actions, or 3) destroys, intrudes on or otherwise deprives \( Q \) of \( Q \)’s legitimate property.

I believe that this analysis gets all the above mentioned cases right. The slave owner who offers not to beat her slave (if she does (not) \( A \)) is acting coercively, as she thereby also threatens to cause the slave physical pain if she does not comply. The fact that the violence is expected or normal is irrelevant to whether it is coercive. Reclaiming a gift would be coercive as it would deprive someone of her legitimate property. Finally, it also explains why pushing someone is a form of coercion without affecting their will, and why a threat of a boycott can cause pressure on someone’s will without being coercive.

This concludes my analysis of the central commitments libertarianism. These central commitments can be supplemented and combined with several additional principles, creating a ‘broad church’ of different libertarian positions. While these additional principles fall outside the scope of this paper I briefly describe some of them in order to give the reader a more complete understanding of libertarianism.

First, my analysis of coercion relies on the notion of legitimate property. In order to make libertarianism fully applicable we therefore also need some notion of what makes a distribution of property legitimate. It follows from the non-aggression principle that once property is owned it can be transferred from one owner to another, as long as this transfer is not coercive. In other words, gifts and trades are allowed, while theft and extortion are not. However, the non-aggression principle gives no guidance as to how a good became owned in the first place. Libertarianism consequentially relies on some principle of initial acquisition of resources. Several such principles of rightful initial acquisition of resources have been suggested, dividing libertarians into “right-libertarians”, who argue that the world is originally un-owned, and “left-libertarians”, who argue that the world is initially owned in some egalitarian manner. Right-libertarians consequentially assume that non-privately owned
resources may be unilaterally and un-proportionally privately appropriated, while left-libertarians deny this. (Vallentyne and van der Vossen 2014)²

A second distinction can be made between ‘conservative’ and ‘liberal’ libertarians. Liberal libertarians embrace social values generally considered left-wing or liberal, while conservative libertarians embrace cultural conservative values. While they agree in adopting the non-aggression principle, and an aversion against state coercion, they disagree in their views of civil society beyond the state. (Block 2015, 34-5)

Another important issue is whether libertarianism is part of or exhausts morality. A libertarian could argue that there is a duty to e.g. visit your sick grandmother at the hospital, but deny that this duty is enforceable with coercion. However, whether such non-enforceable duties exist, and if they exist, what they are, is a matter of controversy among libertarians. (Vallentyne & Steiner 2000, 4-5) Libertarians also diverge on the question of why one ought to accept the non-aggression principle. For example, a utilitarian could embrace it on the grounds that it is part of an optimal utilitarian codex (Compare with John Stuart Mill’s “harm principle” 2011, 25-8). A Kantian could on the other hand argue that aggressors treat people merely as means, and fail to treat them as ends in themselves (Nozick 1974, 32-4. Compare with Kant 2009, 52 Ak 4:429). Others argue that the non-aggression principle is a self-evident axiom which does not need any further justification by external principles (Rothbard 1978, 215).

This concludes my overview of libertarianism. I want to emphasize two things with this section. First, that libertarianism is a ‘broad church’, which is unified by a distinct smallest common denominator in the non-aggression principle. Secondly, that coercion should be understood as enforcement when read in the context of the non-aggression principle. I believe that the non-aggression principle is inclusive enough to include all views properly called libertarian. I also believe that it cannot be made any more inclusive without also including views which should not be considered libertarian. The arguments I make in this text only relate to this smallest common denominator. The virtue of this approach is that it makes the arguments applicable to libertarianism as a general view.

III. Paternalism

In this section I briefly outline paternalism. I begin by analyzing paternalism, and argue that an action is paternalist iff it is an exercise of power over a person which is intended to benefit the person. I add more substance to this account by describing five key assumptions which are tacit in paternalism. Thereafter I broaden the presentation by also discussing how paternalists may differ from each another, how paternalism is usually practiced, and by mentioning some common arguments raised against paternalism.

Paternalism can pre-theoretically be understood as an interference with a person which is intended to benefit the person. However, not all beneficial interference is paternalistic. For example, if $P$ gives $Q$ a gift this may benefit $Q$. But while this would be benevolent it would not be paternalistic.

We therefore need some additional criteria which distinguish paternalism from mere benevolence. Several such criterions have been suggested. One possible criterion is that paternalism infringes on the person’s liberty or rightful sphere of autonomy. A problem with these accounts is that they are normative and thus hinge on some normative idea of legitimate liberty or rightful sphere of autonomy (Grill 2012). Assume that it turns out that moral nihilism is true. It would then be false that we have any rightful sphere of autonomy. Consequentially, there would be no such thing as paternalism, if these analyses of paternalism are correct. But this strikes me as very counterintuitive. I therefore believe that we ought to reject these criteria.

Another suggested criterion is that the interference is unwelcome, or contrary to the person’s will (Dworkin 2017). I do however believe that this criterion is too strict. Consider the child $Q$ who does not like vegetables. I order to make $Q$ eat her vegetables, the parent $P$ offers $Q$ a reward, provided that she eats her vegetables. Consider also $R$ who wants to stop smoking, but fails to do so. In order to help $R$, $S$ hides $R$’s cigarettes. While these actions seem to be paternalistic, they do not seem to be unwelcome or contrary to $Q$’s or $R$’s wills. (And neither does the reward infringe on $Q$’s liberty or rightful sphere of autonomy.)

What then is the relevant difference between $P$ giving $Q$ a gift, and $P$ giving $Q$ a reward if $Q$ eats her vegetables? I believe that the relevant difference between the cases is that the latter is an exercise of power, while the former is not. $P$ exercises power over $Q$ when $P$ causes $Q$ to perform an action or behavior $A$ that $Q$ would otherwise not have done. In this text I therefore
assume that an action is paternalist iff it is an exercise of power over an individual which is intended to benefit the individual.

While paternalism is often benevolent, it differs from “mere benevolence”, as the benefit which $Q$ receives depends on $Q$’s behavior, and not only on some compensation $Q$ receives externally. Paternalism also differs from “benevolent usage”. $P$ uses $Q$ when $P$ makes $Q$ behave in a way which benefits some third-party $R$. For example, $P$ can convince $Q$ to give more money to charity, thereby benefitting $R$. As $P$’s motive in this case is not to benefit the target of the power exercise, $Q$, the action is not paternalist.\(^3\)

We can add more substance to this account of paternalism by considering five assumptions which are tacit in paternalism. The first assumption of paternalism is that our interests are not uniform, but rather that we have a plurality of interests. One example of such different interests would be long-term interests and short-term interests. Other examples include values versus mere preferences, fundamental versus derived interests, and real-life interests versus those interests one would have if one were fully rational and informed. In order to illustrate the tacit assumptions of paternalism, I use the example of a short-term interest in eating cake and a long-term interest in staying healthy.

The second assumption is that there can be conflict between these interests. This seems obvious in the case of the short-term interest in eating cake and the long-term interest in staying healthy. By succumbing to the short-term interest in eating cake, at least repeatedly, the long-term interest in staying healthy will be undermined.

The third assumption is hierarchy, i.e. that one of these interests ought to be given priority when conflicts arise. Assume that $Q$’s overall interest is better satisfied by staying healthy than by eating cake. In such a case the long-term interest in health will have a hierarchically superior position compared to the short-term interest in eating cake. The interests which ought to be given priority compared to all others, and hence is at the top of the hierarchy, I call the “prime interests”.

The fourth assumption of paternalism is that people sometimes fail to realize or maximize their prime interest. This is also intuitively plausible. If $Q$ is subject to \textit{akrasia}, she may fail to

\(^3\) Richard Thaler and Cass Sunstein call nudges used to benefit a third party “libertarian benevolence” (2003, 1185).
act in accordance with her prime interest in staying healthy, and instead succumb to her short-term interest in eating cake.

The fifth and final assumption is that an external actor $P$ may affect $Q$’s behavior in such a way that it better satisfies $Q$’s prime interest. This fifth assumption also seems reasonable. For example, $P$ can hinder $Q$ from succumbing to akrasia by denying her any cake. By doing this $P$ makes $Q$ act in a way which better satisfies $Q$’s prime interest.

To summarize my description: according to paternalism people have a multitude of interests, some of which are more important than others. These may come into conflict, where the person fails to act in accordance with her prime interest. In these cases it is possible for an external agent to affect the person’s behavior in such a way that it better suits her prime interest.

Since the objective of paternalism is to enhance our prime interest, the nature of our prime interest is imperative to paternalism. How do they, for example, relate to our self-perceived, explicit interests? A distinction is sometimes made between weak and strong forms paternalism. Paternalism is weak when it only interferes with the means which an agent uses in order to reach her explicit ends. Paternalism is strong when it not only interferes with the means an agent uses to reach her ends, but also with the explicit ends themselves. Strong paternalists thus assume that people can be mistaken about their prime interests. For example, consider $Q$, who prefers, and consequently consumes, an unhealthy but tasty diet. In this case weak paternalism does not warrant that a paternalist interferes with $Q$ to make her change to a healthy yucky diet. A strong paternalist could however claim that $Q$ is mistaken about her ends, and that we consequentially may interfere with her diet against her explicit preferences. Strong paternalism is logically stronger than weak paternalism, as strong paternalism implies weak paternalism while weak paternalism does not imply strong paternalism.\(^4\) (Dworkin 2017)

\(^4\) A related distinction is that between forms of paternalism which are objectivist and subjectivist about the good life. Objectivist paternalists believe that there is a certain type of life which is optimal for all persons, independently of our attitudes. Our prime interest is consequentially to approximate this optimal way of living, regardless of what our explicit interests are. Subjectivist paternalists reject the objectivist claim that there is an objectively optimal way of living. Rather they believe that what is good for an individual depends on, or is chosen by, the individual herself. (Grill 2012). Note however that the two distinctions do not perfectly converge. For example, the position that our prime interests are those interests we would have had if we were fully rational and informed is subjectivist. But since we can be mistaken about what we would want if we were fully rational and informed, it is also compatible with strong paternalism. Similarly, an objectivist could claim that the good life must be autonomous, and reject strong paternalism for that reason.
Paternalism can be practiced with several different methods. One common method is coercion. For example, a paternalist $P$ can forbid $Q$ from eating cake, threatening with coercive sanctions if $Q$ does so. $P$ can also simply seize $Q$’s cake, and thereby hinder $Q$ from eating it. Another method is economic incentivizing. Incentives can either be positive, providing some additional reward if the person acts in a desired way, or negative, withdrawing or denying some reward unless the person acts in a desired way. For example, consider a shopkeeper who raises the price of tasty unhealthy food, and lowers the price on healthy yucky food. She thereby makes her customers buy more of the healthy food and less of the unhealthy, and thus promotes their prime interest in health. A third common method of paternalism is informing or convincing. For example, by informing $Q$ about the negative long-term effects cake eating has on our health $P$ can make $Q$ eat less cake.

Several arguments have been raised against paternalism. John Stuart Mill famously argued against coercive paternalism on the grounds that it causes uniformity, and thus hampers individual and social development (Mill 2011, 73-4). Another objection is that paternalism does not respect the autonomy of its target (Quong 2010, 97-100). Further, paternalism is often perceived as patronizing, as it relies on the premise that the paternalist is better informed, or more rational than the target (ibid., 100-6). Many libertarians agree in these critiques, and argue that coercive paternalism is offensive towards its target (Boaz 1997, 211). Strong paternalism may be considered even more intrusive than weak paternalism, as the strong paternalist not only interferes with the means the target uses to reach her ends, but also with the ends themselves.

Another objection is that of symmetric paternalism. People find themselves in very different situations. Consequentially, a paternalist action which promotes the prime interest of one person may harm the prime interest of another. Paternalism can therefore, at least partially, be self-defeating when applied symmetrically. Consider for example $Q$ who is highly allergic to healthy food. If the paternalist $P$ established a general tax on unhealthy food, this would benefit the majority of people, whose consumption pattern thus changes towards a healthier diet. However, $Q$ is not benefitted as she must either eat the healthy food which she is allergic to, or pay a higher fee for her usual diet. There has been an exercise of power over $Q$, but it has not benefitted her. This could be due to the fact that $P$ was unaware of $Q$’s condition, making the tax a form of intended but failed paternalism. $P$ could also have known about $Q$’s condition but decided to implement a general fee anyways, as this would promote the health of the majority. In this case the action, vis-á-vis $Q$, is not paternalist at all, but rather a
“collateral damage” in a general paternalist action. This can be contrasted with asymmetrical paternalism, which benefits those who are not following their prime interest, without harming those who already are acting in accordance with their prime interest (Camerer et al. 2003, 1219). One example of such asymmetrical paternalism is nudging.

IV. Heuristics, Biases & Nudging

In this section I discuss and describe nudging. The background to nudging can be found in behavioral economics and psychological research on heuristics and cognitive biases. I consequentially devote much of this section to describing these theories. Thereafter I turn my attention to nudging itself, and discuss why it might be attractive to paternalists.

A heuristic technique is a decision making process or problem solving technique that aims to quickly give a satisfactory, but not always optimal result, comparable to a ‘mental short cut’. The cost of the decision making in terms of cognitive capacity and time is thereby minimized. Examples of such heuristics would be rules of thumb, going by ‘gut feeling’, or using stereotypes. A cognitive bias is a systematic error in, or systematic deviation from rational thinking. One common form of bias is confirmation bias, a tendency for us to look for information which confirms our currently held views, and reject information which challenge our currently held views. Consequentially we sometimes erroneously hold on to unjustified views while avoiding more well-founded views. These features of human psychology challenge the traditional assumption in economics that people act rationally and calculating to maximize their preference satisfaction, a caricature sometimes called the “homo economicus” or “econs” (Thaler & Sunstein 2008, 6-8). Instead it has given rise to its own field in economics, namely behavioral economics, which study how these deviations from rational thought affect the economy and economic decision making.

I explain heuristics and biases by using the dualist psychological model popularized by Daniel Kahneman. According to Kahneman human decision making and problem solving fundamentally consists of two different systems, which he simply calls System 1 and System 2. System 1 is intuitive and immediate. It works automatically with little effort and without any conscious control. System 2, on the other hand, works slowly and with much cognitive effort. It demands a certain level of concentration, which also generates a feeling of conscious control and command. But as soon as the concentration ceases, so does the activity. Since it is
subject to our conscious control, we tend to identify ourselves with System 2 rather than System 1. (Kahneman 2013, 27-30)

An example of a System 1 action would be to decode a simple sentence in one’s native language; such as “the sky is blue”. It takes little to no effort and it is impossible not to decode it once it has caught one’s eye. Other types of System 1 activities include perceiving annoyance in someone’s tone of voice, or making a judgment of a stranger based on stereotypes. An example of a System 2 activity would be to decode a simple sentence in a language which one has moderate knowledge of but does not master, such as “het gras is groen”. A native English speaker may not immediately realize what the Dutch phrase means, but with some consideration she would probably correctly calculate that it meant “the grass is green”. Other typical system 2 activities include complex logical or mathematical reasoning and actively searching for something. (ibid., 28-9)

When faced with a problem System 1 suggests an answer automatically and effortlessly. System 2, on the other hand will remain still unless consciously activated. One reason for this is that System 2 works on a limited budget. We can only focus on one or a few tasks at a time, and after working concentratedly for a while System 2’s resources get exhausted. After an extended time of advanced mathematics people simply lose their ability to focus and to use their System 2 (ibid., 49-51). System 1 uses very little concentration and consequently has vast resources. But in order to stay efficient it must use different heuristic techniques and biases to simplify complex problems. When System 1 faces a question which it is unable to answer immediately it does not stay put. Rather it switches to an approximate heuristic “proxy question” which is easier to answer, and provides a solution to that proxy question. Consider for example the question “what should I have for lunch?” This is a multivariable problem involving factors such as what is on the menu, what the prices are, what is more nutritious, what my personal preferences are, and so on. The way System 1 handles this complex issue is to simplify it into a heuristic question, such as “what is tasty?” or “what is my usual order?” and give an immediate answer to this heuristic question. (ibid., 113-6)

Despite their negative connotation heuristics are extremely important for our day to day lives. Imagine that I actually set out to choose the perfect lunch. I would have to consider all dishes at all nearby restaurants. I could make a table where I attribute different weight to different variables and try to find perfect information about each variable on each option. Such an enterprise would take hours, and even though I probably would get a marginally better lunch,
the cost in terms of time and energy for calculating which lunch is best would probably be greater than, and therefore offset, the gain I get in terms of a better lunch. I would thus, all things considered, be better off by just heuristically picking a second-best option, instead of spending hours calculating which option is the best.\(^5\)

Even though heuristics tend to make judgments significantly quicker and cheaper in terms of cognitive workload they also tend to produce systematic errors. These systematic errors in our reasoning are called cognitive biases. For example, System 1 tends to interpret new information in light of previous information, making the first information an anchor for new information. Consequentially the sequence in which information is presented can affect our judgements, even though this is irrelevant for the matter at hand. If I tell you that Alan is intelligent, diligent, impulsive, critical, stubborn and envious, but that Ben is envious, stubborn, critical, impulsive, diligent and intelligent you are likely to form a better opinion about Alan than about Ben, despite the fact that you have the exact same information about both persons (Kahneman 2013, 94).

Another problem is that System 1 judges probability by availability. If an event can be recalled easily and vividly people tend to overestimate its frequency (ibid., 149-51). For example, many people are worried about terrorist attacks or plane crashes, despite the fact that these are extremely uncommon, since vivid examples of these are easily accessible by our memory. Another bias is that of loss aversion, a tendency for us to avoid losses to a larger extent than we want to acquire an equivalent gain. For example, we tend to think that it is worse to lose a sum of money than it is good to gain an equivalent sum of money (ibid., 315-22).

These cognitive biases cause irrational and sometimes even dangerous decisions and behavior. The fact that we do not consider risk factors that are mentally unavailable makes us take unnecessary risks. We also unjustifiably tend to think of ourselves as exceptions from statistical probability. For example, many smokers seem to think that they themselves are unlikely to get health issues due to their smoking, even though they admit that the average smoker will, without giving any convincing reason to distinguish their prospects from that of the average smoker. Similarly, our tendency to misjudge economic prospects makes us invest

\(^5\) Researchers emphasizing the usefulness of heuristics are sometimes referred to as the “Fast and Frugal” school of heuristics study. For an example, see (Gigerenzer & Todd, 2001).
pensions and savings in sub-optimal funds, while loss aversion hinders us from selling falling securities in due time.

This psychological account is important for our discussion for two reasons. First, the model of human psychology as divided into an automatic and effortless System 1 and a concentrated and resource-intensive System 2 nicely conforms with the tacit paternalist assumptions which I listed in section III. It illustrates how a person can have a multitude of interests, namely System 1 interests which are created automatically and System 2 interests that arise through careful thought and consideration. These can come into conflict as the heuristic questions (e.g. what my usual order is) and the fundamental question (e.g. what I really ought to eat) can have different and incompatible answers. There also seems to be a clear hierarchy between the two. I identify with System 2, and my System 2 judgments and interests are better justified than my System 1 equivalents. It therefore seems reasonable to prioritize System 2 interests before System 1 interests. Finally, it also indicates that we sometimes fail to satisfy our System 2 interests, but that others can aid us to better satisfy them.

Secondly, the fact that we often tend to act in accordance with heuristics and biases rather than considered judgments enables a form of power exercise which we have not considered previously. For example, people are more prone to buy groceries that are placed in eye level in the store. A shopkeeper can consequentialy influence what people buy by how she arranges the wares in the store. Since the shopkeeper thereby makes her customers act in a way which they would not have otherwise done, the shopkeeper is exercising power over the customers. Furthermore, the shopkeeper can use this power in a way which is to her customers’ benefit, understood as their prime interest. For example, she can place healthy groceries at eye level, thereby making her customers buy healthier food. Such an action, which uses someone's heuristics and biases to influence that person’s behavior, and consequentialy can influence irrational humans but not perfectly rational “econs”, is a nudge (Thaler & Sunstein 2008, 8).

Another way to formulate the same idea is to say that $P$ nudge $Q$ when $P$ uses “choice architecture”, i.e. how a choice is presented, to affect $Q$’s behavior (ibid., 6). Choice architecture only regards how a given choice is presented, and not what the substantial alternatives of the choice is. It thereby differs from information and incentivizing. If one alters the choice situation by providing additional information this is a case of informing, not nudging. If one alters the choices, perhaps by offering an extra reward if one of the options is
chosen, one is incentivizing, and not using choice architecture or nudging. Choice architecture rather regards things such as what the default option for the choice is, how information about the alternatives is presented, or how the alternatives are formulated.

Proponents of nudging point towards two supposed advantages of using nudging rather than ordinary paternalistic interventions. First, unlike much of traditional paternalism nudging is not coercive. It may therefore be considered less offensive than these coercive paternalist actions. Many people have a fundamental moral intuition that coercion is prima facie wrong, and should be avoided if possible. As I argued in part II libertarians hold that coercion may only be used in defense, and thus reject coercive forms of paternalism. But as nudging is not coercive, such an argument could not be used against it. Nudging has therefore been described as a form of “libertarian paternalism” (ibid., 4-6).

Secondly, nudging, unlike coercive paternalist methods, is asymmetrical. It will in other words only affect persons who are making sub-optimal decisions, while those who already conform to their prime interest will not be affected (Camerer et. al. 2003, 1219-22). Consider once more the person who is highly allergic to healthy food. If we use a coercive form of paternalism to make people eat healthier this person will have to buy groceries she is allergic to, contrary to her prime interest. If we instead employ a nudging technique by placing the healthy food in eyesight more people would buy the healthy food, but the person who is highly allergic would still be able to buy her unhealthy food. Nudging is in this case Pareto optimal, as we have benefited the prime interests of some without harming the prime interest of anyone.

To summarize these two arguments, nudging affects choices without affecting options. If we believe that it is good to maintain options, then nudging seems to be a good substitute for coercive forms of paternalism.

V. Analysis

The purpose of this thesis is twofold. First to investigate how libertarianism, paternalism and nudging relate to each other. Secondly, to analyze whether, and if so when, paternalism is compatible with libertarianism. So far into the paper I have explained the concepts of libertarianism, paternalism and nudging. This section provides the central analysis of the paper.
I analyze the concepts by formulating them in a common terminology of methods and motives of power. I argue that whether an exercise of power is a nudge depends on its method while whether it is paternalist depends on its motive. Libertarianism is the view that coercive methods can only be used with defensive motives. Any paternalistically motivated non-coercive action is thus libertarian paternalist. I also argue that there several types of actions, such as informing, incentivizing and nudging, which satisfy these criteria. Consequentially I draw the conclusion that paternalism is compatible with libertarianism.

I conclude this section by also discussing a few interesting additional implications of my analysis. First, my analysis implies that asymmetrical paternalism is narrower than libertarian paternalism, as there are symmetrical paternalist actions which are compatible with libertarianism, but no asymmetrical paternalist actions which are not compatible with libertarianism. Secondly, my analysis indicates that strong and weak paternalism are equally compatible with libertarianism. Neither does libertarianism morally distinguish between acting paternalistically towards a person and using a person for a third party’s benefit.

Before I proceed I briefly summarize my paper so far. I have described libertarianism as the view that coercion may only be used in defense. \( P \) coerces \( Q \) iff \( P \) uses violence or the threat of violence against \( Q \). \( P \) uses violence against \( Q \) iff \( P \) 1) causes \( Q \) physical harm or pain, 2) physically restrains or enforces \( Q \)’s actions, or 3) destroys, intrudes on or otherwise deprives \( Q \) of \( Q \)’s legitimate property. \( P \) acts paternalistically against \( Q \) iff \( P \) exercises power over \( Q \) with the intention of altering \( Q \)’s behavior in a way that is beneficial for \( Q \). Finally, \( P \) nudges \( Q \) iff \( P \) uses choice architecture in order to influence \( Q \)’s behavior.

V.I. Formulating notions in terms of power

In order to analyze the relationship between paternalism, nudging, and libertarianism I first make a brief analysis of power. I take power to be the ability to make an event or a situation occur, which would otherwise not have come about. Interpersonal power is thus the ability to make another person perform an action or behave in a way which that person would otherwise not have done. I analyze power along two orthogonal dimensions; method and motive.
The first dimension of power is what method\(^6\) \(P\) uses to exercise power over \(Q\). There are several possible ways through which \(P\) could influence \(Q\)’s behavior, and we can imagine a sequence of such methods ranging from the “hardest” to the “softest”. A hard method would be for \(P\) to enforce \(Q\)’s \(A\)-ing or prohibit \(Q\)’s non-\(A\)-ing using violence or the threat of violence. A softer method would be to incentivize \(Q\) into performing \(A\) by rewarding \(Q\)’s performance of \(A\). An even softer method of power would be for \(P\) to use choice architecture, thereby making \(Q\) more likely to \(A\). Finally, \(P\) can also convince \(Q\) that the performance of \(A\) is in \(Q\)’s best interest, thereby bringing about \(Q\)’s \(A\)-ing.

The second dimension along which I analyze power is motive. The list above does not tell us anything about why \(P\) exercises power over \(Q\). For example, the fact that \(P\) has exercised power over \(Q\) by locking \(Q\) into a room, says nothing about why \(P\) has locked \(Q\) into the room. We can imagine several different possible motives. Perhaps \(Q\) is upset and has threatened to harm \(P\), making \(P\) lock \(Q\) into the room out of self-defense. \(P\) may also have locked \(Q\) into the room because \(P\) is convinced that \(Q\), if not locked up, would go and eat unhealthy cake. \(P\) has then exercised power over \(Q\) with the intention of benefitting \(Q\), and has thus acted paternalistically towards \(Q\).

For each motive there are a number of different methods which could be used. Likewise, for each method there are several different possible motives. These two dimensions are orthogonal, which means that there is no necessary connection between any particular method and any particular motive. I make the assumption that for any action there is only one primary motive. In other words, if the primary motive behind an action is altruistic, it is not defensive, and vice versa. Using this simple schema we can reformulate nudging, paternalism and libertarianism in a common terminology, in order to investigate how they relate to each other.\(^7\)

I argue that \(P\) nudges \(Q\) iff \(P\) uses choice architecture in order to influence \(Q\)’s behavior. This seems to be a method of exercising power, rather than a motivation. It is analogous to “prohibition” rather than “self-interest”, which is indicated by the fact that a nudge could be performed out of self-interest, but could not be prohibitive. Nudging can thus be performed with several different possible motives. Consider the shopkeeper who promotes certain wares by placing them in eye-level. This could be used to promote more expensive goods, thereby

---

\(^6\) “Power tools”, if you will.

\(^7\) The appendix contains this information in tabular form.
serving the self-interest of the store. It could also be used to promote healthier goods, and thereby benefit the prime interest of the customers. Nevertheless, the action seems to be a nudge, regardless what the motive behind the action is.

Whether or not an exercise of power is paternalist is merely a matter of the motive behind the action. When $P$ warns $Q$ that the cake $Q$ is about to eat is very unhealthy $P$ is using a very soft method of power – information – but $P$’s motivation is still to benefit the target of her power exercise – $Q$. $P$ could also use a harder method of power. Perhaps $P$ could have cuffed $Q$ and thereby physically prevented her from eating the cake. While this method would be significantly “harder” the action would still be paternalist, as its motive is to benefit its target.

According to libertarianism coercion may only be used in defense. Libertarianism thus involves both a motive – defense – as well as a method – coercion. To be more precise, it uses a certain motive as a condition for justness in a certain method of power. Concerning non-coercive actions libertarianism does not pass judgment. It does however follow from libertarianism that it would be unjust to use coercion in order to prohibit or hinder any of these non-coercive power methods, as one would then use coercion with a non-defensive motive. Consequentially all non-coercive power methods are compatible with libertarianism.

As I have described libertarianism there are no positive libertarian imperatives. An action can thus only have neutral status, and be compatible with libertarianism, or negative status, and be incompatible with libertarianism. Libertarian paternalism would thus naturally include any paternalist action which is compatible with libertarianism. In other words, any libertarian paternalist power exercise must satisfy the criteria to 1) be primarily motivated by a desire to benefit the target of the action and 2) to be defensive, i.e. primarily motivated by a desire to protect oneself or others, if the method of the action is coercive.

I assumed that any action only has one primary motive. If an action is paternalistically motivated, it cannot be defensively motivated, and vice versa. Consequentially no paternalist actions are defensive. Any coercive paternalist action is therefore incompatible with libertarianism. However, as all non-coercive power exercises are compatible with libertarianism, regardless of their motivation, all non-coercive paternalist actions must also be compatible with libertarianism. Consequently, satisfying the above mentioned criteria is equivalent to satisfying the criteria to 1) be primarily motivated by a desire to benefit the target of the action and 2) not being coercive.
V.II. Are there cases of libertarian paternalism?

In the previous part of the analysis I formulated the terms “paternalism”, “libertarianism”, “libertarian paternalism” and “nudging” in a common terminology. According to my analysis an exercise of power is libertarian paternalist iff 1) it is primarily motivated by a desire to benefit the target of the action and 2) it is not coercive. I now instead turn to the substantial question of whether there are actions which satisfy these criteria. I have mentioned four different methods of power. The first is coercion. Coercion is only compatible with libertarianism if it is defensive. However, if it is defensive it will not be paternalist. There are consequently no coercive paternalist actions which are compatible with libertarianism. Let us instead consider the three other methods of power; nudging, informing and incentivizing.

According to my taxonomy an action is a nudge if it uses choice architecture to influence a person’s behavior. Consider the case of the shopkeeper who places healthy food on eye-level, and thereby nudges her customers into buying and consuming a healthier diet. This is a power exercise, as it makes people behave in way which they would not have done otherwise. As the power exercise is performed with an intention to benefit the customers it is paternalist. Is it coercive? According to some accounts of coercion, coercion is essentially pressure on someone’s will. Nudges do in some sense affect or pressure the will of their targets. However, I have argued that it is not this sense of coercion which libertarians are interested in. Rather, I have argued that an action is coercive if it enforces some behavior with violence or the threat of violence. Nudging does none of these. I therefore conclude that nudges are compatible with libertarianism, and that nudges performed with a paternalist motive consequentially are libertarian paternalist.

The third method of power I consider is information. Consider Q, who wishes to eat cake, but is oblivious to the long-term health effects of doing so. P is a nutritionist and informs Q that eating cake would have very bad long-term effects on Q’s health. Q consequently abstains from eating cake. P’s informing of Q has made Q act in a way that Q would not have otherwise done. It is thus a power exercise. It could also have been performed with the intention of benefiting Q, thereby making it paternalist. Finally, it does not in any way rely on violence or the threat of violence, and is thus fully compatible with libertarianism. I therefore conclude that informing, when performed with an intention to benefit the target of the action, is libertarian paternalist.
The fourth and final method of power I consider is incentivizing. Incentives can be positive, where some sort of compensation is granted for behavior which the incentivizer wants to encourage. Incentives can also be negative, where the incentivizer denies or withdraws some benefit if the person behaves in a way which the incentivizer wants to discourage.\(^8\)

Is negative incentivizing compatible with libertarianism? Consider the case of a shopkeeper \(P\) and a customer \(Q\). \(P\) knows that she is likely to harm the prime interest of her customers by selling them cakes. Raising the price of cakes would deter many of her customers from buying these and thereby reduce the general cake consumption. \(P\) consequently raises the price of cakes. This is a power exercise as \(P\)’s action makes the customer \(Q\) behave in a way that \(Q\) would not have otherwise done. As the motivation behind the power exercise is to benefit its target it is also paternalist. It does not involve any violence or threat of violence, and is thus compatible with libertarianism. Consequentially, it is a case of libertarian paternalism.

How about positive incentivizing? Consider the child \(Q\) who does not want to eat her vegetables. In order to make \(Q\) eat her vegetables the parent \(P\) offers \(Q\) a reward, provided that she eats her vegetables. This is an exercise of power, as it causes \(Q\) to eat her vegetables, and thereby behave in way which she would not have otherwise done. As the power exercise is intended to benefit \(Q\) it is paternalist. Finally, the action does not constitute a case of violence or threat of violence, thus making it fully compatible with libertarianism. It is thereby a case of libertarian paternalism.

Perhaps the fact that positive incentivizing and other non-coercive paternalist actions are compatible with libertarianism should not be all too surprising. Some have argued that it is compatible with libertarianism to sell one’s own body parts (e.g. Frederick 2010). While this question exceeds the scope of this essay, it would be strange to allow people to sell their kidney to others, but forbid that people get paid by others in order to take good care of their kidneys. If libertarianism allows \(P\) to exercise non-coercive power over \(Q\) in order to benefit \(P\), it would be odd to disallow \(P\) to exercise non-coercive power over \(Q\) in order to benefit \(Q\).

I therefore conclude that paternalism is compatible with libertarianism. More specifically, it is compatible iff the method used to exercise the paternalist action is not coercive. I also

\(^8\) Sometimes taxation and fines are also considered forms of negative incentivizing. According to my taxonomy these actions should however rather be labeled as coercive methods of power, as they entail that one deprives the target of their (presumably) legitimate property.
conclude that the libertarian paternalist ‘tool-box’ is larger than usually thought, as it in addition to nudges also contains informing and incentivizing methods.

The questions I set out to answer in this paper is how libertarianism, paternalism and nudging relate to each other, and whether and when paternalism is compatible with libertarianism. However, in addition to answering the main thesis of this paper, my analysis also has additional interesting implications which fall outside the immediate scope of the paper. In the final part of my analysis I briefly discuss a few such implications.

V.III. Additional consequences

The purpose of this paper is to analyze how paternalism, libertarianism and nudging relate to each other, and whether paternalism is compatible with libertarianism. My analysis has showed that there are several forms of paternalism which are not coercive, and hence compatible with libertarianism. In this subsection I briefly discuss a few additional interesting implications which fall outside of the main purpose of the paper, but which follow from my analysis.

Asymmetrical paternalist actions are paternalist actions which only affect those who are benefitted by the power exercise, without having some additional harmful effect on anyone else (Camerer et. al. 2003, 1219). Consider the paternalist $P$ who uses the coercive paternalist method to fine everyone who consume unhealthy food. $Q$, who is allergic to healthy food, would not be benefited by changing to a healthy diet. She would however not be able to avoid the fine. This action would thus affect people symmetrically, regardless of whether they are actually benefitted by the action. If $P$ instead nudges people into buying healthy food by placing it on eye-level in stores, $Q$ is not affected substantially as she can simply look for her unhealthy food on the other shelves. The nudge is thus asymmetrical. The coercive method is not only symmetrical but also incompatible with libertarianism. Similarly, the nudging method is both asymmetrical and compatible with libertarianism. One could therefore assume that libertarian paternalism and asymmetrical paternalism converge.

However, I do not believe that this is the case. Consider our case of negative incentivizing. The paternalist shopkeeper $P$ increases the price of unhealthy food in order to benefit the prime interest of her customers. $Q$, who is allergic to healthy food, will be unable to change her consumption pattern. Instead, $Q$ will have to pay the increased price for unhealthy food,
despite the fact that eating unhealthy food best satisfies \( Q \)’s prime interest. The negative incentivizing is thus symmetrical.\(^9\) It is however compatible with libertarianism. According to libertarianism the shopkeeper is allowed to demand any prices she wishes for her goods, for whatever reason she likes, as long as she does not aggressively coerce anyone. There are therefore cases of libertarian paternalism which are not cases of asymmetrical paternalism.

Are there cases of asymmetrical paternalism which are not compatible with libertarianism? Any and all paternalist actions which are incompatible with libertarianism are coercive. The question might therefore also be posed as “are there coercive asymmetrically paternalist actions?” I believe that this is not the case. If an action is asymmetrical, one has a possibility to opt out from it. It is in that sense voluntary to be subjugated to it. “Voluntary coercion”, however, seems to be a contradiction in terms. If I have an opportunity to “opt out from being coerced”, then I am not really being coerced. I therefore conclude that while there are cases of libertarian paternalism which are not cases of asymmetrical paternalism, there are no cases of asymmetrical paternalism which are not cases of libertarian paternalism.

Another additional consequence regards the difference between two forms of paternalism. In section III I distinguished weak paternalism from strong paternalism. Weak paternalism interferes with the means a person uses to reach her explicit ends. Strong paternalism interferes not only with the means a person uses to reach her ends, but also with the ends themselves (Dworkin 2017). Weak paternalism can consequentially be considered less intrusive than its strong counterpart. One could therefore make the hypothesis that weak paternalism is more compatible with libertarianism than strong paternalism is.

I do however believe that this hypothesis is wrong. Weak and strong paternalism are forms of paternalism in general. I have argued that any paternalist action is compatible with libertarianism iff it is not coercive. This applies to weak paternalism and strong paternalism equally. If a paternalist action is coercive it will be incompatible with libertarianism regardless whether it weak or strong. Similarly, if the action is not coercive it will be compatible with libertarianism regardless of whether it is weak or strong. I therefore conclude that weak paternalism and strong paternalism are equally compatible with libertarianism.

---

\(^9\) Perhaps one could argue that the case is not genuinely symmetrical, as \( Q \) could easily go to some competing store to get her low-price unhealthy food. This counter-argument can however easily be defeated if we simply assume that \( P \) is in a position of monopoly over unhealthy food.
I also made a distinction between paternalism and “benevolent usage”. Any exercise of power which is intended to benefit the target of the action is paternalist, while benevolent usage is the exercise of power which is intended to benefit a third person. For example, P could nudge Q into donating more money to charity, in order to benefit R. This would not be a form of paternalism against Q, as P’s motivation is to benefit R rather than to benefit Q. One could make the argument that benevolent usage is more intrusive than paternalism is, as paternalism is intended to benefit its target, while benevolent usage is not. Therefore, one could argue, paternalism is more compatible with libertarianism than benevolent usage is.

According to my analysis, however, this is not the case. The only factors that matter according to libertarianism, is whether the action is defensive and whether it is coercive. Neither benevolent usage nor paternalism are defensive. It is therefore true for either of them that they are compatible with libertarianism iff they are not coercive. P may thus nudge, incentivize or inform Q into benefiting R. The fact that the usage is benevolent does not seem to matter either. P may just as well nudge, incentivize or inform Q into benefiting P. Therefore, the argument generalizes to all non-coercive cases of “usage”, i.e. all non-coercive power exercises. I therefore draw the conclusion that usage and paternalism are equally compatible with libertarianism.

VI. Summary and conclusions

It is commonly believed that paternalism is at odds with libertarianism (Boaz 1997, 211). Recent literature has suggested “nudging”, as a form of paternalism that is compatible with libertarianism, even referring to it as “libertarian paternalism” (Thaler & Sunstein 2008, 4-6). In this paper I have analyzed how libertarianism, paternalism and nudging relate to each other, and whether, and if so when, paternalism is compatible with libertarianism.

I characterize libertarianism as a ‘broad church’ which is unified in a smallest common denominator in the non-aggression principle, according to which coercion may only be used in defensive purposes. In this context coercion should be understood as the enforcing of behavior with violence or the threat of violence. I characterize paternalism as an exercise of power which is intended to benefit the person it is targeted against. Finally, I characterize a nudge as an action which utilizes choice architecture to influence another person’s behavior.
As a first step in my analysis I formulated these concepts into a common terminology of methods and motives of power. These two dimensions of power are orthogonal, meaning that no motive necessitates any method, and vice versa. In such terms nudging is a method of power. A power exercise is paternalist iff it is motivated by an intention to benefit the target of the action. Libertarianism is the view that one method of power – coercion – may only be used with one motive – defense. Consequentially all non-coercive actions are compatible with libertarianism. Any non-coercive paternalist action is thus libertarian paternalist.

As a second step I analyze whether there are any paternalist actions which are compatible with libertarianism. Nudges, incentives and information are methods of power which are not coercive. They are thus compatible with libertarianism. Further, as they can be performed with paternalist motives, there are paternalist actions which are compatible with libertarianism. I therefore conclude that paternalism and libertarianism are compatible, and more specifically that they are compatible iff the method used to exercise paternalism is not coercive.
VII. References


[28]


• Mill, John Stuart (2011) [1859]. On Liberty. Luton;Palo Alto:: Andrews UK Ltd


### VIII. Appendix

#### Table 1.

<table>
<thead>
<tr>
<th>Method / Motive</th>
<th>Defensive (X)</th>
<th>Paternalist (Y)</th>
<th>Other (Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Coercive</td>
<td>X1</td>
<td>Y1</td>
<td>Z1</td>
</tr>
<tr>
<td>2. Incentivizing</td>
<td>X2</td>
<td>Y2</td>
<td>Z2</td>
</tr>
<tr>
<td>3. Nudging</td>
<td>X3</td>
<td>Y3</td>
<td>Z3</td>
</tr>
<tr>
<td>4. Informing</td>
<td>X4</td>
<td>Y4</td>
<td>Z4</td>
</tr>
</tbody>
</table>


Nudge: 3X-Z. Libertarian paternalism: Y2-4.