

MITIGATING THE THREAT THAT VOLUNTARY BIOENHANCEMENT POSES TO AUTONOMY

James Eder

A Thesis Submitted in Partial Fulfilment of The Requirements for the Degree
of Master of Arts in Philosophy in the
University of Canterbury
2023

Abstract

New technologies are likely to emerge in the coming decades that significantly enhance humans' biology. But there are many examples in recent human history of new inventions, such as leaded petrol and asbestos, that are released to the public without a full understanding of their negative impacts. The purpose of this thesis is to anticipate how bioenhancement will negatively impact autonomy and suggest strategies that mitigate such harm.

Some scholars have controversially argued that bioenhancement should be made compulsory, while others have responded by saying that doing so would grossly violate people's autonomy. But I highlight that even the voluntary use of bioenhancement poses a threat to autonomy. Just as voluntarily drinking alcohol can undermine our judgement and decision-making, so too can voluntarily bioenhancing ourselves with other drugs or technology.

This thesis analyses theories of autonomy to assess the impact of bioenhancement on autonomy. I analyse how autonomy is impacted differently depending on whether bioenhancement targets human cognition, emotions or behaviour.

Nudge theory proposes that decision-making can be influenced by exploiting a contemporary understanding of human psychology and behaviour to guide individuals toward specific choices, without undermining their autonomy. I examine the relationship between nudge theory and autonomy. I argue that designing bioenhancements as "bionudges" can mitigate the risk of undermining human autonomy and instead promote it.

Acknowledgements

Firstly, thank you to my supervisors, Dr Carolyn Mason and Diane Proudfoot, for your advice and encouragement.

Thank you to the UC lecturers of English and Philosophy. I owe so much to your wisdom, passion and often underappreciated hard work. You have made university a place where I can be myself and do work that I love. Each of your undergraduate courses played a little role in writing this thesis.

Thank you to Roseanna Brailsford for being such a caring and passionate tutor, Stephen Rowe for your friendship, and thank you to Dr Carolyn Mason for letting me so frequently barge into your office over the last seven years to talk Philosophy.

Thank you to future-James for paying off this gigantic student debt. Past-James, I will never forgive you for withdrawing course-related costs every year.

Finally, thank you Lydia Dudson for always pushing me to do my best, for listening to me ramble, and for your love.

Table of Contents

| | |
|-------------------------------------------------------------------------------------------------------|-----------|
| Chapter 1. Introduction | 6 |
| 1.1. Theories of autonomy and their relationships to bioenhancement | 7 |
| 1.2. Using bioenhancement to alter cognition, emotions and behaviour..... | 8 |
| 1.3. Using nudge theory to mitigate negative impacts of bioenhancement on autonomy | 10 |
| Chapter 2. Theories of autonomy and their relationships with bioenhancement | 14 |
| 2.1. Understanding and defining autonomy | 14 |
| 2.2. Decisional and practical autonomy..... | 17 |
| 2.3. Local and global autonomy | 18 |
| 2.4. Substantive and procedural autonomy..... | 20 |
| 2.4.1. Substantivism..... | 20 |
| 2.4.2. Proceduralism | 24 |
| 2.5. Hierarchicalism..... | 26 |
| 2.6. Coherentism | 28 |
| 2.6.1. Coherent integration | 32 |
| 2.7. Conclusion..... | 33 |
| Chapter 3. Using bioenhancement to alter cognition, emotions and behaviour | 34 |
| 3.1. Introduction | 34 |
| 3.2. Defining cognitive, emotional and behavioural bioenhancement | 34 |
| 3.3. Assessing different targets of bioenhancement against decisional and practical autonomy | 35 |
| 3.3.1. Cognitive bioenhancement | 36 |
| 3.3.2. Emotional bioenhancement | 37 |
| 3.3.3. Behavioural bioenhancement | 38 |
| 3.4. Assessing the impact of different targets of bioenhancement on authenticity | 38 |
| 3.4.1. Cognitive bioenhancement | 39 |
| 3.4.2. Emotional bioenhancement | 39 |
| 3.4.3. Behavioural bioenhancement | 40 |
| 3.5. Conclusion | 40 |
| Chapter 4. Using nudge theory to mitigate negative impacts of bioenhancement on autonomy | 42 |
| 4.1. Introduction | 42 |
| 4.2. Nudging is not necessarily harmless to autonomy | 43 |

| | |
|-----------------------------------------------------------------------------------------------------|-----------|
| 4.3. Manipulative and transparent nudges | 43 |
| 4.4. Heteronomous influence and cosmonomous choice environments..... | 47 |
| 4.5. Reason-giving nudges are less harmful to autonomy than non-reason-giving nudges | 49 |
| 4.6. Influencing type 1 processing threatens autonomy more than influencing type 2 processing | 51 |
| 4.7. The nudge axis and bionudges | 54 |
| 4.7.1. Transparent bionudges | 56 |
| 4.7.2. Non-transparent bionudges..... | 58 |
| 4.8. Settled and unsettled preferences..... | 59 |
| 4.9. The impact of nudging on global autonomy | 62 |
| 4.10. Conclusion..... | 64 |
| Chapter 5. Conclusion..... | 67 |
| Chapter 6. References..... | 70 |

Chapter 1. Introduction

In this paper I examine how the use of bioenhancement technologies impacts autonomy. Bioenhancement involves temporarily or permanently changing a person's biology with the intention of improving a person's natural abilities or giving them new abilities. For example, this can be done by using pharmaceuticals that affect a person's brain chemistry, gene editing, implanting computer chips into a person's brain, or by giving them prosthetic body parts. The expanding understanding of the sciences and innovation of technology over recent decades has been followed by an expansion of philosophical analysis into their applications in enhancing human abilities.

Philosophers such as Persson and Savulescu have argued that bioenhancement should be compulsory (Persson and Savulescu, 2014). Others have responded with concerns that a program of compulsory bioenhancement would compromise the autonomy of members of the public (Sparrow, 2014; Wilks, 2018). There are many arguments about the effect of compulsory bioenhancement on autonomy, but few, if any, about voluntary bioenhancement.

I examine how the *voluntary* use of bioenhancement technologies impacts autonomy. While voluntarily drinking alcohol can enhance our mood and social interactions, it can also undermine our judgement and the degree to which we are in control of our decisions. In light of this, we should assess how the voluntary use of bioenhancement affects our decision-making and actions. For as long as liberal democracies remain liberal and democratic, it is unlikely that they will establish compulsory programs of bioenhancement. If bioenhancement technologies become available to the public in liberal democracies, it is more likely that they will be accessed via the black market, regulated markets or the free market. This is why I have chosen to analyse voluntary bioenhancement's impact on autonomy.

In fact, I argue that bioenhancement drugs and technology are already being widely used. People use caffeine and sugar to improve their cognition and mood; others use illicit drugs to improve their mood or creativity; and various studies of college students in the US have found that the nonmedical use of prescription stimulants such as Adderall and Ritalin ranges from 6.9% to as much 34% of participating students, and that the most common motivation for using them is to improve academic performance (Cook, Kurtz-Costes and Burnett). Proponents of the extended mind thesis – who state that what constitutes a person's mind includes the tools that they use to assist with cognitive tasks – argue that the use of many everyday technologies amounts to cognitive enhancement (Heersmink). When we start putting these technologies inside our bodies, they become bioenhancement.

Unfortunately, sometimes new technologies are hastily released without a full understanding of their impacts. This is particularly apparent when governments and companies are influenced by market pressures to permit the sale of a new technology. In this respect, I worry that bioenhancement technologies will be no different from the likes of asbestos, leaded petrol and cigarettes. Technologies such as brain chips, augmented reality and 'smart drugs' are becoming much less confined to the science fiction genre as humans seek to enhance themselves with their

tools (Carmigniani and Furht 2011; Rose, 2002; Talha, 2020). It is important that we understand the impacts of these technologies as much as possible before they are invented or distributed, so that we can anticipate the issues and mitigate them.

I analyse the impact that bioenhancements have on autonomy based on an assumption that the bioenhancements achieve their intended effect. For the sake of my analysis, I assume that a cognitive bioenhancement really does enhance cognition, an emotional bioenhancement succeeds in altering a person's experience of emotions, and that a behavioural bioenhancement does produce the expected behaviour. However, I accept that bioenhancements will have unintended effects. For example, while I assume that a cognitive enhancement will succeed in increasing reasoning and deliberative ability, I do not assume that this cannot also cause an unintended effect of delayed decision-making. Of course, if a cognitive enhancement fails to improve, or indeed reduces cognitive ability, then this would undermine autonomy. My discussion focuses how autonomy is impacted when scientists, chemists, engineers and technicians design bioenhancements that do achieve their primary intended outcome.

I also defer discussion of what truly constitutes enhancement. Some analyses of the philosophy of enhancement are critical of arguments whose premises assert that some altered ability is an objective improvement, without limiting the claim by reference to some particular activity or goal. While an athlete's use of steroids may be considered an enhancement in respect to their physical performance, the impact of steroid-induced mood swings on the athlete's mental states might damage their personal life. It is difficult to say that particular alterations of abilities are enhancements in and of themselves. In this paper I discuss the enhancement of cognition, emotions, and behaviour. My description of the enhancement of these human capacities is also subject to this kind of criticism; it would be presumptuous to say that a human having profound and deliberate control over their emotions is necessarily and solely an enhancement. Critics might argue such control degrades a human's dignity or deprives them of the full human experience. My analysis focuses on the kinds of alterations of human capacities that are *typically* deemed to be enhancements, and more importantly on how these alterations affect autonomy. I defer questions about whether such alterations truly constitute improvement, except insofar as they affect autonomy.

1.1. Theories of autonomy and their relationships to bioenhancement

In this paper, I examine theories of personal autonomy and apply them to aspects of bioenhancement. I mainly refer to autonomy insofar as it means the capacity of a person to exercise self-rule, or the degree to which a person actually exercises that capacity.

Autonomy has decisional and practical dimensions. Decisional autonomy refers to an agent's ability to assess their situation, identify and deliberate over their options, and select a course of action. Practical autonomy refers to an agent's ability to act on that decision. Cognitive bioenhancements that alter a person's reasoning ability may impact their decisional autonomy but not their practical autonomy, and vice versa for bioenhancements that alter only a person's behaviour.

Autonomy can be understood as procedural or substantive. Procedural autonomy holds that the expression of autonomy consists only of the process by which an agent comes to a decision. If an agent completes particular kinds of considerations and the process of deliberation is not impaired by another person, then a decision can be considered procedurally autonomous. Substantive autonomy, on the other hand, holds that autonomy depends not only on the process by which a person reaches a decision; the content of the person's decision matters too. These conceptions of autonomy explain how bioenhancements that improve a person's reasoning may increase their procedural autonomy, and how bioenhancements that affect only a person's behaviour may increase their substantive autonomy if their new behaviours align with their desires.

Autonomy can also be understood through either hierarchical or coherentist frameworks. These frameworks help us discriminate between more or less authentic desires and therefore help us to determine whether a decision is autonomous when a person has conflicting desires. Hierarchicalism visualises desires as existing on the rungs of ladders, and holds that a person can have higher order desires that pertain to lower order desires. Coherentism views desires as existing at the intersections on a web, where desires that are logically connected to many other desires are more autonomous than those less connected. Under these frameworks, bioenhancements increase autonomy if they increase how much a person reflects or acts on their higher order or more coherent desires.

Further, part of what makes a decision autonomous is the agent's assessment of the decision against their values and preferences and the formation of their own attitude toward the choice. This is important because it distinguishes decisions into which a person has no input, from decisions in which an agent has wilfully declined to participate. For instance, when a person defers to a doctor's advice, or lets someone else pick the cake flavour for their birthday party, they autonomously relinquish their autonomy about a particular decision because someone else is better equipped to make the decision or because the consequences of the decision are not great enough for the agent's efforts of deliberation to be worthwhile. This demonstrates the counterintuitive possibility that person could give informed consent for a bioenhancement that undermines their autonomy in regard to certain decisions, but whose decision is autonomy-friendly or even autonomy-enhancing.

1.2. Using bioenhancement to alter cognition, emotions and behaviour

These theories of autonomy provide the terms, tools and frameworks that enable me to assess how different forms of bioenhancement, such as cognitive, emotional and behavioural bioenhancement, impact a person's autonomy. Cognitive enhancement involves improving the capacities that we use for gaining, processing, storing, and retrieving information. Emotional enhancement involves altering the types of emotions we experience toward certain behaviours, people, and topics, the degree to which we feel them, and the control we have over them. Behavioural enhancement involves changing a person's behaviour, without first altering their

emotions or cognition. I analyse on how bioenhancing cognition, emotions or behaviour each differently impact autonomy.

I use the concepts of identity and alienation, along with the decisional/practical autonomy distinction, to assess how altering these human capacities impacts on authenticity and subsequently on autonomy. In this respect, identity refers to the elements of the agent's objective character that make them the person that they are. If bioenhancement changes a person too much, it may destroy their identity and therefore undermine their autonomy. Alienation, on the other hand, concerns itself with the subjective attitudes an agent has toward the changes caused by bioenhancement. If an agent feels alienated from themselves as a result of bioenhancement, then it has undermined their authenticity, *ceteris paribus*.

Cognitive bioenhancement can increase decisional autonomy by improving a person's ability to gather information relevant to a decision and their ability to deliberate. It has no negative effect on practical autonomy, and it may increase practical autonomy insofar as it helps an agent overcome barriers that prevent them from acting on their made decision.

Profound cognitive bioenhancement may undermine a person's sense of self if they feel alienated from their new powers of cognition and their subsequent beliefs and decisions. If the cognitive bioenhancement is sudden and profound, then it may destroy their identity.

While emotions themselves are not necessarily involved in autonomous decision-making, they can distract agents from their deliberation or from acting on their decisions. Crude forms of emotional bioenhancement that simply reduce or increase the prevalence of certain emotions across all kinds of scenarios pose a threat to autonomy. This is because even unpleasant or sometimes harmful emotions such as anger are needed to respond appropriately in certain situations. Humans experience a range of emotions and feel them at varying intensities and for different durations. What emotions, intensities, and durations are best for autonomy will vary from scenario to scenario. Therefore, emotional bioenhancements must be delicate and focused in order to avoid undermining a person's autonomy. The most autonomy-friendly forms of emotional bioenhancement are those that make a person more aware of their emotions and increase their ability to control them.

Emotional bioenhancement poses greater risks to a person's authenticity than cognitive bioenhancement. This is because changing the way a person feels about certain behaviours and decisions, without changing their reasoning for endorsing or opposing them, means the person is less likely to understand why they desire the things that they desire. If one cannot understand why they feel a certain way about their behaviours or decisions, then they may feel alienated from their emotions; their emotions may become inauthentic. Further, if the way we feel about certain people, topics, and behaviours comprises an important or large proportion of our identity, then emotional bioenhancement that changes these feelings may destroy our identity.

Behavioural bioenhancement does not threaten decisional autonomy. This is because it affects only the behaviours that result from decision-making. However, behavioural bioenhancement

imposes significant limitations on practical autonomy because it influences our ability to act on decisions.

Behavioural bioenhancement also poses a threat of self-alienation. Because the effects of behavioural bioenhancement can involve preventing an agent from behaving in a way that they have chosen, the agent can feel alienated from their behaviour. Therefore, behavioural bioenhancements should aim to help people behave in ways that align with their higher order or more coherent desires. While agents may still feel some alienation from their behaviour, insofar as it does not accord with their immediate decision-making, they will not feel significantly alienated from their behaviour if it is a behaviour that they wish they endorsed. On the other hand, behavioural bioenhancement poses only a minor risk to a person's identity if a person's behaviour makes up only a small part of their identity. When an athlete is paralysed, we typically accept that part of what makes them who they are is gone, but we do not think their identity is destroyed. Further, behavioural enhancements are unlikely to be as profound or as permanent as paralysis, so there is good reason to believe that behavioural bioenhancement does not pose a significant threat to identity.

It is important to understand how autonomy is impacted by the bioenhancement of different human capacities because this enables us to bioenhance in ways that prevent or mitigate negative impacts on autonomy.

It is important to note that the drugs and technologies, and their actual effects, to which I refer in this paper, are not necessarily going to impact autonomy as much future drugs and technologies. For example, while studies such as Battleday, et al. find that Modafinil enhances cognitive performance in normally functioning adults, they find that the effect does not amount to a major cognitive enhancement. Farah highlights that there is still much unknown about bioenhancement and that a range of drugs and technologies currently believed to enhance cognition, may not be so enhancing after all (Farah, 379-80). While there are not currently any 'wonder drugs' that profoundly improve cognition, we *do* know that we can influence brain chemistry, and subsequently cognition, through the use of pharmaceuticals. My references to various drugs and technologies in this paper are intended to illustrate that it is possible to impact cognitive function via pharmaceuticals and technology, not that these particular drugs would have the same degree of effect that I describe in my arguments about autonomy.

1.3. Using nudge theory to mitigate negative impacts of bioenhancement on autonomy

In addition to some of the considerations above, I propose using nudge theory to inform the design and application of bioenhancements to mitigate negative impacts on autonomy.

Nudge theory involves using a particular understanding of human psychology and behaviour to design software, physical environments and decision pathways, to influence the way a person makes decisions and the behaviours that they exhibit. In nudge theory literature, these deliberately-designed environments are called 'choice architecture', and the designers, 'choice architects'. Making an option a default is a common method of influencing people to select it.

Nudge theory is intended to improve decision-making without undermining a person's autonomy. I analyse arguments about nudge theory and apply them to the use of bioenhancement. I call bioenhancements that are informed by nudge theory "bionudges".

A key element of nudge theory is that nudges should never remove or forbid options from an agent, and the influence of the nudge should be almost effortless to resist. While proponents of nudge theory argue that a nudge that undermines autonomy is not a nudge *by definition*, in practice, choice architects can get it wrong. I argue that the degree to which nudges undermine a people's autonomy varies by case. Choice architects can misjudge the influence that a particular nudge exerts on a person's decision-making. Further, choice architects may misjudge the susceptibility to influence of the person they nudge, their 'nudgee'. Therefore, simply designing bioenhancement to resemble nudges is insufficient if we aim to bioenhance without undermining autonomy. We need to assess each bionudge's impact on autonomy.

Some scholars argue that nudges are manipulative because their influence is not obvious to the nudgee. They argue that if a nudgee is unaware that their choice environment is designed to influence the way they make a decision, then the influence is manipulative. Bovens argues that to avoid manipulating nudgees, choice architects should inform people that they are being nudged. However, Bovens also argues that doing so would undermine the purpose of the nudge, because nudges work better when agents are not aware that they are being influenced. Contrary to Bovens, some studies have found that making nudges transparent does not reduce the likelihood that a nudged option is selected. Bionudges can be made more transparent by alerting the agent when they are taking effect, or by designing them solely to improve an agent's reasoning or provide them with explicit reasons to make a certain decision.

One might argue that nudges undermine a person's autonomy in virtue of the fact that they involve influence coming from another person. However, undesigned or random choice environments also influence agents, despite this influence not coming from another person. If a supermarket designer randomises product placement, customers are still more likely to select whatever is closest to eye-level. Some environments also cannot be made free of influence. However, some nudges add influence into a choice environment, rather than simply replace it. Adding 'SALE' signs around a product increases the existing level of influence to purchase that product and this increases the risk posed to autonomy. *Ceteris Paribus*, bionudges that add extra influence to a choice environment, undermine autonomy.

Further, some nudges can influence by targeting humans' irrational schema or cognitive weaknesses. For instance, predatory loan companies whose advertisements target people's short-term bias, undermine their customers' autonomy. This is a concern for autonomy because the choice architects intentionally engage thought processes that do not produce autonomous decisions. Provided that cognitive bionudges would improve a person's powers of reason, they are unlikely to undermine autonomy in this respect. However, while behavioural bionudges do not directly act on a person's reasoning and decision-making, they can still play on a person's irrational schema. A behavioural bionudge that briefly forces an agent to begin a certain activity, influences them via their status quo bias to continue and complete that activity.

While some nudges increase the influence in a choice environment, or prey on humans' irrational schema, their influence may offset the existing choice environment that undermines a person's autonomy. For example, adding 'SALE' signs only on healthy products increases the influence in the choice environment, but it offsets the autonomy-undermining evolutionary urges to eat foods high in fat, sugar and salt.

Nudges can target type 1 or type 2 thought processing. Type 1 processing is automatic and reactive, whereas type 2 processing is reflective and deliberate. Generally, type 2 processing is associated with more autonomous behaviour because it involves deliberate thinking, and concern about autonomy tends to centre on type 1 processing. However, Levy argues that type 1 processing is rational thinking and that acting under the influence of type 1 nudges is like deferring to expert testimony. Levy argues that when we are given a default option, we interpret it as a recommendation. But Levy's argument is disanalogous. When we defer to testimony we do so deliberately, not automatically and reactively.

We may, however, deliberately allow ourselves to be influenced by type 1 nudges when the decision is one that is not important to us. When I allow myself to be influenced by the packaging or product placement of bread in a supermarket, I do so autonomously because I understand that it is not worth spending my cognitive resources on selecting the optimal type of bread. By voluntarily allowing ourselves to be influenced by type 1 nudges for certain decisions, we increase our autonomy, not because we are deferring to testimony, but because we free up cognitive resources for decisions about which we actually care. In this respect, type 1 nudges can be autonomy-friendly.

Hansen and Jespersen use an axis to visualise nudges that target differing degrees of type 1 or type 2 processing, and have differing degrees of transparency, to understand their impacts on autonomy. Accordingly, I analyse the impact of four types of nudges: transparent type 2 nudges, non-transparent type 2 nudges, transparent type 1 nudges, non-transparent type 1 nudges.

Hansen and Jespersen deem only non-transparent type 2 nudges to be manipulative because their non-transparency makes it difficult for agents to recognise that they are being influenced and because the type-2 processing and reasoning that an agent might otherwise use to overcome the nudge is under the nudge's influence. However, I argue that we can resist the influence of a non-transparent type 2 nudge. This is because we are still able to use type 2 processing to reflect on our information and deliberate on our decision. A doctor's framing of the risk of a surgery as 95% chance of success non-transparently influences us via type 2 processing to view the option of surgery more positively than if the doctor had said it had a 5% chance of failure. But this framing does not preclude us from using our type 2 processing to recognise that a 95% chance of success means there is a 5% chance of failure.

This analysis of nudge theory provides useful insights on how we may bioenhance without undermining autonomy. While nudging is not necessarily harmless to autonomy, considerations

of influence, transparency, thought processing and our bionudgee, enable autonomy-friendly, and even autonomy-promoting, forms of bionudging.

This paper demonstrates that even voluntary bioenhancement poses a threat to self-rule. This is the case when the impacts of the bioenhancement undermine the capacities required for deliberation, or the likelihood that a person engages them. However, bioenhancements may promote self-rule where they increase a person's ability or inclination to autonomously deliberate or act in accordance with their higher order/more coherent desires. Ethics boards, policy makers and companies can use this analysis of bionudging to develop strategies of bioenhancing that mitigate threats to autonomy and instead promote it.

Chapter 2. Theories of autonomy and their relationships with bioenhancement

2.1. Understanding and defining autonomy

Autonomy is about controlling yourself, from reaching your decision to acting on that decision. It may seem obvious that we control ourselves very well in this respect. But we see arguments about how much a person is autonomous arise in many debates, from criminal justice to medicine, from navigating relationships to navigating democracy.

The term autonomy derives from Ancient Greek 'auto', meaning 'self' and 'nomos', meaning 'rule' or 'government' (Sneddon, 3). To be autonomous is to rule oneself, rather than to be unruled or ruled by others. Originally this term was used to refer to politically sovereign states in Ancient Greece (Feinberg, 27). Besides the political use, 'autonomy' is now often used to describe an individual's ability to exercise control over themselves. Autonomy is contrasted with 'heteronomy', which means being controlled by someone other than oneself (Sneddon, 3). One can also be uncontrolled, or controlled by natural non-personal forces, which is called *oudenonomy* and *cosmonomy*, respectively (Sneddon, 3). One can observe expressions of undermined autonomy in regular conversation. For instance, 'she made me do it' or 'it was the alcohol; not me', are examples that reflect an intuition that our decisions and actions are not always maximally under our own control. These definitions of different types of control are important because of that intuition, and so if we are rational beings with ends that we seek to achieve, we may take interest in the kinds of ways we can maximise self-rule.

Within this concept of personal autonomy, the term 'autonomy' is used in a variety of different, albeit similar and often overlapping, ways. There is disagreement between scholars about 'correct' definitions and conceptions of autonomy. There are four main kinds of uses for the term 'autonomy'.

First, autonomy can mean the capacity to exercise self-rule (Feinberg, 1986, 28). Much like one's capacity to sing, the capacity to self-rule varies by a matter of degrees. One can be more or less autonomous in this sense, where autonomy is one's capacity to contemplate decision-making. For instance, in order to autonomously purchase a house, one needs various capacities such as the ability to reflect on one's desires and values, to assess the house location and amenities in relation to their desires and values, assess the cost of the house, and consider long-term needs and desires in relation to the house's features, their future income and mortgage repayments. If we remove the persons' ability to assess the house location, we *reduce* the degree to which they are able to autonomously purchase the house. If we educated the person to be able to understand the property market and anticipate its future, then we would be increasing the degree to which they autonomously purchase the house.

This differs slightly from a second kind of use, where autonomy refers to the 'actual condition' of self-rule (Feinberg, 1986, 31). A person may have, or may typically have, the capacity for exercising autonomous deliberation, but their ability to exercise that capacity may be restricted

by certain circumstances. A person may be considered autonomous on the first kind of use because they possess capacities such as reasoning and reflection, but their actual autonomy may be undermined in the second kind of use, because they are drunk or are being manipulated.

Feinberg notes a third kind of use, where autonomy is referred to in the respect that it is a morally desirable condition to possess (Feinberg, 1986, 44-5). Philosophers may refer to autonomy as a property which makes part of an ideal character and use the term as a synonym for similar, yet different, desirable properties of character like rationality or independence (Feinberg, 46). As we will see later, autonomy generally involves elements of rationality and independence. However, some accounts permit the possibility of irrational yet autonomous decisions and externally controlled but still autonomous decisions, so on some accounts neither rationality or independence are necessary for autonomy.

A fourth kind of use of the term 'autonomy' refers to an individual's sovereign authority to self-rule, that is one's right to be self-ruling (Feinberg, 1986, 47-9). Like the third kind of use, this one possesses a more explicitly normative aspect; where the first and second uses are slightly more descriptive in their conception of self-rule. Granted, the first use of the term 'autonomy' is going to be somewhat normative in that answers to normative questions – where autonomy is a key factor – are going to depend on what kind of conception of the capacity for autonomy is used. But questions at the centre of this first type of use are fundamentally descriptive rather than evaluative. These kinds of questions include 'what kinds of influences on decision-making reduce the degree to which it is self-ruled'; and 'what methods can we use to increase the degree to which a decision is self-ruled'. But the fourth type of use of the term 'autonomy' is more evaluative because the question at the centre of its discussion is 'what makes an agent *sufficiently* autonomous'. This question is vague; one is inclined to ask 'sufficiently autonomous for what?' Answering this question requires some normative or value judgement. What degree of autonomy is sufficient will vary. Consenting to medical procedures likely requires a much higher level of autonomy in order to be considered sufficiently autonomous, than consenting to cookies on a new website. Asking what makes a person a sovereign authority is a normative question whose answer will include some reference to the first kind of use (their capacity to self-rule), but it will assert some normative standard where a person is autonomous enough to be allowed to self-rule.

These different kinds of uses do not need to be considered as competing answers for the 'correct' definition of 'autonomy', especially given that they often overlap (Brailsford, 30-1). Rather, we ought to recognise that there will be a best kind of use depending on what kind of discussion taking place. In this paper, I mainly use the first and second uses of the term as I discuss the ways in which we might generally have the capacity to self-rule, and how certain enhancements or circumstances would increase or decrease the extent to which one actually deliberates autonomously. I use the fourth kind of use in *Chapter 2. Theories of autonomy and their relationships with bioenhancement*, as I analyse what different conceptions of autonomy set as the minimum standard for considering some decision or person autonomous and in 4.6. *Influencing type 1 processing threatens autonomy more than influencing type 2 processing*, as I analyse how a person may deem certain trivial decisions unworthy of lengthy and comprehensive deliberation.

Other terms are used as synonyms for autonomy or aspects of autonomy. This affects how autonomy is understood and can define the implications of an argument if used without care. Philosophers might use terms like choice, liberty/freedom, and rationality, when examining autonomy but these terms are not perfectly synonymous with autonomy (Brailsford, 42-3).

Choice is not synonymous with autonomy. While having one choice rather than many may undermine autonomy, too many choices can make decision-making too difficult and therefore undermine a person's capacity for self-rule; increasing choice does not necessarily increase autonomy (Brailsford, 45-6). Further, increasing autonomy in regard to choice is not just about reaching an optimal number of choices depending on the situation, because the types of choice also matter. If I am trying to decide what takeaway restaurant from which I want to eat, and assuming that roughly five choices of restaurant are optimal, being presented with five cheap pizza restaurants is unlikely to increase my autonomy as much as it would being presented with five different types of cuisine.

Likewise, liberty and freedom are not quite the same as autonomy (Brailsford, 47). The story of Odysseus and the sirens is a classic example of how more freedom does not necessarily equate to more autonomy (Colburn and Irvin, 72-3). In the tale, Odysseus asks his shipmates to block their ears, and to tie him to the mast so that he can listen to the sirens' entrancing songs without losing control and throwing himself overboard. Odysseus' liberty is reduced because he restrained and cannot move. But because Odysseus truly does not want to throw himself overboard, he maximises his autonomy by being able to satisfy both his desire to hear the sirens and his desire to continue his journey home.

Further, the concept of rationality possesses a lot of overlap with autonomy, but they are ultimately different concepts. While autonomy is about making decisions that are in accord with one's wishes, rationality is about making decisions that are in accord with one's interests. In simple terms, autonomy is about making one's own decisions, and rationality is about doing what is best for yourself. Decisions are often either both autonomous and rational, or both non-autonomous and irrational, but there are exceptional cases where these concepts diverge. However, the degree to which these concepts are different varies depending on whether one endorses substantive or procedural autonomy, as I discuss in 2.5. Substantivism.

Given that these examples of how more choice and more freedom can, in fact, reduce autonomy, and how rationality is not necessarily synonymous with autonomy, it is important avoid using the terms interchangeably with autonomy when analysing scholars' work and forming arguments.

I will now provide an overview of what the capacity of autonomy entails. While some actions we physically perform ourselves, the decision-making that led to that action can be less than entirely our own. Influences like coercion and manipulation are some ways by which other people can intentionally undermine another's autonomy; by threatening or deceiving someone with the intention that it will make them reach some decision or another, a person undermines another's autonomy. This is because the manipulator/coercer reduces the degree to which a person is able

to deliberate on what they really want to do, by altering the factors, or their perception of these factors, for each option. I revisit these kinds of intentional reductions of autonomy in 4.3. Manipulative and Transparent Nudges.

Autonomy can be undermined even when other people are not involved. Part of exercising your autonomy involves carefully considering whether you truly want to pursue a particular course of action and thinking about whether you have a conflicting desire that you would prefer to satisfy. But some states of mind are unlikely to result in such a thorough consideration of your desires. The intoxication and subsequent impulsiveness one experiences after drinking alcohol may influence a person to pursue a course of action without taking time to consider if they have any conflicting desires. Even though the decision-making was, in one sense, entirely their own (because no one else was involved in making that decision), in another sense it was not their own because it is not a decision they truly endorse. On the other hand, we can see how the effects of caffeine might increase person's capacity to focus and critically deliberate on their desires and decisions, thereby increasing their autonomy. However, some philosophers, as I shall discuss, endorse differing conceptions of autonomy, and so they take different positions on whether autonomy is hindered by reduced or degraded deliberation. I require a deeper understanding of the concept of autonomy in order to analyse how bioenhancement can impact self-rule.

2.2. Decisional and practical autonomy

Autonomy has a decisional dimension and a practical dimension (Pugh, 16). The decisional dimension pertains to an agent's decision-making and the practical dimension pertains to the agent's ability to act on those made decisions. The decisional dimension involves a cognitive aspect and a reflective aspect (Pugh, 16). Regarding the cognitive aspect, agents must have some beliefs and understanding about how pursuing each of their different options will satisfy their desires or accord with their personal values. For the reflective aspect, agents must consider whether they endorse those desires or values, which motivate those potential pursuits. Decision-making that includes these considerations is more autonomous than decision-making that omits them.

Bioenhancement can be used to increase our decisional and practical autonomy. Cognitive enhancers that increase inhibitory control and extend the time one spends reflecting on decisions, such as Methylphenidate and Modafinil, may help to improve the degree to which people reflect on their desires and values before making decisions (Roberts, 2020). Increased inhibitory control may increase practical autonomy by preventing people from being distracted or tempted away from acting on their made decisions. Bioenhancers like Donepezil that have been linked to improved episodic memory, may increase the likelihood that people recall the consequences of similar past decisions so they can apply their experiences to new decisions (Chuah, 2009). Further, bioenhancers that improve working memory and problem solving, like Methylphenidate, may improve the degree to which people recognise how those memories inform their present deliberation on selecting a desirable course of action (Linssen, 2014). While cognitive enhancers such as those in the papers cited above have not been shown to have profound effects on cognitive function, they show promise for research into drugs that do.

Practical autonomy, that is the ability to act on decisions, has an aspect of positive freedom and an aspect of negative freedom (Pugh, 16). Positive freedom refers to your freedom to act on your decisions in virtue of the abilities that you possess. For instance, you may be practically autonomous in your decision to join a rugby club because you are fit, and you are able to drive to the trainings and games. Negative freedom refers to your freedom to act on your decisions because of an absence of any forces or conditions that would limit your ability to act. A coach refusing to admit me to the club because the teams are already full would compromise my negative freedom.

In an enhancement context, practical autonomy is likely to be impacted when a bioenhancement directly influences a person's behaviour. I will discuss the different targets of enhancement in more detail during the next section. Practical autonomy will be impacted by behavioural bioenhancement because if an enhancement disposes people to behave in certain ways this will increase or decrease the likelihood that they actually act on their decisions. In everyday life there are situations where a person's practical, rather than decisional, autonomy is undermined. For instance, Isla might have carefully come to the decision that she will walk to work the following day instead of driving. But she may get into the car the next morning and pull out of the driveway because of an ingrained sub-conscious habit formed through months of taking the car to work. Enhancements that impact practical autonomy, without influencing decision-making, resemble this example. Pugh's decisional/practical distinctions and sub-categories give me a good starting point to analyse autonomy and enhancement in more depth, in section 3 Targets of Enhancement.

2.3. Local and global autonomy

Autonomy can be understood as being local or global. Local autonomy relates to small specific periods of time (like the night of a party), particular decisions (like the decision to buy a beer), or specific kinds of decisions made over indefinite periods of time (like decisions that are related to purchasing alcohol) (Pugh, 17). On the other hand, global autonomy refers to an agent's autonomy over the course of their life or about long-term projects (Pugh, 17).

Some may understand this distinction as resembling the division between the autonomy of particular choices and the autonomy of agents (Pugh, 4). We can view autonomy as a property of an agent, independent of any particular decisions or behaviours, where arguments focus on the degree to which an agent is autonomous. Alternatively, we can view autonomy as a property of decisions and behaviours.

Dworkin argues that it makes sense to conceive of autonomy only as a global property of an agent and not of particular decisions (Dworkin, 16). This is because, Dworkin argues, it is difficult to see how an agent could switch from being autonomous to nonautonomous in a matter of moments. Indeed, in understanding autonomy as being a property only of choices we may neglect the way in which psychological elements like character, values, and the agent's history affect the degree to which one autonomously deliberates. But Pugh views the autonomy of a choice as depending

on the autonomy of the agent in relation to that choice; autonomy of choice is about the autonomy of the agent, by extension (Pugh, 4).

I argue that the local/global distinction is useful. This is because it helps us understand, for instance, our intuition that someone struggling with an alcohol addiction can live a life generally autonomously, but can switch from autonomous decisions about what clothes to wear, to less-than-autonomous decisions about whether to drink alcohol. Further, it is useful to think about the autonomy of particular kinds of choices, because on some accounts, the level of autonomy required for a choice to be autonomous varies from choice to choice. That is, people tend to autonomously choose to opt-out of broad reflection and thorough deliberation for decisions that they deem trivial. For instance, the amount of information one acquires and the amount of reflection one carries out when making an autonomous decision in relation to dinner-plans is going to be significantly less than that of choosing a treatment plan for a serious illness. By focusing only on properties of the agent, we may ignore important aspects of the choices that impact how we assess autonomy.

In some cases, it is important to understand how an agent may or may not be globally autonomous. For instance, when government policy is being written that will affect people with cognitive disabilities or children, it is important that policy writers understand how and to what degree affected people can exercise autonomy, but not necessarily in regard to any particular kind of decision, rather their capacity to exercise autonomy in general. Understanding global autonomy is important because we may also wish to examine how local decisions can impact global autonomy or one's diachronic projects (Pugh, 18). For instance, the local decision to participate in a boxing match once might be considered autonomous, but a long career involving hundreds of autonomous local decisions to enter the ring might cumulatively impact a person's global autonomy by causing brain damage. On the other hand, we can sacrifice our local autonomy in order to increase our global autonomy (Pugh, 19). Understanding global autonomy is also important for understanding how global autonomy can impact local decisions (Sneddon, 47). For instance, some of the retired boxer's local decisions may be reached via an undermined autonomy because of their brain damage. Global autonomy helps us understand how the history of a decision, or rather how an agent's history and who they are, can impact whether a local decision is autonomous (Sneddon, 47).

Some bioenhancements can have short-term effects or affect only the ability to do a few discrete activities, and therefore they impact only local autonomy. For instance, the effects of many pharmaceutical enhancements last only a few hours. The potential control over very specific skills and abilities offered by future developments in brain chips, are an example of how enhancement technology can directly affect discrete local autonomy (Vassanelli, 2011). However, the implantation of a brain chip or the regular use of pharmaceuticals may offer more permanent effects than the one-off consumption of a pharmaceutical. A brain chip that improves one's capacity to solve math equations may improve one's local autonomy in regard to math-related decisions, but it may also improve one's global autonomy in regard to one's diachronic project becoming an engineer.

Someone might use an enhancement that reduces their local autonomy, in order to increase their global autonomy. For instance, recovering alcoholics are sometimes prescribed Disulfiram, and can take the drug so that if they succumb to their urge to drink any alcohol, they quickly experience unpleasant effects like nausea, vertigo, and heart palpitations (Stokes, Abdijadid, 2022). While taking the Disulfiram reduces the alcoholic's local autonomy, insofar as they struggle act on their desire to binge-drink, it also increases the alcoholic's global autonomy by helping them overcome an addiction that undermines their general capacities for decision-making. Understanding the distinction and interplay of global and local autonomy helps us understand how, in practice, we can increase or decrease particular types of autonomy through the use of bioenhancement technology.

2.4. Substantive and procedural autonomy

One significant division between philosophers of autonomy lies in the split between substantive autonomy and procedural autonomy. Substantive autonomy requires that made decisions involve some content that adheres to a particular value, while procedural accounts require that an agent follows some particular method in reaching their decision (Pugh, 5; Friedman, 19; Sneddon, 20). Substantive autonomy includes all of the conditions of procedural autonomy, plus adherence to certain values of the 'ideal of autonomy' (Friedman, 20). Subsequently, Friedman places substantive autonomy higher on a spectrum of autonomy than proceduralism (Friedman, 20-1). In other words, meeting the standard of substantivism is going to require more autonomy than what is required to meet the standard of proceduralism. According to Friedman, the debate between substantivists and proceduralists concerns the minimum criteria for autonomy, with substantivism setting a higher standard (Friedman, 21). Indeed, substantivism is a harder threshold to meet simply given that it involves the same conditions that proceduralist autonomy requires plus extra conditions. However, if the values and ideals required by substantivism do not enhance one's capacity for self-rule, then it cannot truly be considered a higher standard of autonomy than proceduralism. If the values and the ideals of autonomy that substantive accounts require do not increase one's capacity to decide for themselves, and proceduralist requirements are the entirety of factors that affect one's degree of autonomy, then substantivism is not a higher standard of autonomy than proceduralism; it is just proceduralism plus non-autonomy-based conditions.

2.4.1. Substantivism

A generic example of a substantive account of autonomy is one supported by Kant, which I will briefly explain. Kant's understanding of personal autonomy resembles the ancient use of the term autonomy, as being a self-governing state (Johnson and Cureton). A state is autonomous when its citizens are unified by and ruled by a set of laws, laws that are chosen or endorsed by those same citizens (Johnson and Cureton). For Kant, autonomy is self-legislation, that is, following laws set by oneself. But in order for those laws to be truly one's own, they must come from one's own rational will; they must come purely from one's reason-based decision-making (Johnson and Cureton). This is because, for Kant to consider a decision to be autonomous it must be 'physically and psychologically unforced', which rules out natural forces in the brain such as urges,

obsessions and disordered thoughts (Johnson and Cureton). This is why rationality is so important in Kant's conception of autonomy. But because Kant argues that the rational will causes actions, there must be a universally valid law that connects the cause and effect. This law cannot be a natural law because that would make it an external force that governs one's rational will (Johnson and Cureton). Therefore, Kant argues that it must be a special law to which all rational wills adhere (Johnson and Cureton). All rational wills adhere to the same law freely because, as they are pure reason, they are independent from the particulars of their circumstances and free from the external and natural forces that would subvert their self-government. This law, 'the Categorical Imperative', holds that you should 'act as though the maxim of your action were to become by your will a universal natural law' (Johnson and Cureton). It is a moral law, because it applies to everyone and holds that they *should* act in a particular way. For Kant, decisions are autonomous only if they come from the rational will, which will be adhering to the categorical imperative. Therefore, normativity is embedded in Kant's conception of autonomy. For Kant, autonomy is substantive because one's decisions must come from the rational will, and our way of knowing whether it comes from the rational will is to examine whether the actual content of the decision aligns with the categorical imperative.

However, philosophers of bioethics argue that using substantive accounts such as Kant's risk a kind of elitism. This is because if the quality of one's autonomy depends on the degree to which one can reason, then we may value some people's autonomy over others' and too easily resort to paternalism in cases where patients lack in some capacity for reason. Paternalism involves making decisions on a person's behalf, which are intended to be in the person's best interests (Beauchamp and Childress, 176). Given that children and some people with mental disabilities may lack some capacity for reasoning and rationality, they might not be considered sufficiently autonomous to make their own medical decisions on substantive accounts of autonomy (Cauffman and Steinberg; Beauchamp and Childress, 177, 179). In these cases, medical professionals might intervene on behalf of the patient and make decisions that they believe are in the patient's best interests (Beauchamp and Childress, 176). These paternalistic interventions risk elitism because the agent's decisions are dismissed and their autonomy is deemed insufficient because the values upon which the agent bases their decisions are not shared by the prevailing substantive account of autonomy and those applying it (Brailsford, 33-4).

Further, as I will discuss when I analyse 'hierarchicalism', it can be difficult for us to decide what is best for another person without imposing our own set of values onto that person who may not share them. This is a concern if policy makers, ethics boards and companies seek to ensure that bioenhancements accord with people's desires and values. One might argue that the risk of paternalism for those who lack some capacity for reason is a concern when proceduralism is the prevailing account of autonomy proceduralism, because those who lack reasoning skills may not fulfil the requirements for proceduralism. Certainly, it is true that those who lack reasoning skills are more likely, *ceteris paribus*, to fail the requirements of proceduralism than those who do not lack reasoning skills. But they are even more likely to fail the requirements of substantivism because substantive accounts that hold rationality as a central value for autonomy, such as Kant's decisions based on 'pure reason', require a very high standard of both the process and content of decision-making compared to proceduralism which holds only the process of the decision-

making to a high standard of rationality. Paternalistic intervention is therefore more likely to be required if substantivism is the standard of autonomy that patients must meet.

Alternatively, some philosophers state that agents lacking in the ability to reason can be autonomous in different ways (Brailsford, 32). These different kinds of autonomy fall under a conception of autonomy known as 'coherentist autonomy', which I will shortly discuss. As for Kantian accounts, deliberating and acting only via pure reason and refusing desires to hold any influence over decisions, is arguably an unattainable standard of decision-making for any human to achieve, in part because humans are not perfectly rational, but also because the coherence of our decisions with our desires is a basic element of what makes our decisions autonomous (Thaler and Sunstein, 7, 19; Brailsford, 33). The risk of elitism posed by using substantive accounts of autonomy exposes some of the ways in which substantivism fails to capture the concept of autonomy. It seems counterintuitive to conceive of autonomy – the capacity to make one's own decisions – as requiring that those decisions align with some particular value, if the reason that we value autonomy is because we expect people to make decisions based on their *own* set of preferences.

Some substantive accounts of autonomy hold that one cannot autonomously choose to relinquish their practical autonomy and enter a life of slavery, or kill themselves because these accounts require that the content of the decision must not eradicate their own autonomy (Pugh, 6). Because Kant's substantive account, for instance, holds that our autonomy produces our moral obligations, and given that we cannot universalise this act – we cannot rationally will that everyone enters into a life of slavery because there would be no free humans to own the unfree slaves – then entering into a life of slavery is inconsistent with our rational will, which is to say that it is inconsistent with our autonomy (Pugh, 4). According to Kant, those actions that eradicate our own autonomy are incompatible with our autonomous rationality.

I have examined how one might sacrifice local autonomy to increase global autonomy, but there are also cases where we are rational to sacrifice our own capacity for autonomy, in order to secure or promote another value altogether. For example, it is conceivable that a person could rationally and autonomously decide to sacrifice their future capacity to make their own decisions to avoid terrible pain. When a person opts for general, rather than local, anaesthetic before a minor surgery, they are sacrificing their capacity to make their own decisions for the period of time that they are unconscious. They do so, not to increase their autonomy, but rather in order to pursue another value, that is the value of avoiding pain. One might argue that if the surgery fixes something broken or failing in the human body, undergoing the surgery involves a temporary sacrifice of autonomy in order to maximise autonomy afterwards. Autonomy is not really being sacrificed because the surgery aims to increase the patient's autonomy by healing them. But this issue is about general anaesthetic versus local anaesthetic; the patient can still get the autonomy-increasing benefits of the surgery without sacrificing their capacity for self-rule if they opt for local anaesthetic which only partly numbs the area of the body undergoing the surgery. The local anaesthetic would not eliminate the pain and would keep the patient's mind clear enough to make self-ruled decisions. In spite of this option, it seems rational that the patient should eliminate their pain during the surgery by selecting a general anaesthetic, if possible, because they can get the

benefits of the surgery without any of the pain. Therefore, it is conceivable that we can be rationally self-ruling in sacrificing our autonomy in favour of a different value. But this example involves only a temporary sacrifice of autonomy in favour of a different value, not a permanent eradication of autonomy. This example of a temporary sacrifice of autonomy however, holds as an analogy for the permanent sacrifice of autonomy via euthanasia. It seems possible that a person can be rational and self-ruling in choosing to end their own life prematurely, in order to avoid the terrible pain of a debilitating disease.

Some discussions of voluntarily surrendering one's autonomy risk making autonomy an evaluative term, and associating it with morally desirable decisions (Brailsford, 31, 33). For instance, people are unlikely to say that a soldier who dives on a grenade to save his buddies or a security guard who takes a bullet for their prime minister, does not do so on their own decision. Typically, entering a life of servitude is not a decision which is made autonomously but it certainly seems logically possible for this decision to be made autonomously. For instance, a person might choose to preserve some small degree of autonomy by entering a life of slavery, because the alternative is death. This demonstrates how a person can autonomously choose to profoundly but rationally undermine our own personal autonomy, contrary to popular accounts of substantive autonomy.

This is important to note because some bioenhancements may undermine a person's capacity for, or engagement of, autonomous deliberation. But the decision to use these bioenhancements may still be considered autonomous. This will be relevant when I consider how people may autonomously choose to allow themselves to be nudged toward selecting certain options, in decisions which they deem unworthy of deliberation.

If we were seeking to enhance a human's capacities without undermining their autonomy on a substantive account, then we would need to ensure that the enhancement fulfils two key conditions. First, the enhancement must not undermine the capacities, or the likelihood of engaging the capacities, required for procedural autonomy. This is because such an enhancement would reduce the likelihood that a person procedural requirements of substantivism. Second, the decisions and behaviours that the enhancement influences must not conflict with the value/s central to whatever account of substantivism we endorse. If an account of substantivism holds rationality as a central value of autonomy, then enhancements that do not reduce the likelihood that a person reaches and acts on the decision that best satisfies their interests, will not harm their autonomy. Enhancements that increase one's ability and inclination to use reason will promote one's capacity for self-rule on substantive accounts.

However, there are other accounts of autonomy besides substantivism that do not depend so heavily on one's ability to reason, and do not suffer from the counterarguments above (Pugh, 5-6). Because the promotion of one's personal autonomy and rational decision-making are so important for key substantive accounts of autonomy, and because we can identify cases where it is rational to not only undermine but indeed eradicate our personal autonomy, there is good reason to doubt the strength of substantivism in capturing what it means to be autonomous. Further, if those who are concerned about the elitism of invoking substantive accounts of

autonomy in practice can conceive of ways to express self-rule without using rational decision-making, then we should consider those alternative accounts of autonomy too.

2.4.2. Proceduralism

In proceduralist accounts of autonomy it is the method, not the content, of decision-making that matters. For instance, a procedural account of autonomy might include factors like ensuring that the agent understands their options, takes enough time to reflect on their desires and broader values, and that their decision is not controlled by someone else. These processes may not be required for the decision to be autonomous if the agent (procedurally) autonomously opts out of practising them, as I discuss in 4.6. *Influencing type 1 processing threatens autonomy more than influencing type 2 processing*. Further, while substantive accounts of autonomy require that one's decisions adhere to some particular value, proceduralist accounts state that 'part of being autonomous is choosing to act in accordance with one's own beliefs about value, even if those beliefs are not universally shared' (Pugh, 6). In this way, proceduralism is able to capture something important about the concept of self-rule, that the ends of an autonomous agent's decision-making are selected by the agent themselves and are not bound by normativity.

If we were to enhance a person in a way that does not undermine their autonomy on a procedural account, then we would need to ensure that the enhancement does not reduce the degree to which that the person carries out their deliberation and reflection. For instance, alcohol may be considered a bioenhancer for improved social interaction, but it can undermine autonomy if, as a result of drinking, people do not consider the consequences of their made decisions or fail to reflect on whether their made decisions align with their values and goals. On the other hand, there are a number of pharmaceutical enhancers that may actually increase the degree to which people carry out their deliberation. For example, Methylphenidate, otherwise known as Ritalin, has been linked to improved working memory, which supports 'extensive processing and thoughtful evaluation of information about options' for complex decisions (Del Missier, et al. 2013). Additionally, Modafinil, a sleep disorder treatment, has been shown to increase focus and sustained attention in normally functioning adults (Battleday, et al. p.1867). When making complex decisions, this study found that Modafinil may make it easier for an agent to engage in and stay focused on deliberating over a decision. Finally, Modafinil was observed to improve fluid intelligence, that is flexibility and speed in solving novel challenges and being able to identify relationships between different pieces of acquired knowledge (Battleday, et al. p. 1874). Bioenhancements that improve cognitive skills such as these would also increase autonomy, where autonomy is understood procedurally.

One problem with proceduralist accounts is that they allow for autonomous decisions to include decisions based on values or desires which come from external controlling sources (Pugh, 5-6). For instance, some feminist philosophers consider some decisions from women, despite adhering to the criteria of procedural autonomy, as not being self-ruled because the desires and values that motivate those decisions come from an oppressive patriarchal society (Pugh, 5-6). This is a good reason in favour of using substantive accounts of autonomy, because substantive accounts can better explain how it is that people who are indoctrinated into oppressive or fundamentalist

communities are not autonomous when they make decisions based on values that are not really their own. These values might not be truly their own if those people have been enculturated in a way that ensures that they never have a genuine opportunity to endorse or reject those values. According to Pugh, proceduralism does not give us a means of excluding desires that arise from any societal norm from being considered autonomous (Pugh, 5-6).

However, proceduralism can still account for external influences by requiring that agents critically examine their own values and desires in relation to their society's norms, consider alternatives, and decide whether to accept, modify, or reject those values and desires. It is virtually impossible to avoid being influenced to some degree by the values of one's society, but it is possible to reject these values if they can be critically examined. This solution allows us to say that those who grow up in nationalist regimes or patriarchal societies, who have never been taught to analyse or question norms, are not autonomous in regard to certain decisions.

However, this may be too great a demand on agents as it requires them to consider abstract concepts like societal values before every decision. But reflecting on societal values may be necessary only a few times during one's life, provided that they modify their values according to their judgements. This reflection would be required infrequently because one is unlikely to change such deeply held set of endorsed values, and one's judgements of those norms are unlikely to require reassessment before making day-to-day decisions.

Another problem with proceduralism is that a person may fulfil the criteria for procedural autonomy and make decisions that are uninfluenced by values they receive from their culture, but still do not seem consistent with what the person truly wants. We can imagine Steve, who wants to skate down a very steep hill without wearing a helmet. Steve has reflected on his desires and his values, including the desire to avoid brain damage and death, he has considered and learned about the different ways he might satisfy those desires, and he has taken time to learn about the risks associated with downhill skating and the role that helmets play in negating some of that risk. Yet Steve still chooses to skate without a helmet. Despite practising procedural autonomy, there is no guarantee that Steve will realise that wearing a helmet to curtail the risk of head injury is what he truly wants to do. There is no requirement that an agent actually understands the aspects of their decision under procedural autonomy, it only requires that they practise a consideration of those aspects. For most people it is unlikely that an earnest consideration of each part of the procedural decision-making will result in such a poor decision, but some people will carry out the required procedure and still fail to recognise how their chosen actions do not align with their real desires and proceduralism is impotent in accounting for this. For substantivists, this example shows how substantivism can capture what the agent *truly* wants, where proceduralism cannot. For proceduralists, this example shows that Steve *truly* wants to skate without a helmet, but is irrational for wanting to do so; for proceduralists it illustrates the point where autonomy and rationality come apart.

Others argue that proceduralism does not allow us to say that those with mental disorders that distort and alter our desires, like anorexia or drug addiction, are non-autonomous in relation to those relevant decisions (Pugh, 6). People with drug addiction may satisfy conditions of

procedural autonomy, they may have a good understanding of the way that using the drug damages their body, their mind and their relationships, they may reflect on their desires and values and continue to use the drug anyway. They are considered procedurally autonomous even though the effects of the drug are an external influence which has rearranged the addict's set of desires.

A proceduralist might be inclined to bite the bullet and say that while addicts and those with eating disorders are unlikely to satisfy the conditions for procedural autonomy, it is still possible for them to do so and to still choose to use drugs or carry out disordered eating. After all, despite there being some overlap between the two, rationality and autonomy are still distinct concepts, and proceduralists hold that people can make self-ruled decisions that are not in their best interests (Pugh, 6-7).

The bitterness of biting this bullet can be sweetened somewhat by recognising that just because we view an action as autonomous does not preclude the public health sector or legislators from being justified in intervening for the sake of a different value, like wellbeing. I mention above how medical professionals may make decisions on a patient's behalf when the patient lacks autonomy in some respect. We might still be justified in permitting medical intervention to help drug addicts and anorexic people who are autonomously destroying themselves, but are doing so irrationally because their other interests are not being met, especially while we do not know whether these people are actually practising procedurally autonomous deliberation about their consumption. However, this does not reconcile with our intuition that drug addictions and eating disorders do undermine autonomy.

2.5. Hierarchicalism

In any case, an appropriate conception of autonomy should be able to explain how drug addiction and mental disorders can undermine autonomy, but is a conception that is not embedded with normativity. One other solution to the proceduralist problem above, is to invoke the hierarchical concept of higher order desires.

Higher order desires are desires about desires (Brailsford, 34; Sneddon, 28). An agent might have a first order desire to eat beef, but also a second order desire where the agent wishes they did not want to eat beef. The agent might analyse these desires and recognise that their desire to eat the beef comes from an urge they want to satiate, while the second order desire against wanting beef comes from having some personal moral commitment. Upon reflection, the agent might recognise that they endorse the moral commitment more strongly than the urge, which explains why they have this desire-focused desire. This idea of higher order desires explains how it is that a person could act in a way appears autonomous, despite their autonomy being severely undermined. The drug addict may be aware that using drugs will destroy their body and relationships, still desire to take drugs, but have a higher order desire which involves wishing they do not desire drugs. Notice how desires about desires avoids the difficulty of trying to weigh two conflicting desires against each other. If the addict has the desire to use drugs and the desire to

maintain good relationships with family, then it is more difficult for us to say that the addict is wrong to want one more than the other without imposing our own values.

If an addict has a desire against their desire to use drugs, the hierarchical model allows us to determine that their decision to use drugs is not autonomous. It is common to occasionally have desires that we did not choose and that we are unable to control. Though, we may wish that we did not have these desires if we feel alienated from them or feel like they do not truly reflect what we want. Second order desires are a way of understanding how it is that we can have an urge or desire which we do not endorse and is not authentically our own, *ipso facto* not autonomous. Because we can assess whether we truly want each of our desires, our second order desires will represent a wish that is more highly valued by the agent. It is still possible that an addict does not have a desire to stop desiring drugs, and if so, on this account they would be using drugs autonomously.

A critic might argue that the nature of the drug addiction is such that addicts do not have the ability to contemplate the relationship between their values and their desires. They do not and will not have the chance to properly examine whether they wish not to desire drugs, in virtue of the fact that their addiction compels them to avoid that consideration. But given that some addicts are still able to recognise that they have a second order desire against using drugs, addiction does not necessarily prevent a person from this kind of autonomous deliberation. Of those addicts who do not report having a second order desire to stop using drugs, some may not have the capacity to consider this properly because of the effects of drug use. But it is conceivable that some have the capacity to carry out this consideration, and then conclude that they do not have a second order desire to stop using drugs. It is not difficult to conceive of someone who has lost everything through their drug addiction, is nearing the end of their life, and now wishes only to spend their remaining time using the drug after reflecting on their second order desires. As regrettable as it is, some addicts are procedurally autonomous in their decisions to use drugs.

Further, hierarchicalism faces the same issue I have noted above regarding substantivism; we should be cautious when estimating the higher order desires of others (Brailsford, 37). Because of the nature of substantive accounts of autonomy, policy makers, for instance, who use those accounts may take paternalist positions toward agents whose decisions will harm themselves or reduce their future autonomy (Brailsford, 36). If we are going to analyse how 'enhancing' people's cognition affects their autonomy on a hierarchical account, we need to be very careful not to make false assumptions about their higher order desires. People's conceptions of values will vary both in nature and in the intensity to which they are valued (Brailsford, 37). Brailsford suggests that the problem with higher order desires is not that we may not have them, but rather that it is very difficult to ascertain what they are for oneself, let alone for the objective observer or would-be intervener (Brailsford, 37). This should give us pause when considering how enhancement might increase or decrease a person's autonomy, when using a hierarchical framework.

In terms of bioenhancement, any enhancers that make one more likely to act on their higher order desires will increase their autonomy, on a hierarchical account of self-rule. If people tend to act impulsively on the first order desires about which they have conflicting second order desires, then

pharmaceuticals such as those that improve inhibitory control will increase autonomy (Roberts, 2020).

In his examination of hierarchical autonomy, Sneddon is less concerned with what makes one capable of making decisions for oneself and more concerned with what makes autonomous decisions authentically one's own (Sneddon, 26-8). Sneddon analyses how we know that it is truly *the self* that is ruling, rather than analysing how it is that we are capable of making decisions for ourselves. He suggests that we might better understand the nature of autonomy through an analogy to epistemology, and in particular the epistemological practice of justification (Sneddon, 24-26). This practice involves justifying our beliefs with reference to other pieces of knowledge. I may know that there is lemonade in a glass, because I know that I poured liquid from the lemonade bottle into my glass moments ago. Likewise, from the hierarchical perspective, we can justify our first order desires with reference to second order desires. For instance, what makes my decision to play rugby authentically my own, is my second order desire to desire rugby which comes from my deeper and broader desire to improve my fitness.

However, hierarchical models of autonomy lead to infinite regression. If we want to know what makes our first order desires authentically our own, and we refer to second order desires to justify the first order desires, how do we know that our second order desires are authentically our own? In the analogy to epistemology, how can we be sure that my knowledge of pouring the lemonade in the glass earlier is actually knowledge and not some false belief? The lemonade bottle from which I poured the liquid may have been filled with something else. I may be inclined to identify a third order desire that justifies my second order desire to get fit. Sneddon says that our autonomous thoughts can be understood as existing in a chain or a ladder, with each thought justifying the one below it (Sneddon, 28). But this pushes the problem infinitely backwards. Hierarchical theories of autonomy run into this issue of infinite regression, where one needs third order thoughts to justify the authenticity of our second order thoughts, and fourth order thoughts to justify the third order thoughts and so on and so forth (Sneddon, 28). If there is no value at the top of the ladder which we can be certain is our own, then we cannot be sure of the authenticity of any of them.

Therefore, not only does hierarchicalism risk a harmful paternalism when used in practice, it also suffers from infinite regression in theory. But there is yet one more concern with hierarchicalism: it does not appear to give us an accurate model of the self. The nature of selfhood and thus of authentic self-rule, Sneddon argues, is constituted by a complex relationship between desires, thoughts and attitudes (Sneddon, 34). Visualising the self as a rigid ladder of mental properties seems to miss the complex and multi-relational nature of the self.

2.6. Coherentism

Alternatively, Sneddon examines coherentism, another theoretical approach to epistemic justification that might better align with our conception of authentic self-rule. In this model of epistemic justification, a belief could be justified by its logical relationship to several other beliefs. So, my belief (1) that the chairs around the corner of the library are red is a justified belief because

of my other beliefs that (2) all the other chairs in the library are red, and (3) that I have been told by someone who has been around the corner that the chairs are also red. This style of justification might also be useful for assessing whether our mental properties are authentically our own.

In this model of authentication, the self exists as a network of individual mental properties like thoughts, beliefs, attitudes and desires. It is a web rather than a ladder, where each node is a mental property that shares a relationship with other nodes, and where nodes closer to the centre of the web are going to be higher order mental properties, because they relate to so many other mental properties (Sneddon, 29-30). The self, rather than being constituted by some specific superseding mental properties, is instead constituted more holistically by the relationship between all of the mental properties.

This conception may better represent the relationship between desires and the self. My desire to play rugby is authentically my own, not because of one subsuming desire to improve my fitness, but because it coheres with many other mental properties of mine such as: positive attitudes toward team sports, my beliefs about what playing rugby consists of and about my capacity to play rugby, and other desires to make friends, improve fitness and binge drink on Saturday nights. There are likely many more mental properties that relate to my desire to play rugby, but this complexity represents how our desires are authentic and how we can justify that we do, in fact, want what we want. On this account, if a desire fits coherently with other desires and beliefs, then it is an authentic desire.

But coherentism struggles to determine what desire is more authentic when a person has a desire that they feel more motivated to fulfil than a conflicting desire with which their mental properties share more connections. Let's say that Joel really wants to start reading books, and he also wants to buy the newest instalment of a videogame franchise he already enjoys, but he only has enough money for one of those things. Joel's desire for the videogame is more his own than his desire to start reading books because it relates to more of his existing mental properties. This is because he has more beliefs and specific desires relating to the activity he knows more about. But it could still be the case that he *wants* to start reading books more than he wants the videogame, he just does not know as much about reading books as he does about this videogame. Under coherentism, it is unclear how Joel can best exercise his autonomy, when weighing more authentic desires against more personally valued desires.

Under coherentism, enhancements that make us more likely to act on our more coherent desires will make us more autonomous. Like the example above suggests, enhancements that make us more conservative with our decision-making, more likely to stick to options with which we are more familiar, will make us more autonomous on a coherentist account, *Ceteris Paribus*.

Like the hierarchical account, however, the coherentist position suffers from its own problem of circular logic (Sneddon, 30). Consider a very simple network of desires. Let's suppose that desire A is justified in part by desires B, C, D and E. Let's say that A is the desire to play rugby and B is the desire to improve fitness. A is justified because of B, and B is justified because of A. To know if A is authentic, we need to know that B is authentic, but to be sure that B is authentic, we need

to know that A is authentic and if we cannot find some valid independent justification, our network of desires suffers from a circularity of reasoning (Sneddon, 30). Sneddon argues that there is no circularity, because given that higher order desires are closer to the centre of the web, central mental properties confer autonomy onto the wider mental properties to which they are connected (Sneddon, 31). On the other hand, first order mental states on the edges of the web do not confer autonomy onto the more central mental properties. The conferring of autonomy goes one-way and so there is no circularity of justification.

Further, and more powerfully, Sneddon argues that regardless of any issue of circular logic, the analogy to epistemology is inaccurate and inappropriate for understanding autonomy (Sneddon, 31). Epistemic coherentism justifies one's single belief in terms of a network of beliefs. The issue of circular justification is quite rightly a threat to the justification of beliefs and the possession of knowledge. Whereas, the coherentist model of autonomy does not only justify the authenticity of a mental property by referring to other authentic mental properties, but also by referring to the self as it is conceived of as a web of mental properties. As Sneddon puts it 'a first-order desire should not be construed as autonomous on the basis of its relation to another autonomous mental state, or even to a web of autonomous mental states. It should be construed as autonomous because of its relation to the self', that is, the web as a whole (Sneddon, 31). This is in contrast to the hierarchicalist account, which does not offer the ladders of desires as conception of the self, but rather simply a series of desires, the highest of which the self endorses, but from which the self is entirely independent.

One problem is that, if the coherentist model is correct, then if a neuroscientist was able to surgically implant a desire, and that desire fits coherently with the web of mental states, it would be considered an autonomous desire, despite being heteronomously implanted (Sneddon, 34). For instance, if I have no desire to play football, then upon implanting the desire to play football it would cohere with other mental states, like my desires to get fit and make friends, my general positive attitude towards team sports, and my beliefs about football and my ability to play it. Even though I would otherwise have had no such desire, this implanted desire would be considered autonomous.

This might not necessarily conflict with our intuitions about autonomy. After all, is there any relevant difference here between surgical implantation and, what we would not typically deem autonomy-undermining, a friend encouraging you to play football? Perhaps, encouragement to form a desire allows the agent more opportunity for deliberation than a surgically implanted desire. Nonetheless, let's suppose that the neuroscientist implants a desire which does *not* cohere with the web of desires, that is, the desire to play golf. But let's also suppose that the neuroscientist implants a series of other desires which make the desire to play golf cohere (Sneddon, 34-5). They might implant the desire to play a sport which one can practise and play on one's own, a positive attitude towards slow-paced individual sports, and certain beliefs about how golf works as a sport. Any decisions or actions that result from this desire, like playing golf or buying golf clubs, would surely be heteronomously controlled, and thus not autonomous, despite cohering in the network which constitutes the self.

Further, at some point, the neuroscientist's alteration of my web of mental states will constitute a changing of my self, of my identity, and this represents a profound threat to my autonomy. Too much alteration might mean not that I would be acting on a compromised autonomy, but that I would have ceased to exist and have become someone else altogether. If the self is constituted by this web of mental properties, and all of these properties are taken out and replaced with others, then the resulting person is certainly not the same as the former. The decisions of this person cannot be said to belong to the former person.

But our web of mental states may change dramatically over the course of just a few years involving influential events. We typically do not consider this a literal destruction of the original self because these changes are likely to be autonomous insofar as we endorse or accept them. But for those changes that I would not endorse, the change caused by a descent into a drug addiction for instance, it is possible that my network could be changed so much that I cease to exist. Likewise, if a person engages in bioenhancement to such a degree that they fundamentally change who they are, and do so without both an understanding and an endorsement of the changes, then they may compromise their identity, effectively eliminating their own autonomy.

Although, controls on our desires and decisions, such as that of the neuroscientist, are not necessarily a threat to autonomy if they do not destroy one's identity. For instance, a person's decision to marry might be highly controlled by their normal but intense love for their partner, but because the source of the control is not alienating for the agent, the decision is still autonomous (Pugh, 11). Because an agent loves another person so deeply, the agent's decision to marry them is essentially controlled in a deterministic sense because they could not have chosen otherwise. However, the controlling nature of their love is not necessarily undermining for their autonomy because the agent endorses the source of the control and how it influences their decision-making.

Walker suggests that just because an action is controlled in some sense, does not preclude it from being autonomous (Walker, 602). Instead, what might matter is what the source of control is and whether we accept and endorse it (Walker, 602). For instance, a drug addict may reflect on their desires to use drugs, and decide to use drugs, and still feel some kind of alienation from the control exerted by the drugs over their desires, and so the addict, unlike the lover, has a compromised autonomy.

Let's examine how this discussion plays out in an enhancement example. Sharma wants to enhance his ability to build good habits for climate change prevention. He has one desire to prevent climate change and also the belief that climate change is likely to cause severe issues in the future. On the coherentist model this desire is not currently autonomous because Sharma has no other cohering desires such as the desires to change diet, limit plastic consumption and vehicle use, and lobby the government to establish regulations, and the beliefs and attitudes that would relate to these desires.

But if a neuroscientist implanted these related desires, without Sharma's knowledge, so that the desire to prevent climate change was coherent, the desire to prevent climate change would no longer be incoherent. If Sharma had second order desires that resemble 'I want to desire [doing

all of those things that prevent climate change]’ then the neuroscientist’s intervention might feasibly be considered autonomy-increasing on coherentist accounts, rather than undermining. But to have that second order desire, one might say that Sharma would need to already possess those related beliefs and attitudes toward the behaviours required to prevent climate change. This is because it might be difficult to see how one could want to desire things of which they are not aware, or do not understand. But I could want to perform heart surgery, know nothing of how to perform heart surgery, and I may not desire to carry out the learning; yet I could still want to desire to learn how to perform heart surgery. This kind of second order desire is possible, and so likewise for Sharma having second order desires about things he does not entirely understand. The implantation of these eco-desires would not undermine Sharma’s autonomy because they would cohere with that second order desire.

But if Sharma does not possess the second order desire and instead only possesses the first order desire to prevent climate change, whether the implantation of related desires and beliefs undermines Sharma’s autonomy would depend on whether he feels alienated from the decisions that they produce. To explain this, consider a slightly different scenario. Like the last example, the neuroscientist plans to implant the mental properties relating to climate change prevention without Sharma’s knowledge. In this example, the neuroscientist waits to tell Sharma about the implantation until after Sharma has spent a week doing the things required to prevent climate change. If, upon being told about the secret implantation, Sharma would suddenly feel alienated from his decisions – if he would consciously reject those new desires – then those decisions would not be autonomous and the neuroscientist would have undermined, rather than supported Sharma’s autonomy. This is because, without other desires against which we can assess the coherence of implanted desires, we must defer to an agent’s self-appraisal.

2.6.1. Coherent integration

If the neuroscientist implants the desires and never tells Sharma, then Sharma may never have the opportunity or the prompt to analyse, and potentially modify or reject those desires. We could ensure that we are not mistaken about Sharma’s autonomy by telling him about the implanted desire and seeing if Sharma endorses the same behaviours after deliberating. But according to Sneddon, the fact that the desire was implanted, alone, does not necessarily mean that the subsequent decisions are not autonomous (Sneddon, 25, 35-6).

Sneddon specifies this way of understanding coherentist autonomy as ‘integration’ (Sneddon, 38). The earlier implantation of the desire to play football would integrate with my other desires, and upon considering this desire, I would probably possess a second order endorsement of this first-order desire. Whereas the implantation of the desire to play golf would not, because it conflicts with my other desires, and I would probably feel some alienation or disavowal toward that desire.

Sneddon distinguishes functional-integration from content-integration (Sneddon 38-9). My implanted desire to play football would be content-integrated because the content of the desire shares similarities with my other first order desires. For instance, football is a ball sport with lots of kicking and passing and so an implanted desire to play football would resemble my desire to

play rugby. But sharing a similar content and thus being 'content integrated' is insufficient to make the implanted desire as autonomous; it must be functionally-integrated. A desire being functionally integrated means that the desire integrates with the other desires by way of causal relationships rather than by a similarity of content. So my implanted desire to play football could be considered autonomous because it shares relevant causal relations with my desires to get fit, play a team sport and make friends; it integrates causally into my web of desires. Sneddon argues that functional-integration is not only necessary for autonomy, but sufficient (Sneddon, 39).

But Sneddon's claim that functional-integration is sufficient for autonomy, misses an important procedural aspect of autonomy. If a neuroscientist implants a functionally coherent desire that I would never have otherwise had, and I am never made aware of this desire and do not assess whether I endorse it, then the desire lacks a degree of autonomy. The implanted desire may not significantly undermine my autonomy because it coheres with my set of desires, but it does not increase my autonomy, except insofar as it helps me do something that I endorse, and so it is unreasonable to say that functional-integration is all that is required for a desire to be one's own. If I have not assessed whether I endorse this new desire or given consent to have this desire implanted, then the desire is not entirely my own and acting upon it is not fully autonomous.

2.7. Conclusion

This section has laid out different conceptions and issues with theories of autonomy, and on this basis, how bioenhancement may impact autonomy. I have examined the frameworks that I will use to analyse the impact on autonomy of bioenhancements that target cognition, emotions or behaviour, or that take the form of a bionudge. Understanding the practical and decisional distinction is important for understanding how autonomy is affected in different ways when cognition, emotions or behaviour is enhanced. Recognising that desires can be of a higher order or more coherently integrated within one's identity helps us to understand how behavioural bionudges that intervene on a person's actions can be consistent with their autonomy. Understanding the difference between substantivism and proceduralism informs assessments of the impact of cognitive enhancement on one's reasoning and deliberation, and thus on their autonomy. This analysis of autonomy offers a foundation from which we can assess more specific forms of bioenhancement, namely cognitive, emotional, and behavioural bioenhancement, and bionudges.

Chapter 3. Using bioenhancement to alter cognition, emotions and behaviour

3.1. Introduction

In this section I set out how autonomy is impacted by bioenhancement depending on the type of human capacity that the bioenhancement targets. I focus on three primary targets of bioenhancement: cognitive enhancement, emotional enhancement and behavioural enhancement. While some scholars of enhancement assess the ethics of enhancing phenomena such as desires, preferences, dispositions and motivation, I have chosen to focus on cognition, emotion and behaviour (Jebari, 2014; Jotterand and Levin, 2019). This is because it is easier to distinguish cognition, emotions and behaviour as distinct aspects of a person than the alternatives listed above, as I shall soon explain.

While other scholars' analyses cover a broad ethical assessment of enhancement, my analysis focuses solely on the impact that enhancements have on personal autonomy. My focus on cognition, emotion and behaviour lends itself to this kind of analysis because these phenomena overlap conceptually with the alternative phenomena above to some extent. For instance, we can reasonably expect little meaningful difference in the impact on autonomy between enhanced motivations and enhanced preferences.

A lot of research has focused on the ethics of using specific technologies to enhance human capacities and there is a lot of scholarly work that examines the wider potential impacts of having enhanced capacities (Crutchfield, 2016; Savulescu and Persson, 2014; Sparrow, 2014). But there are few ethical assessments of the targeting of specific capacities, and still fewer assessments of how altering them impacts personal autonomy. The following analysis examines how autonomy is impacted when one enhances cognition, emotions, or behaviour, respectively. Here, autonomy consists of two distinctions:

- (1) decisional and practical autonomy, where the former involves one's ability to reach a decision and the latter is one's ability to act on the decision, and
- (2) two forms of authenticity, that is,
 - (2a) identity, such as the thoughts, feelings and actions a person typically has/does, and
 - (2b) the degree to which they identify and relate to their thoughts, feelings and actions.

I use my analysis of autonomy above to understand how targeting these different aspects impacts autonomy.

3.2. Defining cognitive, emotional and behavioural bioenhancement

In Bostrom and Roache's terms, cognitive enhancement involves improving 'capacities that we use for gaining, processing, storing, and retrieving information' (Bostrom and Roache, 14). For the purposes of my analysis, this can include specific conscious aspects of capacities including the aspects of attention, such as selective attention, attentiveness and attention span (Matthews

and Wells, 1999, pg. 171-2); various aspects of memory and recall such as memory consolidation, and visuospatial, numerical, and working memory (Funayama, et al. 2015, pg. 255; Squire, et al. 2015, pg. 1; Baddeley, 1992, pg.556-7); reasoning ability, such as problem solving, case-based reasoning, memory-based reasoning, analogy-based reasoning, exemplar-based reasoning and instance-based reasoning and the speed at which reasoning is carried out (Corchado, 1996). As I have noted, there is burgeoning research which examines how various pharmaceuticals impact these capacities and from my discussion in the previous section, we can see how an agent requires these cognitive capacities to exercise decisional autonomy.

Emotional enhancement involves changing the ways we feel about particular behaviours, activities, persons or concepts (Jebari, 255), or altering what we feel, in general. Enhancements of emotions can include changing the intensity, duration and kinds of emotions a person feels. Further, emotional enhancement could include increasing the awareness and control a person possesses over their emotions. Research of the enhancement of emotions often centers on limiting emotions like anger/aggression and the promotion of feelings such as trust and happiness (Jebari, 255). Anti-anxiety medication and anti-depressants are effectively forms of emotional enhancement pharmaceuticals already being widely used. While there may be disagreement about whether the use of these drugs constitutes a restorative treatment that returns a person to normal functioning, as opposed to an enhancement, an examination of the “enhancement vs treatment” distinction is not necessary for the purposes of my argument (Daniels, 2000).

Behavioural enhancement involves changing a person’s behaviour without altering their emotions or cognition. It may be difficult to see how a person’s actions can be influenced independently from their emotions or cognition, however some people’s typical behaviours are not the result of affective or cognitive states. The example of Isla acting out her unconscious habit of driving her car to work demonstrates how people can act in conflict with their emotions and cognitive decision-making. Other such behaviours include involuntary flinching, mirroring mannerisms in social interactions and the kinds of behaviours induced by type-1 nudges that I will discuss in the following section on Nudge Theory. A crude but clear example is the use of electrodes that stimulate muscles, forcing them to contract, thereby moving body parts without influencing emotions or cognition. Jebari notes that behavioural bioenhancement technology might include a brain chip that ‘intervenes’ at the point in which a person has decided to execute an undesirable action, and prevent them from doing so by releasing counteracting chemicals into the brain or bloodstream, or via electrical stimuli (Jebari, 255).

3.3. Assessing different targets of bioenhancement against decisional and practical autonomy

In this sub-section I analyse the impact that enhancing cognition, emotions and behaviour will have on autonomy. In doing so, I use the decisional/practical distinction from the previous section, where the decisional part of autonomy involves the deliberation, judgement and selection of a course of action, and the practical element is the ability to act on that made decision. Using this distinction gives us a lens through which we can assess the variable impacts of cognitive, emotional and behavioural enhancement.

3.3.1. Cognitive bioenhancement

One's capacity to self-rule can be positively impacted by cognitive enhancement. If one is better able to gain, process, store, and retrieve information, or better able to reason about consequences and weigh these against one's goals, then they are better able to make self-ruled decisions. In a global sense, the agent could acquire more knowledge, understand it more deeply and be better able to recall it and use it when it is needed to support decision-making.

There are two questions to answer when assessing how an enhancement of one of these cognitive skills will affect autonomy. The first question asks how the enhancement will impact one's capacity for autonomous self-rule, or how it will affect the decision-making toolkit one possesses. As I have noted in the paragraph above, cognitive enhancement can increase autonomy by improving the cognitive tools one uses to make decisions. The second question asks how the enhancement will impact the likelihood or extent to which the agent will exercise that capacity or use that toolkit.

To be clear, this question differs slightly from the question being answered in the following paragraph that asks how an agent's practical autonomy, their ability to *act* on their deliberation, is affected. Rather, the current question asks how one's motivation/ability to use reason when deciding is affected by cognitive enhancement. Cognitive enhancement not only improves an agent's decision-making toolkit, it also increases the likelihood or extent to which they use it. If deferring to deliberation produces more favourable results after the enhancement than before (because the agent is better at it), they may view the use of careful deliberation more positively in future decisions and be more motivated to decide based on their reasoning. In this way cognitive enhancement's impact on decisional autonomy is compounding; the more one's capacity for reason is enhanced, the better they are able to identify how rationally deliberating upon decisions is favourable. The more positively they view using rational deliberation, the more they are inclined employ it.

As for practical autonomy, cognitive enhancement may increase one's ability to act on their made decisions. Cognitive enhancement may have a positive impact on an agent's positive freedom for two reasons. First, cognitively enhanced agents will have a greater ability to do the problem solving required to turn their decision into a desired result. For example, my positive freedom to play for a rugby club could be undermined by my lack of knowledge about what rugby consists of, but with greater powers of learning and information retention/recall, I am in a better position to take the steps required to action my decision to play rugby. Second, cognitive enhancement may have a small positive impact on one's negative practical autonomy. This is because, if the strength of one's reasoning is improved, then the weight it holds in the mind of the agent may be significant enough to overcome or even eliminate an internal influence, such as urges or emotional influence, that would otherwise be a threat to their negative freedom. For example, let's say that I had decided to lend a sum of money to my sister to fix her car, and then upon meeting my sister I am struck by sudden feelings of greed and distrust after recalling a news story of a person scamming their family members. My enhanced reasoning might verify that my sister's car is indeed broken

and that my sister is poor, recognise that my sister has paid back debts in the past, and importantly, understand that this is relevant information for my decision. If this is the difference between my overcoming those irrational influences – the feelings of greed and distrust – that harm my negative freedom, then the cognitive enhancement increases my negative freedom.

3.3.2. Emotional bioenhancement

Emotions themselves are not required for or directly involved in autonomous decision-making. They do not add relevant information to scenario that requires a decision. The role that emotions play in affecting the autonomy of a decision falls on the degree to which it helps or hinders one's cognitive deliberation. The nature of an emotional enhancement's impact on autonomy depends on how and what emotions are altered and how emotions are related to cognition and actions. Some emotions can undermine a person's ability to execute decisional deliberation as they distract them from what is important in deliberation, i.e. reasons. For instance, incidental emotions such as anger and sadness, excitement and joy have been shown to impact on people cognitive abilities and decision-making around eating, impulse control, setting prices for products, and helping and trusting others (Dunn and Schweitzer, 2005; Grunberg and Straub, 1992; Lerner, et al., 2004; Manucia, et al., 1984; Tice, et al., 2001; Yiend, 2010). Feeling depressed undermines a person's ability to judge their situation and consider their options properly (Barlow et al., 245; Leahy et al., 377). On the other hand, experiencing some emotions in certain situations may improve decision-making. For example, feeling apprehensive and skeptical may enhance decision-making in situations where careful consideration is required, but it may harm decision-making if the agent is in a situation where a quick decision is needed. If an agent is typically prone to fits of anger, and these fits undermine rational deliberation, then reducing the severity of that emotion may increase their decisional autonomy. To reiterate, emotions do not provide a person with relevant information or reasons to act, but they can affect a person's ability to assess any relevant reasons and deliberate upon them.

Similarly, the impact that an emotional enhancement exerts on one's practical autonomy depends on the details of the case. Some emotions might undermine one's ability to act on made decisions, such as the feelings of greed or skepticism when I lend my sister money. Feelings of love and joy when I see my sister may make me more inclined to act on my made decision. While in some situations feeling angry or anxious may undermine a person's ability to select and appropriately act on a decision, in others that anger may be vitally important (Litvak, et al., 2010). When a person is threatened with violence and must defend themselves, it is important that they feel the anger or anxiety necessary to intimidate or escape the attacker. Aristotle acknowledges that excessive anger is not virtuous, but he also notes that anger can be directed by reason, and he is critical of those who do not feel sufficient anger, anger of the right nature, anger toward the right things, or for the right duration (Aristotle, 121). Enhancements that improve the manner in which we feel anger, or toward what we feel anger, could improve our ability to act on the reasons that drive that emotion and our physiological response to that anger.

3.3.3. Behavioural bioenhancement

Behavioural enhancement does not impact decisional autonomy. This is because the person is free to carry out the decision-making process as usual with no influence from the enhancement technology, triggering the technology only if they attempt to act on a made decision that is not permitted by the technology. Therefore, behavioural enhancement would have a neutral impact on decisional autonomy.

On the other hand, a behavioural enhancement that intervenes and prevents action would have a significant impact on an agent's practical autonomy. Such a technology could exert a major threat to one's negative freedom because it could be a barrier to the person's actioning of a made decision. If a person is capable of making good decisions but struggles to act on them, then a behavioural enhancement that forces them to act on their decisions could promote their practical autonomy.

But such an absolute intervention is not necessary for a behavioural enhancement. For instance, instead of intervening and entirely preventing or forcing an action, the bioenhancement could prevent the agent from doing the undesirable action once or twice before allowing them to continue, or it could intervene and trigger a delay period, after which the agent is allowed to perform the undesirable behaviour. This would be the bioenhancement equivalent of person opting to 'self-ban' from a casino. Under the New Zealand Gambling Act 2003, casinos are required to ban people for a period of up to two years from entering the casino, if the person self-identifies as a problem gambler and asks that they be banned (Gambling Act 2003). The person's ability to act on a decision to gamble is therefore hindered, but not eliminated; they can still gamble online or at other physical gambling locations. A bioenhancement intervention of this kind might have a more negligible effect on autonomy and qualify as being an autonomy-friendly 'nudge'. I will discuss nudges in the Bionudge section.

3.4. Assessing the impact of different targets of bioenhancement on authenticity

Another way in which autonomy could be impacted by the enhancement of cognition, emotions and behaviour is the enhancement's influence on one's authenticity. There are two ways in which an enhancement can undermine an agent's authenticity: (1) The traits and character that make up the agent's identity are altered so suddenly and dramatically that their identity is destroyed and they cease to be the person they were, and (2) the agent is still considered the same person as they were before the enhancement, but they feel alienated from the changes to their thoughts, emotions and behaviour.

The makeup of one's character changes frequently throughout their life, and we do not consider this to be harmful to their identity. These changes are sometimes intentional and sometimes unintended, but typically we accept that a person's identity can change without thinking that their identity has been undermined. For a bioenhancement's change to one's character to be considered harmful to their identity, it must be a dramatic change that does not cohere with the

person's existing web of mental properties. As I explain below, some kinds of bioenhancement are more likely than others to change a person's character in this way.

3.4.1. Cognitive bioenhancement

It is challenging to see how enhanced cognition would create any significant sense of alienation from oneself. Perhaps this could occur if the enhancement is profound and, like Charlie's cognitive enhancement in the novella *Flowers for Algernon*, the agent is alienated by and unhappy with their new powers of reason and their subsequent outlook and beliefs (Keyes). However, many of us may typically experience a more subtle sense of alienation when, for instance, we shock ourselves with our incidental ability to recall a distant memory or a piece of obscure information. In these cases we feel a sense of alienation from our cognitive feat but because we endorse the results, and perhaps because the sense of alienation is not profound, we do not feel that our authenticity is compromised in these moments. Provided that one's current reasoning and capacities for learning and recalling information are not central to the person's sense of self, and they endorse the results of the enhancement, cognitive enhancement may only pose minor negative effects to one's authenticity. On the other hand, it is also possible that an agent believes they have cognitive capacities of a quality better than they actually do; the enhancement of these qualities could increase the degree to which they identify with their thoughts and behaviours because their actual capacities match the capacities they believe they possess.

However, the enhancement of cognitive abilities may compromise a person's identity regardless of how they feel about their altered identity. If one's cognitive abilities make up a significant portion of their objective identity (if such a thing exists), and an enhancement sufficiently improves a person's ability to learn, process, store and recall information, and reason about topics and decisions, then it is arguable whether the enhanced agent is really the same person they were before the enhancement at all.

3.4.2. Emotional bioenhancement

Emotional enhancement is more likely to undermine one's authenticity than cognitive enhancement. This is because our experience of our emotions, the emotions we feel toward certain activities, concepts, ourselves and others arguably make up a larger proportion of our identity, than the tools we use to process information. If an enhancement altered the way a person engaged with the people around them, the way they felt about the activities in which they typically engage, and the feelings that influence their values and stances on various normative issues, then it is reasonable to expect a person to feel a sense of alienation. In cases of extreme alteration agents may be concerned that they have completely destroyed their identity. Jebari raises the concern that reducing the experience of undesirable emotions like disgust and aggression may result in unexpected impacts on personalities and values such as the degradation of individuals' values regarding equality and justice (Jebari, 255). Jebari argues that if one cannot feel sufficient disgust and aggression after an enhancement, then they may lose the drive to punish criminals of serious crimes, or scold peers who break social codes (Jebari, 255). If one's normative values

and judgements on issues such as these are altered by an enhancement, we might expect an agent to feel alienated from their new outlook.

It is possible, however, that cognitive enhancement could also result in this kind of personality change in an indirect way if the improved cognitive capacities result in an agent changing their understanding and beliefs regarding issues of value and normativity. For instance, if one becomes better able to learn and understand the histories of colonisation and indigenous peoples, and their nation's property laws, they may take on a very different set of feelings toward those issues, and subsequently to their values relating to race and equality.

3.4.3. Behavioural bioenhancement

The form of enhancement most likely to result in an agent feeling a sense of alienation from themselves is behavioural enhancement. This is because the agent will be influenced to behave in certain ways that have not compelled themselves to do. In some cases they may have carried out some form of deliberation, settled on a course of action that they endorse and then find themselves unable to act on the decision on which they have settled and would otherwise be able to complete.

However, of the forms of enhancement discussed here, behavioural enhancement may pose the smallest threat to a person's identity because their cognitive abilities, values, judgements and feelings remain entirely unaffected. Behaviour is not a part of one's personal identity; we do not think a person's identity is destroyed, that a person has ceased to be, when they are physically paralyzed in an accident. But when a person suffers a brain injury and their cognitive capacities, memories or emotions are severely impacted, we are more inclined to say they are a different person. While the agent may feel alienated from their behaviours, their personal identity will survive the enhancement entirely.

3.5. Conclusion

Enhancements that act on cognition, emotions and behaviour have different impacts on one's autonomy. Cognitive enhancement poses the smallest threat to autonomy across the board. It will likely increase one's decisional autonomy and have marginal or no effect on practical autonomy and authenticity. Emotional enhancement's effect on autonomy is difficult to anticipate because the nature of the effect depends on what exactly is being enhanced and the situations in which decisions and actions are being made. Emotional states can have a significant influence on one's ability to deliberate on decisions and on their ability to act on them. Of the three targets of enhancement, emotional enhancement is the most likely to compromise a person's authenticity by undermining their identity and triggering a sense of alienation. The only form of emotional enhancement whose effect on autonomy we could have some confidence about is an enhancement that gives an agent greater control over their emotions. Enhancements that target behaviour pose contrasting impacts to their decisional and practical autonomy respectively; they pose no threat to decision-making whatsoever but they pose the greatest threat to practical autonomy. Likewise with authenticity, a person's identity may suffer only minor impacts from

behavioural enhancement, whereas they will experience profound alienation from their reduced abilities.

Analysing the impact of bioenhancements that target specific human capacities is important because the impact of enhancing different capacities varies significantly. This analysis sets up the following assessment of bionudges and their potentially more subtle influence on autonomy.

Chapter 4. Using nudge theory to mitigate negative impacts of bioenhancement on autonomy

4.1. Introduction

An emerging theory popular among policy makers is Thaler and Sunstein's 'Nudge' (Thaler and Sunstein). Nudging is a method of intentionally designing a person's environment to make certain decisions or behaviours more likely, using scientific understanding of human psychology. Examples of nudging include placing supermarket items at customer eye-level to encourage their purchase, and electronic traffic signs that record car speeds and broadcast them back to the driver to reduce excessive speed. Nudges do not always push nudgees toward a certain option, some encourage self-reflection, like cooling-off periods before a gun purchaser receives their gun or offering gamblers the option of voluntarily banning themselves from the casino (Hausman and Welch, 123). The essence of nudging is encouraging agents to make a certain decision or helping to facilitate deliberation by structuring the decision environment based on a particular understanding of human psychology and behavioural economics (Hausman and Welch, 123; Wilkinson, 343).

A key aspect of nudged choice architecture is that it should never remove or forbid options from the agent, and it should be almost effortless to resist the nudge and choose otherwise (Thaler and Sunstein, 6). This theory is popular among politicians of different ilk because it allows policy makers to shape the public's decision-making in ways that promote wellbeing and economic prosperity, supposedly without harming autonomy or conflicting with libertarian ideals. Nudge theory relies on a pragmatic understanding of human psychology and behaviour. In particular, Thaler and Sunstein focus on the shortcomings of human psychology – the cognitive biases, irrational tendencies and instinctive behaviours – that lead us to make bad decisions, but whose harmful effects could be counteracted by deliberately designing people's choice architecture (Thaler and Sunstein, 19). Given this focus, it is worth examining how we could nudge people via bioenhancement, without undermining their autonomy.

My discussion of nudging examines to what degree nudges increase or decrease autonomy. Many discussions around nudge theory explore the effect that nudging has on autonomy and whether governments can therefore be *justified* in creating policies that nudge the public. They examine the potential benefits of the nudge and weigh them against the potential losses of autonomy and decide whether there is a worthwhile trade-off. I examine to what degree nudges increase or decrease autonomy, but I defer discussion of whether that effect is justified. This is because democratic governments are already justified in coercive restrictions of autonomy, such as prohibitions of murder, theft, and other harmful crimes, in order to protect the welfare of their citizens. However, Hausman and Welch suggest that nudges may be more insidious than existing legal structures because they can amount to covert mind-control rather than transparent coercion (Hausman and Welch, 130-2).

In this section I examine how nudging impacts on autonomy. I also explore how nudge theory applies to my analysis on cognitive, emotional and behavioural bioenhancement, and the

relationship this bioenhancement would have with autonomy. I will call nudges carried out via bioenhancement technologies 'Bionudges' as I apply the insights from my analysis of nudge theory to autonomy and bioenhancement.

4.2. Nudging is not necessarily harmless to autonomy

In principle, nudging is neither harmful nor harmless to autonomy; the effect that nudging has on autonomy varies by case. Thaler and Sunstein claim that nudges are by-definition not harmful to autonomy because they should not make any options costly for the agent to select, either financially or through effort. But Thaler and Sunstein focus on the effect nudging has on liberty and freedom, which is only part of what is required to exercise autonomy (Engelen and Nys, 140-1). Nudged choice architecture might not remove options from a person, but it could undermine their ability to deliberate. On the contrary, some scholars argue that nudging, by-definition, *is* harmful to autonomy because it involves intentionally exploiting humans' irrational schema to shape their decisions independent of the agent's reasoning (Wilkinson, 343). But others include in their definition of nudging alterations of choice architecture that give reasons to the agent or help to engage the agent's reflective thought processes that we comfortably associate with autonomous decision-making (Wilkinson, 348; Mills, 32; Levy, 12). Further, Wilkinson argues that nudging does not always rely on the defects or shortcomings of human cognition, and defines nudging as designing choice architecture based on an understanding of human psychology and behavioural economics (Wilkinson, 343). Further, as I discuss in 4.5. reason giving and non-reason giving nudges, on coherentist accounts of autonomy, some nudges that operate via arational schema do not undermine autonomy because they cohere with a person's desires.

Nudges might be intended by choice architects to have no effect on autonomy, but they might misjudge the strength of their nudge or the suggestibility of their nudgee. If we can be more or less autonomous, and the degree to which different choice architecture influences an agent's decision-making can vary, then how much a nudge undermines or promotes an agent's autonomy will differ from nudge to nudge, and even nudgee to nudgee. A nudge in theory, might be a 'shove' in practise. Given this concern, my discussion does not centre around simply whether or not nudging harms autonomy, but rather *how* it affects autonomy, and what steps we can take to make that effect positive.

4.3. Manipulative and transparent nudges

Some scholars argue that nudging involves *manipulating* agents into making certain decisions (Hausman and Welch, 123-136; Mills, 495-509; White, 2013). Wilkinson examines what conditions make a nudge non-manipulative. He summarises manipulation as 'intentionally and successfully influencing someone using methods that pervert choice' (Wilkinson, 347). Wilkinson outlines the following as conditions that preclude a nudge being manipulative:

1. If and when nudging does not use a perverting method, it is not manipulation.
2. If and when nudging is not a form of intentional influence, it is not manipulation.
3. If and when nudging does not succeed in altering behaviour, it is not manipulation.

4. If the target consents in the appropriate way to nudging that would otherwise be manipulation, the nudging would not infringe on the target's autonomy and would not therefore be manipulative.

This set of conditions provide a good reference for evaluating whether a nudge is manipulative. However, it is important to note that a nudge can undermine autonomy without constituting manipulation. For instance, in regard to condition 1, a choice architect might design the options to be randomised and presented in full detail to avoid manipulating the agent to select any particular option. But as we have discussed, an oversaturation of choices and inclusion of irrelevant choices can undermine a person's autonomy by making it harder for them to assess the options and to commit to a final decision. In regard to condition 3, a nudge may be unsuccessful in making a person select a specific option, but if the nudge uses a method that undermines decision-making it could still undermine autonomy because it made it harder for the agent to deliberate and reflect. Further, a nudge might rationally persuade, clarify options or engage a person's autonomous reflective thought processes, and so its method might actually induce or support autonomous decision-making, despite doing so with the help of irrational schema.

In regard to condition 2, Wilkinson notes that if the choice architect does not intend to make a person choose a particular option, then the nudge is not manipulative. Wilkinson argues that the element of nudge theory which requires no options be forbidden and that avoiding the nudge must be effortless, is essential for the prevention of manipulative nudges (Wilkinson, 351-2). This is because if choice architects provide an adequate opt-out from the nudge, then they cannot be said to be intentionally perverting the agent's ability to choose for themselves. For instance, an expensive brand of beer might be perched near the main flow of foot traffic in the supermarket, but if the agent can see cheaper beer a few metres away, then it is difficult to say they are being manipulated to buy the expensive beer which is close at hand. The form of influence would not take the manipulative form of 'A intends B to choose X', but rather the non-manipulative form of 'A intends B to choose X, unless they do not want to choose X' (Wilkinson, 351-2). There is intention on the part of the choice architect, but not the kind that aims to subvert the agent's autonomy. However, a non-manipulative nudge can still undermine autonomy; a choice architect might not intend to shove (rather than gently nudge) agents toward an option, but their nudge might end up being stronger than they intend. If the cheaper beer was placed the length of a rugby field away from the flow of foot traffic, then it might be appropriate to say that while the choice architect does not intend the customer to choose the expensive beer, the customer's practical autonomy is still undermined because their ability to act on a decision to purchase cheaper beer is reduced.

To prevent manipulation and generally protect the public's autonomy, Thaler and Sunstein argue that governments (or whoever is nudging) should be prepared to publicly and transparently defend the method and the purpose of the nudge (Thaler and Sunstein, 245). Thaler and Sunstein's explanation of this is vague. They are unclear whether governments should actually disclose these details, whether they should simply be willing to do so, or even whether they must be successful in defending their nudge to the public. Hausman and Welch argue that because nudge

theory can be abused by governments, we should take Thaler and Sunstein's recommendation to its greatest extent and require that nudged agents are made aware that they are being nudged. However, others argue that many nudges are already transparent to agents, and that for defaults and product placement, the agent is already aware that these are nudges for their decision-making (Bovens, 217; Hansen and Jespersen, 20). If the nudge is transparent in its influence, then agents are free to recognise this, avoid the nudge and, if they believe the nudge is a shove, they can broadcast their judgement of that nudge and attempt to have it removed (Bovens, 217).

Bovens argues that while we might better safeguard agents' autonomy by proactively telling agents they are being nudged, this would make the nudges ineffective (Bovens, 217). He argues that nudges 'work best in the dark' (Bovens, 217). He claims that if someone is aware of how their decisions are being influenced by choice architecture then they would be less inclined to select the option toward which they are being nudged. But Bovens gives no reasons or evidence for this claim. Further, one should also consider the possibility that being told how one is being nudged might snap a person of their unreflective thought processing and engage more reflectively and deliberately with their decision. If the option toward which they are being nudged is a good option for that person, and a transparent nudge prompts reflective thought, the agent might still be inclined to select that option. This kind of nudge would help, not hinder, autonomous decision-making. But if, as some people argue, humans have a limited cognitive resources and reflective thinking can be understood as a scarce resource, then widespread transparent nudges attempting to engage people in reflective thought may become ineffective (Engelen and Nys, 142). Nudges are supposed to make human decision-making easier, not harder, after all (Thaler and Sunstein).

Some preliminary studies researching the efficacy of proactively transparent nudges have found that transparency does not reduce the likelihood that an agent selects the nudged option (Bruns et al., 2018; Kroese et al., 2015; Loewenstein et al., 2015; Steffel et al., 2016). In Bruns et al.'s study, healthy food is placed at a supermarket checkout (where customer blood-sugar levels are lowest and impulsive food purchases are most likely) along with a sign that reads 'we help you make healthier choices' (Bruns, et al., 2018). Despite disclosing the purpose of the product placement, transparently nudged customers were no less likely to select the healthy food than the control group. But this study does not explain how blood-sugar levels influence food purchasing habits; agents may be inclined to buy whatever food is at the checkout, regardless of nudge transparency, given that a person's blood-sugar is lowest toward the end of a shop and a person with low blood sugar is more inclined to buy food (Goodwin, 271).

A 2020 study examines how proactively transparent nudges affect agents' self-perceived autonomy (Wachner, et al., 2020). The study recognises that nudges can be more or less transparent, for instance agents can be informed about the method of the nudge, the purpose of the nudge, who designed the nudge, and the phrasing of the transparency message can vary (Wachner, et al., 2). Because of this, it is difficult to be certain that transparency does not alter the effectiveness of nudges until more studies corroborate that finding. In Wachner, et al.'s study, participants are asked whether they want to complete a short survey or the default long survey (Wachner, et al., 5). Participants were told that they would not be paid any more than the standard 1 British pound for completing the longer version but that they will helping to improve future

questionnaires (Wachner, et al., 5). The control group were only nudged via the default option I have described; the test group was informed that the purpose of the default was to encourage more people to fill out the longer questionnaire to help; and a second test group was informed of the purpose of the nudge *and* that typically people are unaware that by setting a default option, they are much less likely to opt-out and instead complete the shorter questionnaire (Wachner, et al., 5). The study found that group 2 and group 3 participants were not less likely to fill out the nudged longer questionnaire, compared to the control group (Wachner, et al., 9). Further, participants in all three groups reported high levels of self-perceived autonomy, with no notable difference between each group (Wachner, et al., 10). It is important to note that self-perceived autonomy is not necessarily an accurate view of actual autonomy because a person who has been unknowingly manipulated will self-rate their autonomy higher than they perhaps ought to. Although, how an agent feels about their decisions should give us a general indication of the state of their autonomy.

As I have noted, it is possible that proactively transparent nudges prompt agents to use their more reflective thought processes, and if the nudged option already coheres with the agent's desires then prompting the reflective thinking will not change the outcome. Purchasing healthy food options and helping researchers carry out a survey are things that participants might typically endorse anyway. If researchers want to get a better idea of whether transparent nudges are effective and influence self-perceived autonomy, then studies should attempt to transparently nudge people toward selecting options which are not in their interests. In other words, researchers need to ensure that it is really the nudge that maintains efficacy when it is transparent, and not the transparency message that prompts reflective thought. If the transparency message is indeed prompting reflective thought, we might expect agents to select options besides the one they are being nudged toward, if that nudged option is not actually in their interests. If the point of transparent nudges is to prevent malevolent manipulative nudges, we should examine exactly those kinds of nudges.

It is easy to see how a bionudge could influence a person's decisions without their knowledge. If a person's cognition, emotions and behaviour are slightly altered, then a bionudge's influence may not be great enough to detect, particularly given that they operate in the agent's body and mind. Bioenhancements that have a drastic influence on decision-making and behaviour, such as sudden spikes in reasoning and mathematical ability, or being uncharacteristically calm in a tense situation, will be relatively transparent to agents. But transparent bioenhancements may have too drastic an influence to be considered bionudges at all.

However, if a person gives informed consent to a bionudge, then they will have some understanding and awareness of how their decision-making is likely to be influenced, however subtle it may be. Therefore, decisions influenced by a voluntarily endorsed bionudge, however subtle the impact, will have some degree of transparency for the bionudgee.

Further, some bionudges can be made deliberately transparent, particularly if they are administered via brain chips. This is because, theoretically, brain chips can send an explicit alert in the form of a message or notification that informs an agent that the bioenhancement is taking

effect. It is also possible for the message sent by a brain chip to constitute the bionudge itself. This could take the form of prompts that encourage a person to take more time to consider their options, or to be mindful that the bionudgee's blood-sugar levels are low and how this may impact their emotions, decisions and social behaviour. These information-giving and reason-giving bionudges would not lose their efficacy in influencing decision-making by being made transparent because their efficacy *depends* on the agent being aware of the bionudge.

While there is some hope for nudges and bionudges maintaining their efficacy despite being proactively transparent, broader research is needed to confirm this. Certainly, as I have noted, some nudges and bionudges, such as those that give reasons or clarify options, will not lose efficacy through transparency. But with such broad possibilities and circumstances for different nudges and different styles of transparency messages, it is yet unclear whether nudge efficacy and self-perceived autonomy are truly unaffected by transparency.

4.4. Heteronomous influence and cosmonomous choice environments

Scholars worry about nudging being manipulative because manipulation is viewed as uniquely harmful to autonomy. Just as one might be concerned about the effect of coercion on autonomy because it involves heteronomy, they may be concerned that nudging is a manipulative form of heteronomy. But we might reasonably ask whether this heteronomous nudging is actually any worse for our autonomy than cosmonomy, that is, the way our existing choice environment and nature limit our autonomy (Hausman and Welch, 133). Indeed, one of the key arguments defending nudging on the grounds of autonomy, holds that humans' decisions are heavily influenced by their choice environment regardless of whether it has been designed (Hausman and Welch, 133). While it may seem that another person influencing our decisions would undermine our ability to self-rule more than unintentional underminings of autonomy, we need only look at specific examples to see why this is not so. For instance, a supermarket designer who opposes nudge theory might randomise the placement of the shop items to avoid influencing people, but shoppers will still be inclined to buy whatever products end up at eye level, or at the checkouts. Both heteronomy and cosmonomy are external influences on autonomous decision-making, and both influences can go unnoticed. Given that *some* choice environment must exist, the degree to which our decisions are affected does not depend so much on what or who is influencing, but rather the methods by which the choice environment influences. If a nudge undermines autonomy, then we might say that the nudge is morally dubious or unjustified, but it does not necessarily undermine autonomy any more than the original choice environment.

Rather than considering whether heteronomy is more harmful to autonomy than cosmonomy in principle, we may ask whether it is more harmful in practise. To some extent, we already expect the world to influence our decisions and for our psychology to have complex but finite capacities of deliberation, and we are therefore better able to account for those influences on our decision-making. But other humans sometimes aim to deceive us, and it is not always clear whether another person is a friend or foe. In light of this, perhaps we should be more concerned about heteronomy than cosmonomy. But like other humans, our choice environment sometimes covertly undermines our autonomy: some people are not aware that humans tend to weigh future costs

much less than present costs, or that unconscious factors like low blood-sugar decrease proclivity to social cohesion (Goodwin, 271). Further, choice architects might design choices with the aim to promote agent welfare or autonomy, by placing healthy food at the checkouts or wrapping unhealthy products with warning labels. Thaler and Sunstein describe a beneficent school cafeteria designer who wishes to increase student welfare by choosing to put healthy food first in the cafeteria line because hungry students are likely to load their plate up most with whatever is first (Thaler and Sunstein, 11). To what degree a bionudge or an undesigned choice environment helps or harms autonomy is going to depend on the particulars of the choice, not on whether the influence is intentional.

While a choice environment is going to be present for all decisions, some nudges will add *extra* influence into the environment or target cognitive weaknesses (Hausman and Welch, 133). In cases where nudges add extra influence into the existing choice environment, people should be more concerned about the nudge's impact on a person's autonomy. For instance, the school cafeteria food must be arranged in some way or another, but adding flashing lights and music near the fruit and vegetables involves adding an influence which was not otherwise present (Hausman and Welch, 133). Nudges that add extra influence to the choice architecture will undermine autonomy more than those that simply replace influences. On the other hand, some nudges that add extra influence may potentially increase autonomy by offsetting the influence of the existing choice environment. For instance, high-rated 'health star rating' stickers on food products influence consumers to buy that product by adding extra influence to the choice architecture, but it offsets the cosmonomous influence of humans' evolutionary biology and psychology to select food items high in sugar (Wiss et al., 2018). If we can better offset the external influences to select different options, then we can make it easier for agents to make their decisions based on genuine desires about those options.

Bionudges can also operate a way that offsets the existing choice architecture that undermines autonomy. If a person knows that they are prone to behaving unfavourably during fits of rage or fear, a bionudge could respond to high levels of relevant hormones and reduce them or trigger a behavioural delay on aggressive or cowardly behaviour. This would offset the unwanted cosmonomous influences of brain chemistry on their decision-making and allow the person to make decisions that they endorse. The 'as-needed' use of anti-anxiety medication is a bionudge that people already use in this way. When a person knows they are about to enter a situation that typically makes them unduly anxious, such as social events or public speaking, they may take a Benzodiazepine pill which counteracts the unwanted cosmonomous influences to allow them to feel better and behave in favourable ways (Westra and Stewart, 2002). This is considered a bionudge rather than a 'bio-shove' because a person on Benzodiazepines has a reduced physiological reaction to anxiety-inducing situations but retains the cognitive function necessary to identify and respond to risk (Westra and Stewart, 2002). Despite adding extra influence to choice architecture, bionudges that equally offset autonomy-undermining influences pose no threat to autonomy.

4.5. Reason-giving nudges are less harmful to autonomy than non-reason-giving nudges

Some philosophers argue that nudges are inconsistent with autonomy because many of them do not give reasons to agents and do not rationally persuade them (Blumenthal-Barby and Naik, 45). If nudges toward selecting certain options do not involve giving reasons to select that option, then Blumenthal-Barby and Naik argue that they do not increase the agent's decisional autonomy. This is because the nudged agent is no better able to rationally deliberate about whether an option coheres with their desires, goals and values. Moreover, some nudges not only omit reasons, they also covertly influence decision-making toward an option by intentionally bypassing the agent's rational deliberation and targeting the agent's arational and sometimes irrational decisional heuristics (Blumenthal-Barby and Naik, 45). For instance, playing quiet soothing music on the airplane as passengers board bypasses humans' rational assessment of risk and makes them feel more relaxed by targeting unconscious schema (Hansen and Jespersen, 21). Further, the framing of risk can affect people's decision-making. Framing the risk of a surgery as having a 5% mortality rate makes patients more likely to refuse the surgery than if the surgery is framed as having a 95% survival rate, even though the information is logically identical (Ploug and Holm, 29). By deciding how we frame risk to agents, we can target irrational schema to influence decision-making. This is a concern because if the nudge does not support rational deliberation, it does not promote decisional autonomy.

Others argue that nudges that target irrational schema or do not appear to give reasons may still be consistent with autonomy. Smoke alarm advertisements may exploit viewers' emotions by showing a family tragedy that could have been prevented by a smoke alarm. But while a purely fact-based advertisement might have given reasons more directly, an emotional advertisement does not undermine autonomy because the fear it induces makes the beliefs that we already hold about the risk of fires more salient in our minds (Wilkinson, 348).

Moreover, Levy argues that humans view framing and defaults as recommendations for decisions. If a doctor chooses to frame risk positively by saying the survival rate of the surgery is 95%, patients will understand this as an implicit recommendation to select the surgery (Levy, 8). He argues that in the absence of our own understanding, it is rational to defer to expert testimony, so selecting the default option or consenting to the surgery amounts to accepting a recommendation (Levy, 8). It is not obvious that all people actually view this framing as a recommendation, but there will certainly be some cases where the default option is viewed by decision makers as a recommendation from choice architects. I will revisit Levy's arguments in the next subsection, 4.6. *Influencing type 1 processing threatens autonomy more than influencing type 2 processing.*

But some nudges do not give reasons or rationally persuade, but also do not covertly exploit humans' irrational schema. For instance, diesel pumps at gas stations in New Zealand are labelled red, and 91 petrol pumps are labelled green, to help customers use the right fuel for their car. The colour of the pump does not give reasons to the petrol customer but plays on a learned unconscious rule that a particular-coloured pump is the one they want to use.

We might also say that nudges that play on a person's irrational schema/biases are *ceteris paribus* harmful to autonomy. This is because they not only fail to give reasons, but they also encourage agents to continue using the irrational schema that typically undermines their assessment of the situation and their favoured courses of action. However, nudges that operate via a person's irrational schema can still promote autonomy in three ways.

First, a nudge that targets irrational schema may not support one's ability to make a decision, but it could increase the likelihood of selecting options that cohere with a person's endorsed values or desires. For instance, people are biased toward the status quo. In many modern cars the traction control system is switched on automatically when the engine is started. There are only some situations where disabling traction control is safer than having it switched on, such as when driving in mud or snow (New Zealand Automobile Association, 2022). Car companies that make traction control a default when a driver starts their car use the cognitive bias toward defaults to increase driver safety. Presuming that drivers want to be safe, car companies target drivers' cognitive biases in a way that promotes their autonomy from a coherentist point of view.

Second, a nudge could target one irrational schema in order to offset the autonomy-harming impacts of another irrational schema. For instance, anchoring bias could be used to encourage people to save more money for retirement, to offset the impact of short-term bias which makes them under-save for their future, or wait until they are much older to begin saving. Consider a bank's website that provides customers guidance how much money to contribute weekly to their retirement fund. The bank may ask how much money the customer estimates they need to save weekly for retirement, then give the following examples of how much other people save, 'Steve estimates that he will need to save \$80 dollars a week. Hannah has decided she needs to save \$110 dollar a week.' This targeting of anchoring bias makes a person more likely to estimate they need to save an amount similar to Steve and Hannah, even though they are given no explicit reason to think that Steve and Hannah have chosen to save an appropriate amount of money. We can offset the short-term bias that influences people to under-save by targeting this irrational schema, encouraging them to save adequately for retirement – something that likely coheres with their desires.

Third, nudges can target irrational schema to encourage a better assessment of options and increase decisional autonomy. As I have noted, short-term bias makes us inclined to put off saving for retirement until later in life. The New Zealand Government targets the public's short-term bias by contributing 50 cents for every dollar a person invests in their Kiwisaver superannuation fund, up to a maximum of ~\$500 each year (New Zealand Inland Revenue Department, 2021). In light of this scheme, people must contribute in the short-term to get this year's free money. They are given a tangible incentive to act now, rather than later. This scheme not only targets biases to get endorsed outcomes. It also targets a bias to improve decisional autonomy, because the short-term real-dollars gain of free money triggers a person's interest in contributing now and that makes salient in their mind the amount of money they will have during retirement if they contribute at the rate required to receive the free ~\$500. The allure of being given free money now, to be used in retirement, draws people's attention to the real costs and benefits of saving for retirement.

Some nudges may target irrational schema and they may not explicate reasons, but this does not necessarily mean that they undermine decisional autonomy.

4.6. Influencing type 1 processing threatens autonomy more than influencing type 2 processing

Nudges can target type 1 or type 2 thought processing, and the impacts of the nudges on autonomy vary depending on what type of processing is being targeted. Type 1 processing involves unreflective, unconscious or reactive decision-making and behaviour (Hansen and Jespersen, 14). Because it is automatic, type 1 processing is rapid and does not deplete cognitive resources as much as type 2 processing (Levy, 4). Type 1 processing can be inflexible, in that even when agents are aware that a cue in their environment is false or inaccurate, the cue might still influence their behaviour (Levy, 4-5). For instance, while we may be aware that taller skinnier bar glasses hold much less liquid than shorter wider glasses, when we drink from a tall skinny glass, we still sense that we have consumed more beer than we actually have (Hansen and Jespersen, 22). Type 2 processing, on the other hand, involves reflection, judgement and deliberate thought and action in decision-making and behaviour (Hansen and Jespersen, 15). It is more flexible in reacting to nuance in situations and more reactive to reasons, so it is associated with intelligent, and thus autonomous, decision-making (Levy, 4-5). It is important to assess how autonomy is impacted by type 1 and type 2 nudges because it informs how bioenhancement can alter cognition, emotions and behaviour in a way that does not undermine self-rule.

Nudges can also target both type 1 and type 2 processing. For instance, establishing a default option is a type 1 nudge, because it does not necessarily prompt reflective thought and plays on humans' status quo bias. Though, if one views the default as a recommendation, then it might target both type 1 and type 2 processing. On the other hand, placing a sticker of a fly in the centre of a urinal is a type 2 nudge to discourage missing the urinal, because it triggers humans' automatic 'visual search processes' and gives agents something at which to aim (Hansen and Jespersen, 15).

Hansen and Jespersen note that nudges that target type 2 processing are considered less harmful to autonomy because even though they function via automatic influence (like triggering our visual search processes), they prompt a more deliberate decision and therefore give agents more of a chance to act contrary to the nudged option (Hansen and Jespersen, 21). Type 1 nudges, on the other hand, are *prima facie* more concerning for autonomy because they influence the selection of an option more directly, and because decision-making based on type 1 processing is less conscious. For example, people tend to consume less food in buffet restaurants when the restaurant uses smaller plates, but the size of the plate directly influences behaviour, rather than the agent's deliberate decisions about how much food they want to eat (Hansen and Jespersen, 15). Type 1 nudges can be understood as influencing behaviour, rather than decision-making, because the influence bypasses deliberation. Hansen and Jespersen stress that the distinction between type 1 and type 2 thought processing is not a strict differentiation, and that nudges can be more or less type 1 or type 2, with some being in the middle (Hansen and Jespersen, 20). But

this distinction gives us a means of quickly assessing how harmful a nudge might be and comparing its harm to autonomy with other nudges.

Levy argues that a lot of human decision-making is carried out via type 1 processing, but he argues that this does not mean it is unintelligent or arational, and therefore nudges that target type 1 processing are not necessarily harmful to autonomy (Levy, 6). If type 1 processing and type 1 nudges use reasons, then we can make a better case for type 1 nudges being consistent with personal autonomy, insofar as our autonomy depends on being able to act on reasons. Levy argues that type 1 intuitive processes are not arational processes, rather they are rational processes from humans' learned or evolutionary past that misfire in new contexts (Levy, 8). For example, candidates that are listed near the top of voting ballots gain an advantage over others below because humans' type 1 processes trigger a slight bias toward those first on the list (Levy, 3). Many ballots are organised alphabetically, so being at the top of the list is not a good reason to vote for a candidate (Levy, 8). But this does not mean that the process that favours those at the top is arational, he argues, rather it is a rational mechanism that misfires in this situation (Levy, 8). Levy argues that this process is rational because options which have been made more salient to agents, such as those at the top of lists (or defaults, or the framing of risk), are viewed by agents as recommendations (Levy, 8). Levy argues that if it is a rational process to be guided by testimony, then it must be a rational process to be guided by implicit recommendation in choice architecture (Levy, 8). Further, he argues, it is rational to defer to testimony when cognitive resources are low, or in other words, when one does not possess the information, energy or care required to deliberate over a decision (Levy, 9; Weber, 114). Other type 1 processes, Levy argues *are* reasoning processes and are therefore consistent with autonomy.

However, Levy's argument is disanalogous. When we defer decision-making because we do not possess sufficient cognitive resources and instead make choices based on expert recommendation, we do so consciously and deliberately. We are cognizant of our cognitive deficiency and based on that reason, we decide to seek or listen to advice. But when choice architecture influences our selection via type 1 processing, our 'referral to recommendation' is automatic and unconscious and so it does not satisfy the common standards of autonomous decision-making as Levy claims. Levy's claim that type 1 processes, such as defaults, resemble deference to testimony is therefore disanalogous.

Further, Levy's description of type 1 processes as being 'rational processes' insofar as they play some causal role in unintentionally producing favourable outcomes, indicates that they are rational in a way that does not apply to the kind of autonomy with which we are concerned. If the processes produce a decision that coheres with our values and desires, then those type 1 nudges are at least consistent with autonomy. But that consistency with autonomy is often not due to type 1 processes being rational or reason-sensitive in any respect, because as Levy admits, these processes misfire in modern contexts. Sometimes we just get lucky when our unconscious thought processes result in outcomes that cohere with our desires.

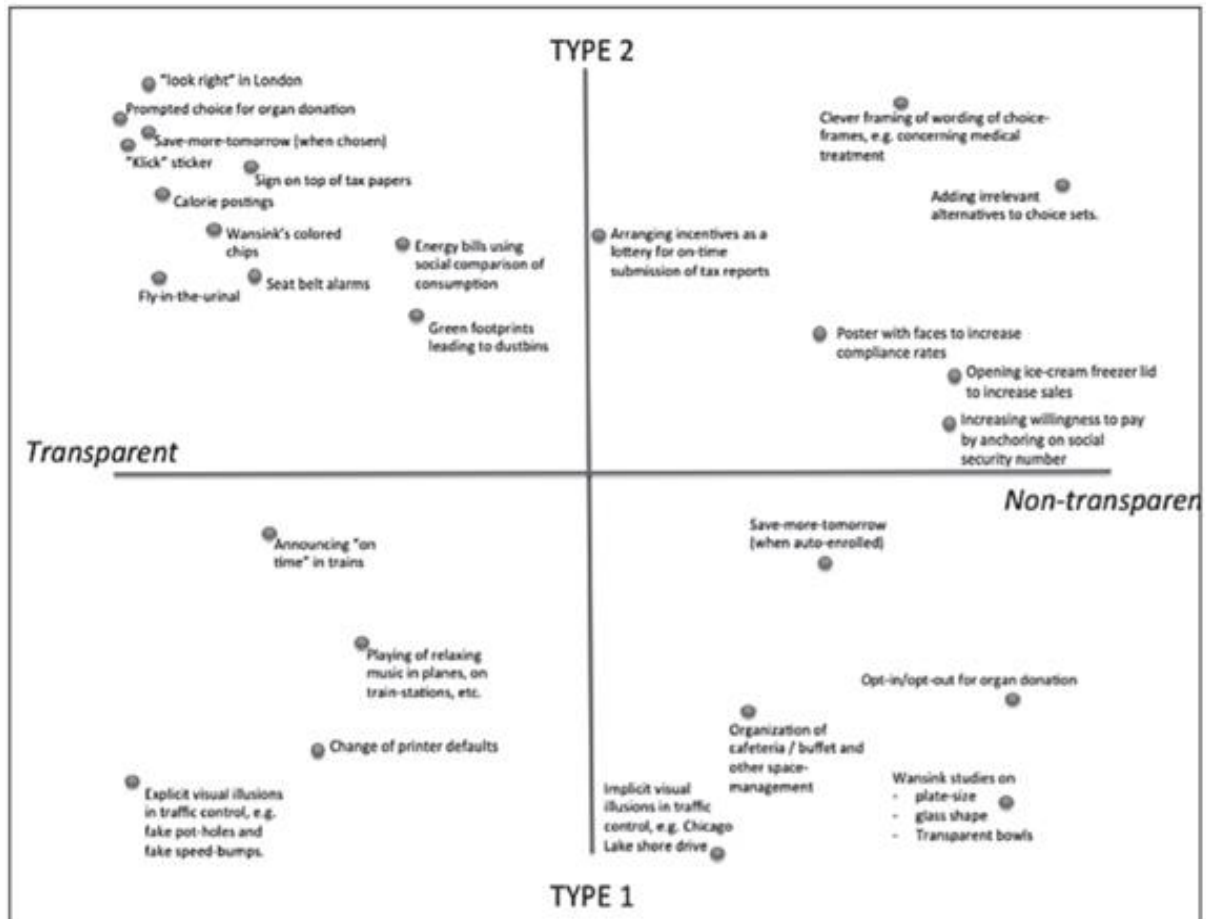
Similarly, Blumenthal-Barby and Naik argue that even though type 1 nudges might not use reasons to influence decision-making, this does not mean they are incompatible with autonomy

(Blumenthal-Barby and Naik, 45). We might deliberately and rationally defer to type 1 processes in certain situations (Blumenthal-Barby and Naik, 45). They argue that we should take a more realistic view of human psychology and autonomy, where most everyday decisions can be considered autonomous even though we do not fully inform ourselves or entirely decide on our own, because we might deliberately endorse this style of decision-making (Blumenthal-Barby and Naik, 45). For instance, I consider that choosing what bread to purchase is a trivial decision and I do not believe it requires rigorous deliberation. Accordingly, I endorse deferring the decision-making to my type 1 processing where I might be nudged toward one option or another. Decisions such as these are considered trivial because all of the options fall within our more psychologically realistic 'perimeters of autonomy', wherein all options cohere with our values and goals (Engelen and Nys, 153). Although a nudge might successfully change our behaviour, this does not mean that the nudge undermines autonomy because the option toward which we were nudged could also fall within our perimeters of decisions that would be made if the person was acting autonomously; it still coheres with our desires and values (Engelen and Nys, 154).

Further, over-reliance on active choosing via our type 2 thought processing could undermine our autonomy more than deferring to type 1 because it prevents us from focusing our cognitive resources on the decisions that matter to us (Engelen and Nys, 142). It might not only be autonomous to defer to type 1 processing, but it might also be rational because it prioritises cognitive resources. Though, in this sense type 1 nudges support our autonomy only indirectly, by freeing up cognitive resources for other decisions, but not by supporting our autonomy for the decision in which we are being nudged (Engelen and Nys, 143). However, if our meta-decision preferences (our preferences about *how* we make a decision) endorse deference to type 1 processing for this decision and even other similar decisions, the nudge does not appear to undermine autonomy (Wilkinson, 351). Therefore, bionudges that operate via type 1 processing can be harmless to autonomy, in decisions where an agent endorses the bionudge.

Hansen and Jespersen add another value to their assessment of nudges by examining not only whether a nudge functions via type 1 or type 2 processing, but also to what degree the nudge is transparent. They argue that nudges can be more or less transparent (Hansen and Jespersen, 18). Proactively disclosed nudges (such as the ones studied in Wachner, et al.'s paper) are extremely transparent. Nudges where the intention and method of the nudge are not disclosed but are still obvious to the agent are transparent. This includes the fly sticker in the urinal, changing printer defaults from one-sided to double-sided to reduce paper use, and unused seat belt alarms in cars (Hansen and Jespersen, 20). Non-transparent nudges are those where the agent would struggle to recognise the intention of the nudge or means by which the nudge works. These nudges include plate and glass sizes and shapes in buffets to reduce consumption, music styles in shops and supermarkets to influence purchasing habits, and the framing of risk or treatment options when giving medical advice (Hansen and Jespersen, 18). Both type 1 and type 2 nudges can be transparent or non-transparent, so Hansen and Jespersen create the axis below to visualise and assess the degree to which nudges undermine autonomy (Hansen and Jespersen, 20).

4.7. The nudge axis and bionudges



On this axis there are four types of nudges: transparent type 2 nudges, non-transparent type 2 nudges, transparent type 1 nudges, non-transparent type 1 nudges. It is important to note that nudges do not necessarily fit perfectly into these categories, they can be more or less transparent and affect both type 1 and type 2 processing. But the axis is an imperfect tool that is still useful for assessing and comparing nudges. The fly sticker in the urinal is a transparent type 2 nudge because it is obvious to the agent that it is intended to prevent messes and because it operates via engaging deliberate decision-making (Hansen and Jespersen, 21). An example of a transparent type 1 nudges is the Danish national train service announcing 'on time' when the train arrives on schedule so that people more easily remember the times the train was punctual and not just the times it was late (Hansen and Jespersen, 21). The purpose of the announcement may be obvious to passengers, but the nudge functions by using humans' availability bias to draw passengers' attention to the punctuality, making memories of punctual train rides more salient in their memories, instead of just the times the train was late.

Non-transparent type 2 nudges are those where agents are making a conscious and deliberate decision but in which the influence of the nudge on that decision is not obvious (Hansen and Jespersen, 22). Displaying images of eyes or faces can make people more inclined to behave pro-socially (Bateson et al., 1-9). An example of this used at the University of Canterbury where

bike racks display an image of human eyes on posters to discourage people from stealing bikes. Finally, non-transparent type 1 nudges influence non-deliberate decision-making in ways that are not obvious to agents and so they seem to pose the greatest threat to autonomy of all four types of nudges (Hansen and Jespersen, 22-3). This includes the use of smaller plate and glass sizes in restaurants, because it influences a non-conscious sense of the amount of food one has consumed, and it is not obvious to agents that the size of the plates has been intentionally selected to reduce consumption.

Hansen and Jespersen argue that transparent type 2 and transparent type 1 nudges are generally acceptable and do not undermine autonomy because their transparency offers agents the chance to opt-out or avoid the nudge (Hansen and Jespersen, 24-5). They argue that non-transparent type 1 nudges pose a greater threat to autonomy because they are covert in their influence, so they make it harder for agents to opt out (Hansen and Jespersen, 26). However, agents are sometimes able to use their type 2 processing to override the influence from the nudge that targets their type 1 processing, and more important decisions tend not to be influenced by type 1 nudges because agents are more likely to be engaging their reflective type 2 processing when making big decisions. Because non-transparent type 1 nudges do not affect overt choice making, Hansen and Jespersen are not concerned about the potential for non-transparent type 1 nudges to significantly undermine the decisional autonomy of agents, though they do define this kind of nudge as the 'manipulation of behaviour' and recommend that public policy makers ensure that the nudged option is in the public's best interests and that they consider informing the public of the nudge (Hansen and Jespersen, 26).

It is non-transparent type 2 nudges that Hansen and Jespersen deem most threatening to autonomy and which they deem a 'straightforward manipulation of choice' (Hansen and Jespersen, 27). They argue that this is because agents cannot revert to their more reflective type-2 processing to avoid the nudge, because it is their type 2 processing that is already being influenced. They argue that non-transparent type-2 nudges cannot be justified because of the threat they pose to autonomy (Hansen and Jespersen, 27). On this view, doctors intentionally framing the risk of failure in a surgery as 'there is a 95% chance the surgery is successful', as opposed to 'there is a 5% chance of failure' to encourage patients to agree to undergo surgery amounts to a straightforward manipulation of choice, even though risk looms larger than it should in our minds due to risk-aversion bias.

Contrary to Hansen and Jespersen's claim, people are able to resist the impact of a type 2 non-transparent nudge. We are able to analyse our own conscious thought processes, critique them, recontextualise them, and adopt new positions as a result. A doctor's framing of risk may make the risk loom larger in our minds than it should, but this does not preclude us from recognising that humans are prone to placing too much weight on the risks of a decision and that the framing of risk can influence us to make irrational decisions. When a product is advertised as being '20% cheaper' than an undisclosed previous price, we are able to use our type 2 processing to ask '20% cheaper than what price?' to work out whether the current price is actually a bargain, despite this advertising nudge operating on our type 2 processing. Further, while non-transparent type 2

nudges might influence overt choice the most out of the four types of nudges, they can still be used to offset the aspects of human psychology that undermine autonomy.

Hansen and Jespersen's transparency/processing-type axis is a helpful measure, but assessing how different nudges impact autonomy is not as simple as seeing where they land on the axis. This is because of my arguments above, but also because type 1 and type 2 non-transparent nudges might nudge us toward selecting options that we do not actually endorse, things that do not cohere with our web of desires and values. Even though the transparency of a nudge enables us to avoid the nudge's influence, and while type 2 nudges engage our more reflective thought processes, they might still push us toward doing things which do not accord with our values and desires. For example, power bills that target our social psychology to reduce coal-powered energy consumption by indicating whether we have used more or less energy than our neighbours might be a transparent type-2 nudge, but if this nudge discourages an environmentalist from charging (and using) their electric car and lawnmower, then it might actually undermine their autonomy because it conflicts with the agent's settled values and preferences.

In this subsection I apply Hansen and Jespersen's axial analysis of nudging to bionudges.

4.7.1. Transparent bionudges

Type 1 – Transparent bionudges

A type 1 transparent bionudge could be a behavioural bionudge that delays action or triggers an agent to begin carrying out a favoured action. It's a type 1 nudge because it would play on a person's automatic status quo bias, wherein they are likely to continue carrying out the action if they have already started it, and because the effect does not impact their conscious thought processes. This would be transparent because an agent would notice if they began doing something that they had not initially decided to do. Moreover, it is a nudge rather than a shove, because when the bionudge's influence subsides the agent may choose not to continue. This bionudge resembles Isla's habits causing her to unconsciously begin driving her car, as she typically does. A type 1 transparent bionudge could also alter a person's emotions, and informs the person that it is doing so. Provided that the effect of the bionudge is obvious to the agent and it plays on their automatic arational thought processes, it qualifies as a type 1 transparent bionudge.

Type 1 transparent bionudges pose few risks to autonomy. As Hansen and Jespersen note, their transparency enables agents to resist the nudge. Further, by influencing only type 1 thought processing, the type 1 bionudge fully ensures the agent is able to revert to their type 2 thought processing to overcome the effects of the bionudge. These elements provide a lot of protection to the agent's autonomy.

Further, type 1 transparent bionudges are going to pose even less risk to autonomy if they take the form of behavioural bionudges, rather than emotional nudges. This is because while behavioural bioenhancement poses some risk of alienating a person from their behaviour, it poses

very low risk to decisional autonomy because it can only impact it insofar as it prevents deliberation; behavioural bioenhancement can increase practical autonomy if it helps people act on their made decisions, and only briefly or marginally undermines it if it nudges them to act against made decisions; and it poses no risk to one's identity.

Emotional enhancement, on the other hand, poses risks to each of those elements of autonomy. This because one kind of emotion can motivate many different kinds of decisions, and in one circumstance experiencing a particular emotion may be appropriate and endorsed, and in a subtly different situation that same emotion may be entirely unwanted and unhelpful. For example, reduced anxiety and fear might be desired for someone whose fear of tsunamis prevents them from kayaking. But overly reducing anxiety and fear may impede one's risk assessment when making a decision to go kayaking in rough conditions, and it may reduce the likelihood that they act on a made decision to wear a life jacket. While current as-needed use of anti-anxiety medication appears to balance this influence effectively, this example demonstrates how bionudges that alter emotions and do not strike a particular balance of influence can have broad and uncertain impacts on decision-making and behaviour (Westra and Stewart, 2002).

Type 2 – transparent bionudges

Whether transparent or not, type 2 bionudges could only take the form of a cognitive enhancement (rather than emotional or behavioural), because by acting on a person's type 2 processing the bionudge's impacts on the agent's ability to analyse and deliberate on decisions. Emotional enhancements and behavioural enhancements do not produce their intended outcomes by influencing a person's conscious decision-making, except perhaps indirectly, insofar as certain emotional states lend themselves to better reasoning or judgement.

A type 2 *transparent* bionudge might take the form of a cognitive bionudge that enhances cognitive capacities such as reasoning ability or working memory and informs the agent that the bionudge is taking effect. It might also include cognitive bionudges that do not enhance capacities but take the form of alerts/prompts to the agent to consider options carefully, or to take more time to consider their decision.

In light of my assessment of the targets of bioenhancement and Hansen and Jespersen's framework for understanding the impact of autonomy on nudges, type-2 transparent bionudges pose the least risk to autonomy overall. This is because (1) their transparency means agents are aware they are being bionudged and this enables agents to resist the nudge, and (2) because they must be cognitive bionudges, and as I have noted in Chapter 3, the effect that cognitive enhancement has on different conceptions of autonomy is either neutral or positive.

4.7.2. Non-transparent bionudges

Type 1 – non-transparent bionudges

Type 1 non-transparent bionudges will only take the form of emotional or behavioural bionudges. Cognitive enhancements cannot be type 1 bionudges because they impact the agent's conscious type 2 thought processes. Emotional bionudges can be transparent or non-transparent because humans can be either aware or unaware of how their emotions are influencing their decision-making and behaviour. Therefore, bioenhancements that alter our emotions, particularly those with effects that are subtle enough to be considered bionudges, are not necessarily going to be transparent to agents.

While one might argue that behavioural bionudges must be transparent to an agent because they can conflict with an agent's made decision, there are some instances where they can be non-transparent. If a behavioural bionudge triggers a person to carry out an activity that they have already endorsed, then while the bionudge may not undermine the person's autonomy, the bionudge may not be transparent to the agent. Additionally, if an agent is 'running on autopilot' and not making deliberated decisions about what they are doing, then the effect of a behavioural bionudge will be imperceptible to the agent, provided that the nudged behaviour is not wildly outside of the norm for the agent. Therefore, type 1 non-transparent bionudges can be either emotional or behavioural bionudges.

Hansen and Jespersen argue that while the non-obviousness of type 1 non-transparent nudges poses a risk to a nudgee's autonomy, the fact that it focuses on influencing type 1 thought processing means that agent is able to use their type-2 processing to overcome that influence. However, because type 1 non-transparent bionudges are more likely to take the form of emotional, rather than behavioural, bionudges, and because emotional bionudges are more likely than other bionudges to have broad and unexpected effects on decision-making, they do pose a substantial risk to a person's autonomy.

Type 2 – non-transparent bionudges

As I have noted, type 2 bionudges can only take the form of a cognitive bionudge. An example of a type 2 non-transparent cognitive bionudge could be one that alters the way that we process information and evaluate options e.g. improved working memory, ability to absorb and retain information and reasoning.

Hansen and Jespersen argue that type 2 non-transparent nudges pose the greatest threat to autonomy. However, given that when they are bionudges they must take the form of cognitive enhancement, there is some reason to believe that they might not be so harmful to autonomy. Hansen and Jespersen are concerned that a person may be unable to avoid the nudge if it acts on our type 2 processing because we would be unable to revert to the type 2 processing to resist it.

But if a cognitive enhancement improves our ability to learn, analyse information and assess our options, then in the same way that we are able to analyse our own logic when making decisions and change our minds, nudges that improve our cognition can enable us to overcome some of the ways our type 2 processing fails us. A cognitive enhancement that makes us take a more rational approach to risk assessment, for example, could counteract the examples about which Hansen and Jespersen raise concern. In light of this, type 2 non-transparent nudges may pose a risk only insofar as they are non-transparent to an agent who is unwilling or unable to assess their reasoning.

4.8. Settled and unsettled preferences

Engelen and Nys examine how nudges affect decisional and practical autonomy depending on the agent's preferences and values. Engelen and Nys' table, shown below, examines seven cases where an agent chooses between a vegetarian option (V) and a meat option (M) (Engelen and Nys, 147). This examination helps us understand how a nudge might undermine our ability to deliberate on our values and preferences when making a decision (our decisional autonomy) and our ability to act in accordance with our values and preferences (our practical autonomy), the latter of which Engelen and Nys refer to as 'autocracy' (Engelen and Nys, 145-7). The decisions in the table below that include an exclamation mark (!) are decisions where a nudge appears to have been the causal difference between the agent choosing (V) rather than (M). The use of question marks (?) show when the settled preference or how the decisional/practical autonomy is impacted is unclear.

| Case | Settled preference | Decision absent nudge | Nudged decision | Autonomy? | Autocracy? |
|------|--------------------|-----------------------|-----------------|-----------|----------------------|
| 1 | V | V | V | Respected | Respected |
| 2 | V | M | V (!) | Respected | Promoted |
| 3 | M | M | M | Respected | Respected |
| 4 | M | V | V | Respected | Not further impaired |
| 5 | M | M | V (!) | ? | ? |
| 6 | ? | V | V | ? | ? |
| 7 | ? | M | V (!) | ? | ? |

Case 1 shows an agent who prefers vegetarian food, who would pick the vegetarian option whether or not they were nudged and who successfully acts in accordance with their preference. Their decisional and practical autonomy is respected because the nudge did not successfully interfere with their ability to generate values and preferences. Case 2 involves a nudge helping the agent act in accordance with their established preferences, and so the nudge promotes the agent's practical autonomy. Perhaps the agent has a strong value for preventing climate change, but struggles to resist the temptation of eating meat. In this scenario, the nudge helps the agent to resist their urge and to act in accordance with their values. Case 3 involves an agent who prefers to eat meat and for whom the nudge for the vegetarian option is unsuccessful. In Case 4 the agent prefers meat but appears to choose the vegetarian option for some reason besides the

nudge. In this case, the agent's practical autonomy is no further undermined by the nudge because the existing choice environment (or the agent's own inability to act in accordance with their preferences), was the reason they chose the vegetarian option (Engelen and Nys, 148).

In case 5, the nudge is the factor that influences the agent to select the vegetarian option. In this case, Engelen and Nys argue that the agent's autonomy is undermined because, due to the nudge, they were unable to act on their settled preference (Engelen and Nys, 148). Engelen and Nys say only that this should not concern us as this kind of case is very unlikely to occur because nudges are, by definition, easily resisted (Engelen and Nys, 148-9). However, there is nothing immediately obvious about any aspect of choice architecture that makes us certain it is a nudge rather than a shove; even nudges that do not intend to undermine autonomy like in case 5 can get it wrong, and further, it is not made clear to us simply by viewing the nudge whether a choice architect intends to undermine our autonomy. As I argued in 4.2. *Nudging is not necessarily harmless to autonomy*, while Thaler and Sunstein may wish to define nudging in principle as not harmful to autonomy, in practice we can find nudges which are more or less harmful to autonomy. We should be wary of nudges that are too strong and that cause people to choose contrary to their values and preferences.

For cases 1-5, Engelen and Nys argue that one's decisional autonomy is unimpaired because the agent already has an established preference and so the nudge cannot influence their generating of a preference, only their ability to act on that preference in the moment (Engelen and Nys, 147-8). But we often have the chance to autonomously assess and reassess our set values and preferences, even after we receive a nudge. Engelen and Nys are mistaken in assuming that a nudge cannot influence a person's decisional autonomy if they already have set preferences, because if I choose (or if a nudge triggers me) to reassess my set preference when I encounter the nudge, then that nudge can influence the generation of a new preference or the reaffirmation of my existing preference. Given that some nudges are more influential than others, and some undermine autonomous deliberation more than others, just because an agent enters a choice environment with settled preferences does not mean their decisional autonomy cannot be undermined by a nudge (Blumenthal-Barby and Naik, 45; Hansen and Jespersen, 15). With that noted, I can examine both decisional and practical autonomy in regard to cases 6 and 7.

Cases 6 and 7 show agents who do not have settled preferences about choosing the meat or vegetarian option. These cases are not uncommon, because we can be 'undecided, conflicted or indifferent' about our preference for this decision (Engelen and Nys, 150). Engelen and Nys explain that when one does not have settled preferences, they are much more easily nudged (Engelen and Nys, 150). If the agent does not deliberate on their preferences before they select an option, it is difficult to say that the nudge undermines their practical autonomy, that is, it is difficult to say the nudge undermines their ability to act on their preferences (Engelen and Nys, 150). If the agent has no preferences, then an agent's selection cannot be said to cohere or conflict with their preferences.

Further, Engelen and Nys argue that even when an agent is undecided, nudges do not undermine their decisional autonomy. This is because (a) if the agent does not deliberate on their

preferences, the nudge cannot be said to undermine that process, and (b) if the agent does deliberate, then their decision-making is influenced by the choice environment whether or not it includes a nudge (Engelen and Nys, 150-1). Therefore, the nudge does not undermine decisional autonomy any more than a non-nudged choice environment. As in case 7, while a nudge could influence the agent's in-the-moment deliberation of their preferences to select the vegetarian option, if the nudge was not present, the agent, in a state of indecision and uncertainty, would be influenced by whatever existing choice architecture *is* present. If this is so, then Engelen and Nys say that we should not be concerned for the autonomy of the indecisive agent.

While point (a) in the paragraph above explains that a nudge cannot undermine the autonomy of an undecided agent if they do not deliberate on their preferences, it fails to account for the possibility that a nudge might be the influence that *precludes* an agent from engaging in deliberation. If, as Engelen and Nys claim, agents are more easily nudged when they do not have settled preferences, the presence of a nudge might be the difference not just between selecting (V) rather than (M), but between an agent pausing or not pausing to reflect on their preferences.

Engelen and Nys might reply, using point (b) from the paragraph above, that even if we did end up deliberating on our preferences, we would be just as influenced by the nudge during this deliberation as we would be by the existing choice environment. But as I have discussed in 4.4. *Heteronomous influence and cosmonomous choice environments*, some nudges *add* influence to a choice environment rather than simply replace it, and further, and in 4.5. *Reason-giving nudges are less harmful to autonomy than non-reason-giving nudges* and 4.6. *Influencing type 1 processing threatens autonomy more than influencing type 2 processing*, I note that some nudges promote and support autonomous deliberation such as reason-giving and type 2 nudges, while others bypass or undermine autonomous deliberation (Hausman and Welch, 133; Blumenthal-Barby and Naik, 45; Hansen and Jespersen, 15). A type 1 non-transparent bionudge that alters a person's emotions, or a type 1 transparent behavioural bionudge that impels an agent to take a certain action, may make them less likely to engage in reflective deliberation, for instance, and this would undermine their decisional autonomy. As such, we have no reason to think that, in principle, a nudge will never undermine one's decisional autonomy any more than the existing choice environment.

Finally, Engelen and Nys's examination of cases 1-7, frame nudges as being either successful or unsuccessful in entirely undermining an agent's practical or decisional autonomy based on their subsequent selection and their settled preference. But a nudge can undermine an agent's decisional autonomy (their ability to deliberate on their preferences) even if it is unsuccessful in changing the agent's preference. For instance, in case 1, the agent prefers the vegetarian option, and receives a nudge to select the vegetarian option, but if the agent chooses to deliberate on their preference for vegetarian options and if the nudge was one that deceives, distracts or makes it more cognitively costly for the agent to deliberate, then even though the agent sticks with their (V) decision, the degree to which they are able to autonomously deliberate could be undermined. Just because a nudge might not actually change someone's selection or the result of their reflection, it does not mean that the nudge does not undermine their autonomy.

4.9. The impact of nudging on global autonomy

So far, I have discussed how nudging can help or hinder one's ability to self-rule in the short-term, but as I have noted, autonomy can be understood in a 'global' long-term sense. Some events in the present (be they sacrifices of short-term autonomy or not), can increase future autonomy and increase autonomy overall. Some nudges, particularly those designed by public policy makers, increase long-term autonomy by increasing or preserving the agent's lifespan, cognitive and physical abilities, money and independence (Goodwin, 87). Capacities for decisional autonomy are improved by preserving or increasing cognition and reasoning, and capacities for exercising practical autonomy are improved by preserving or increasing physical abilities, education, legislation, and access to money.

The New Zealand Government pays money into New Zealanders' superannuation funds in proportion to each person's own contribution, to incentivise them to prepare for their retirement (New Zealand Inland Revenue Department, 2021). Laws around the advertisement, packaging and in-store visibility of tobacco help people to preserve their lifespan, health and money (Smokefree Environments and Regulated Products Regulations Act 2021, 9(1), 16-19; Smokefree Environments and Regulated Products Regulations 1990, 37(1)(b)). Some laws that do not outright prohibit certain behaviours, but require them to be carried out in certain ways, like the sale of tobacco, may undermine our short-term autonomy by nudging us away from acting on our settled preferences, but they can increase our long-term autonomy. While some nudges will decrease long-term autonomy, such as those that nudge people to select options that decrease their health, cognition and money, other nudges will improve long-term and overall autonomy, even though some of them will decrease short-term autonomy to do so.

Further, nudges can increase long-term autonomy by freeing up cognitive resources or 'mental bandwidth' for decisions deemed more important by the agent (Engelen and Nys, 142-3). If minor decisions can be made easy for agents, such as what kind of bread to buy or how quickly one should drive around an upcoming corner, then agents can spend their cognitive resources deliberating on decisions that they believe deserve their attention and where they wish to exercise maximum self-rule. To appreciate how selective deliberation can increase autonomy, consider how undermining it would be to approach every single decision with the caution and reflection of purchasing a house; the practical and cognitive costs we would impose on ourselves would undoubtedly conflict with our wishes.

However, nudges are also used in major, not just minor decisions (Engelen and Nys, 143). Some nudges, like making organ donation an opt-out choice or doctors framing treatment risk positively, relate to decisions that many would deem worthy of their best autonomous deliberation. While some nudges may free up cognitive resources for more important decisions, others may oversimplify or distort decisions that will have lasting and profound implications.

The use of bionudges can also impact long term autonomy. We might understand the effects of a bionudge as working in combination or in conflict with a person's natural desires, motivations and abilities. Imagine that Brian wants to improve his diet. Brian already makes a conscious self-

determined effort, but he still struggles to eat a complete healthy diet. Brian uses an emotional bionudge that makes him feel more positive about eating the kinds of food he wishes to eat. He experiences a decade of healthy eating, and gradually requires less conscious effort to maintain a healthy diet as he increasingly relies on his bionudged emotions about eating certain foods. If he stops using the bionudge, Brian may struggle to muster the willpower he had previously, thus eroding some of his capacity for autonomy. Bionudges could pose a general threat to long term autonomy in this way. While a bionudge might support short-term autonomy, if agents come to rely too heavily on bionudges to achieve their desired outcomes, agents may lose some of their natural ability to achieve those desired outcomes.

On the other hand, bionudges may increase long-term autonomy by building desirable habits. It's possible that when Brian stops using the bionudge he may find it much easier to stick to his health diet than before he used the bionudge because he was able to develop habits that continue to influence his choices. This kind of effect resembles the intended effect of the use of prescription drugs to support people detoxing from more harmful drugs. Methadone is used to reduce the symptoms of heroin withdrawal, and does not trigger feelings of euphoria like heroin does (United Kingdom National Public Health Service, 2021). This allows addicts to 'rewire' the reward neural pathways in their brains so that they can generate pleasure through other, more healthy activities and form associated habits. While Methadone itself can be addictive, and extra support is required when addicts cease using it, methadone is effective in forming positive habits that continue after an addict stops using methadone (United Kingdom National Public Health Service, 2021). We may also look at the use of antidepressants in this way. When a person is depressed they may struggle to motivate to carry out self-maintenance activities related to exercise, eating, hygiene and sleeping that support positive mood. The medium-term use of antidepressants can be used to help people carry out these activities, and form habits around them, before they stop taking antidepressants. Additionally, the use of gastric bands that restrict the stomach size of people with obesity act in this way, where the person eats less, initially because they more quickly feel full, and later because they develop healthier eating habits (Lang et al, 2002).

Cognitive, emotional and behavioural bionudges can each form positive habits in this way, but they may also develop the dependency I noted above that erodes autonomy. While empirical research on the use of bioenhancement would be required to confirm this, we can draw some expectations from existing research on the use of similar things such as methadone, antidepressants, and gastric bands. Bovens and Furedi argue that the long-term effect of living in a world where many decisions are structured to guide agents may be that people are infantilised and their capacity to self-rule is slowly eroded (Bovens, 214-5; Furedi, 2011). Bovens uses the analogy of a teacher who wants to gain the attention of their class (Bovens, 214). The teacher may be successful in the short-term by raising their voice, but if overused, the class may become desensitised and less attentive than they would have been if the teacher had used more sustainable methods. Likewise, Bovens's argues, widespread use of intentional choice architecture could help for a period but make people less inclined to exercise self-rule. If agents are spoon-fed advice and information, and if every complicating or cognitively challenging hurdle to autonomous deliberation is removed, they will become reliant on easy choice architecture.

Bovens and Furedi give no explicit evidence to support this claim. It is difficult to imagine that so many of life's decisions could be structured with choice architecture, but if they were, it is not clear that it would diminish our independence. For instance, choice architecture for some decisions will free up cognitive resources and allow us to consider decisions we might not otherwise have considered. Additionally, if there is a prevalence of nudges that trigger type 2 thought processing and more reflective consideration, we may instead become a population that possesses stronger skills of deliberation and exercises greater independence. It remains to be seen whether broad use of nudging for both minor and major decisions could result in a population with a diminished power and motivation to think for themselves, reflect on values and make decisions.

Further, it is possible that the broad use of nudges that target cognitive deficiencies might make people more aware of their deficiencies and cautious about their decision-making. Our capacity to self-rule is already subject to a number of barriers in the form of cognitive biases and irrational schema. Would the ongoing targeting of those heuristics make them a more prominent influence on our decisions or would targeting those cognitive deficiencies help people become more aware of the ways in which we can take greater control of ourselves? For example, it was learning about the amount of research and planning that goes into supermarket design that made me aware of the ways in which our schema affect decision-making. I am now more conscious of these affects when grocery shopping, but I had to become aware of them in order to increase my self-rule. While some nudges add extra influence into a choice environment, our irrational schema will play a role in our decision-making whether or not we are in an intentionally designed choice environment. For instance, my short-term bias is going to cause issues for me in preparing for retirement, whether or not companies intentionally exploit my short-term bias by tempting me to buy an expensive product and pay high interest to pay it off over time. It is possible, therefore, that we would observe an increased awareness of human psychology from more prevalent nudging, although we can increase awareness and understanding about our psychology and decision-making through methods other than exploitation, like education campaigns. On the other hand, if nudges become the norm, then it is possible that people will take less notice when their choice architecture has been designed to influence their decisions. It is difficult to predict and measure how our autonomy might be affected by widespread nudging, but it is worth evaluating.

4.10. Conclusion

This section provides some frameworks with which we can evaluate a bionudge's effect on autonomy. In a general sense, a single bionudge can impact a person's practical autonomy differently than it impacts their decisional autonomy; the means and method by which the bionudge functions can have an effect on autonomy; the agent's selection of the nudged option can cohere or conflict with an agent's autonomous values and desires; and the outcomes caused by the bionudged option can affect one's capacity for autonomy.

To be more specific, some bionudges inform, give reasons, or make costs and benefits salient to the agent and therefore they support decisional autonomy. The better they do these things, the more they increase autonomy *ceteris paribus*. Some bionudges operate via type 2 processing, which may pose a threat to decisional autonomy if agents cannot use type 2 processing to

overcome the bionudge. But as type 2 bionudges must be cognitive bionudges, they also empower agents' decisional autonomy by improving the cognitive toolkit an agent uses to deliberate. Other bionudges, such as type 1 bionudges, operate via arational influence and some by targeting irrational psychological schema, which *ceteris paribus* have a negative impact on autonomy.

Some of these bionudges exert more influence on agents than others, and so they may constitute a greater undermining of autonomy. The kinds of bionudges that are successfully designed to be difficult to avoid, do not just undermine autonomy but they constitute a manipulation of the agent. That is the case unless an agent voluntarily consents to being bionudged, in which case the bionudge does not undermine autonomy. Bionudges vary in influential strength, and individuals are more or less sensitive to them. The impact on the agent's autonomy depends not only on how sensitive they are, but also on the relationship between the goal of the bionudge and the agent's specific circumstances.

Some bionudges influence people to select options that cohere with their values and desires, and others influence them to select options that do not cohere. Under coherentist accounts, bionudges that do the former increase autonomy and those that do the latter diminish autonomy. Some bionudges offset or help us avoid the negative effects that irrational schema and urges impose on our autonomy, and some of those bionudges do this by leveraging that irrational schema. These bionudges increase autonomy, particularly if nudges that leverage our irrational schema to help off-set the autonomy-harmful effects of other irrational schema are transparent for the agent. The method of some bionudges exploit our irrational schema for positive autonomy outcomes, but could achieve those outcomes with methods that are less harmful to autonomy.

Some bionudges only replace the already existing negative influences on autonomy in the choice environment, while other bionudges increase or add an influence to the existing choice environment. Provided that the bionudges which replace influence do not nudge people toward selecting options which are more harmful to autonomy, and do not nudge in a more manipulative way, they have a neutral effect on autonomy. Whereas those bionudges that add extra influence to the choice environment undermine autonomy, *ceteris paribus*. The method and purpose of bionudges can be more or less transparent for agents, with those more transparent nudges being better for autonomy than less-transparent nudges.

Some bionudges exert influence on agents in decisions which they deem unimportant or pose trivial outcomes for autonomy, those bionudges have negligible negative effects on autonomy. In cases where an informed agent deliberately opts to let the choice architecture influence their decision, the effect on autonomy is neutral. Insofar as these bionudges for unimportant decisions free up cognitive resources for more important decisions, they increase autonomy.

However, if the broad use of bionudges slowly undermines agents' ability to think for themselves, reflect on their values and desires and act on them, then a bionudge may play some long-term role in the collective undermining of autonomy. On the other hand, if the widespread use of bionudging offers opportunities for agents to recognise and learn about how their decisions can

be influenced by choice architecture or form positive habits, then a bionudge may play some long-term role in increasing autonomy. These long-term possibilities are less helpful in determining the effect of a single bionudge on autonomy, but because we are investigating something as influential and unprecedented as potentially widespread bioenhancement, these broad and distant concerns require further assessment.

Chapter 5. Conclusion

Biological enhancement is not just an activity that science fiction writers predict for future humans, it is an activity many people already perform voluntarily and intentionally with illicit, prescription or mundane everyday drugs. New drugs, and technologies that edit genes or engineer the brain are being researched and a world where humans have an ability to profoundly change themselves appears likely. It is important that we anticipate how these tools will impact our lives.

While philosophers such as Persson and Savulescu and their critics worry about how compulsory bioenhancement will impact on humans' autonomy, I believe that scholars neglect the threat that even voluntary bioenhancement poses for self-rule and the need to identify strategies that mitigate this risk.

Recognising that desires can be of a higher order or more coherent within one's identity helps us to understand how behavioural nudges that intervene on a person's actions can be consistent with their autonomy. Understanding the difference between substantivism and proceduralism informs assessments of the impact of cognitive bioenhancement on one's reasoning and deliberation, and thus on their autonomy.

Enhancements that act on cognition, emotions and behaviour have differing impacts on one's autonomy. Cognitive enhancement poses the smallest threat to autonomy across all of my considerations. It will likely increase one's decisional autonomy and have marginal or no effect on practical autonomy and authenticity. Emotional enhancement's effect on autonomy is difficult to anticipate because the nature of the effect depends on what exactly is being enhanced and the situations in which decisions and actions are being made. Emotional states can influence one's ability to deliberate and act on their decisions. Of the three targets of enhancement, emotional enhancement is the most likely to compromise a person's authenticity by undermining their identity and triggering a sense of alienation. The only form of emotional enhancement whose effect on autonomy we could have some confidence about are enhancements that give agents greater awareness of or control over their emotions. Enhancements that target behaviour pose contrasting impacts to their decisional and practical autonomy respectively; they pose no threat to decision-making whatsoever, but they pose the greatest threat to practical autonomy. Likewise with authenticity, a person's identity may suffer only minor impacts from behavioural enhancement, whereas they may experience profound alienation from their inability to act on made decisions, particularly if nudged behaviours do not accord with their higher order or more coherent desires.

Bionudges offer strategies of bioenhancing that do not undermine autonomy. A single bionudge can impact one's practical autonomy differently than it impacts their decisional autonomy; the means and method by which the bionudge functions affects autonomy; the agent's selection of the nudged option can cohere or conflict with an agent's autonomous values and desires; and the outcomes caused by the bionudged option can affect one's capacity for autonomy. Bionudging, rather than also decreases the possibility that changes to the agent are so profound that they destroy their identity or make them feel alienated from their new cognitive abilities, emotions, or behaviour.

To be more specific, some bionudges inform, give reasons, or make costs and benefits salient to the agent, increasing decisional autonomy. The better they do these things, the more they increase autonomy. Bionudges that operate via type 2 processing can pose a minimal threat to autonomy if agents cannot use type 2 processing to overcome the bionudge. But as type 2 bionudges must be cognitive bionudges, they also empower agents' decisional autonomy, offsetting this threat. Type 1 bionudges operate via arational influence and some target irrational psychological schema, which *ceteris paribus* have a negative impact on autonomy.

Some bionudges exert more influence than others, and some agents are more sensitive to influence than others, so they may constitute a greater undermining of autonomy. Bionudges that are designed to be difficult to avoid do not just undermine autonomy, but they amount to manipulation. That is the case unless an agent voluntarily requests or consents to being bionudged.

Some bionudges influence people to select options that cohere with their values and desires, and others will influence them to select options that do not cohere. Some bionudges offset or help us avoid the negative effects that irrational schema and urges impose on our autonomy, and some of those bionudges do this by leveraging that irrational schema. These bionudges increase decisional autonomy, particularly if they are transparent for the agent. The method of some bionudges exploit our irrational schema for positive autonomy outcomes, but could achieve those outcomes with methods that are less harmful to autonomy.

Some bionudges replace existing influences on autonomy in the choice environment, while other bionudges will increase or add an influence into the existing choice environment. The former have no effect on autonomy, *ceteris paribus*. Whereas those bionudges that add extra influence to the choice environment undermine autonomy, *ceteris paribus*. The method and purpose of bionudges can be more or less transparent for agents, with those more transparent nudges being better for autonomy than less-transparent nudges.

Some bionudges exert influence on agents in decisions which they deem unimportant or pose trivial outcomes for autonomy, those bionudges have negligible negative effects on autonomy. In cases where an informed agent deliberately opts to let the choice architecture influence their decision, the effect on autonomy is neutral, perhaps even positive. These bionudges indirectly increase autonomy insofar as they free up cognitive resources for decisions self-adjudged to be more important.

My analysis has focused on how bioenhancements that target either cognition, emotions or behaviour impact autonomy. Further research may be focused on particular technologies and their unique impacts on autonomy. Research should also examine how autonomy is impacted by bioenhancements that target other human phenomena such as dispositions and habits, imagination, or the senses. Additionally, this research could assess how autonomy is impacted by bioenhancements that target specific combinations of phenomena.

Many bioenhancement technologies are still being developed and their full impacts remain unknown. This paper provides us with a framework with which we can assess the effect these technologies will have on personal autonomy, and identifies strategies of applying these bioenhancements in ways that protect and promote our autonomy.

Chapter 6. References

- Aristotle, et al. *The Ethics of Aristotle / (The Nicomachean Ethics) / Chase's Translation, Newly Revised*. Walter Scott Pub., 1890.
- Baddeley, A. "Working Memory: The Interface between Memory and Cognition." *Journal of Cognitive Neuroscience*, vol. 4, no. 3, pp. 281–288.
- Barlow, David H et al. *Abnormal Psychology: An Integrative Approach*. 8th ed., Cengage Learning, 2019.
- Bateson, Melissa, et al. "Do Images of 'Watching Eyes' Induce Behaviour That Is More Pro-Social or More Normative? A Field Experiment on Littering." *PLoS ONE*, vol. 8, no. 12, 2013, pp. 1–9., doi:10.1371/journal.pone.0082055.
- Battleday, R.M., and A.-K. Brem. "Modafinil for Cognitive Neuroenhancement in Healthy Non-Sleep-Deprived Subjects: A Systematic Review." *European Neuropsychopharmacology*, vol. 25, no. 11, Nov. 2015, p. 391., doi:10.1016/j.euroneuro.2015.12.023.
- Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*. Oxford University Press, 2019.
- Blumenthal-Barby, J, and A Naik. "In Defense of Nudge-Autonomy Compatibility ." *The American Journal of Bioethics*, vol. 15, no. 10, 2015, pp. 45–47.
- Bostrom, Nick, and Rebecca Roache. "Ethical Issues in Human Enhancement." *New Waves in Applied Ethics*, 2008, pp. 120–152.
- Bovens. "The Ethics of Nudge." *Preference Change: Approaches from Philosophy, Economics and Psychology*, vol. 42, 2009, pp. 207–219.
- Brailsford, Roseanna. "Exploring the Concept of Respect for Autonomy in the Context of New Zealand Health Promotion." *University of Auckland*, 2016.
- Bruns, H, et al. "Can Nudges Be Transparent and Yet Effective? ." *Journal of Economic Psychology*, vol. 65, 2018, pp. 41–59.
- Carmigniani, J, and Borko Furht . "Augmented Reality: An Overview." *Handbook of Augmented Reality*, 2011, pp. 3–46.
- Cauffman, Elizabeth, and Laurence Steinberg. "(IM)Maturity of Judgment in Adolescence: Why Adolescents May Be Less Culpable than Adults." *Behavioral Sciences & the Law*, vol. 18, no. 6, 5 Feb. 2001, pp. 741–760., doi:10.1002/bsl.416.
- Chuah, Lisa, et al. "Donepezil Improves Episodic Memory in Young Individuals Vulnerable to the Effects of Sleep Deprivation." *Sleep*, vol. 32, no. 8, Aug. 2009, pp. 999–1010.
- Colburn, Ben, and Margo Irvin. *Autonomy and Liberalism (Routledge Studies in Contemporary Philosophy ; 19)*. Routledge, 2010.

- Cook, Carley, et al. "Nonprescription Stimulant Use at a Public University: Students' Motives, Experiences, and Guilt." *Journal of Drug Issues*, vol. 51, no. 2, 2021, pp. 376–390., doi:10.1177/0022042620988107.
- Corchado, J M. "Five Types of Methodological Reasoning Systems – Cognitive Science." n.d..
- Crutchfield, Parker. "The Epistemology of Moral Bioenhancement." *Bioethics*, vol. 30, no. 6, July 2016, pp. 389–396.
- Daniels, Norman. "Normal Functioning and the Treatment-Enhancement Distinction." *Cambridge Quarterly of Healthcare Ethics*, vol. 9, no. 3, 2000, pp. 309–322., doi:10.1017/S0963180100903037.
- Del Missier, Fabio, et al. "The Multifold Relationship between Memory and Decision Making: An Individual-Differences Study." *PsycEXTRA Dataset*, 11 Sept. 2012, doi:10.1037/e519682015-035.
- Dunn, Jennifer R., and Maurice E. Schweitzer. "Feeling, Believing, and Trusting: The Influence of Emotion on Trust." *Journal of Personality and Social Psychology*, vol. 88, no. 5, 2005, pp. 736–748., doi:10.1037/e617892011-029.
- Dworkin, Gerald. *The Theory and Practice of Autonomy*. Cambridge University Press, 2001.
- Engelen, Bart, and Thomas Nys. "Nudging and Autonomy: Analyzing and Alleviating the Worries." *Review of Philosophy and Psychology*, vol. 11, 16 Dec. 2019, pp. 137–156.
- Farah, Martha J. "The Unknowns of Cognitive Enhancement." *Science*, vol. 350, no. 6259, 2015, pp. 379–380, doi:10.1126/science.aad5893.
- Feinberg, J. "Harm to Self." *The Moral Limits of Criminal Law*, vol. 3, 1986.
- Friedman, Marilyn. *Autonomy, Gender, Politics*. Oxford University Press, 2003.
- Funayama, M, et al. "Visuospatial Working Memory Dysfunction from Tapping Span Test as a Diagnostic Tool for Patients with Mild Posterior Cortical Atrophy." *Sci Rep-UK*, vol. 11, no. 1, 2021.
- Furedi, Frank. "Defending Moral Autonomy against an Army of Nudgers." *Spiked*, 20 Jan. 2011, www.spiked-online.com/2011/01/20/defending-moral-autonomy-against-an-army-of-nudgers/.
- Goodwin, Neva R., et al. *Microeconomics in Context*. Routledge, Taylor Et Francis Group, 2015.
- Goodwin, T. "Why We Should Reject 'Nudge.'" *Politics*, vol. 32, no. 2, 4 May 2012, pp. 85–92.
- Grunberg, Neil E., and Richard O. Straub. "The Role of Gender and Taste Class in the Effects of Stress on Eating." *Health Psychology*, vol. 11, no. 2, 1992, pp. 97–100., doi:10.1037/0278-6133.11.2.97.
- Grüne-Yanoff, T. "Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles." *Social Choice and Welfare*, vol. 28, no. 4, 2012, pp. 635–645.

- Hansen, Pelle, and Andreas Jespersen. "Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy." *European Journal of Risk Regulation*, vol. 4, no. 1, 2013, pp. 3–28.
- Hausman, D.M, and B Welch. "Debate: To Nudge or Not to Nudge." *The Journal of Political Philosophy*, vol. 18, no. 1, 8 Jan. 2010, pp. 123–136.
- Heersmink, Richard. "Extended Mind and Cognitive Enhancement: Moral Aspects of Cognitive Artifacts." *Phenomenology and the Cognitive Sciences*, vol. 16, no. 1, 2015, pp. 17–32., doi:10.1007/s11097-015-9448-5.
- Jebari, Karim. "What to Enhance: Behaviour, Emotion or Disposition?" *Neuroethics*, vol. 7, no. 3, 2014, pp. 253–261., doi:10.1007/s12152-014-9204-5.
- Johnson, Robert, and Adam Cureton. "Kant's Moral Philosophy." *Stanford Encyclopedia of Philosophy*, Stanford University, 21 Jan. 2022, plato.stanford.edu/entries/kant-moral/#Aut.
- Jotterand, F, and S Levin. "Moral Deficits, Moral Motivation and the Feasibility of Moral Bioenhancement." *Topoi*, vol. 38, 15 Mar. 2019, pp. 63–71.
- Keyes, Daniel. *Flowers for Algernon*. Harcourt, Brace & World, 1959.
- "KiwiSaver Benefits." *New Zealand Inland Revenue Department*, 28 Mar. 2021, www.ird.govt.nz/kiwisaver/kiwisaver-individuals/kiwisaver-benefits#:~:text=Government%20contribution,1%20July%20to%2030%20June.
- Kroese, Floor, et al. "Nudging Healthy Food Choices: a Field Experiment at the Train Station." *Journal of Public Health*, vol. 38, no. 2, June 2016, pp. 133–137.
- Lang, Thomas, et al. "Impact of Gastric Banding on Eating Behavior and Weight." *Obesity Surgery*, vol. 12, no. 1, 1 Feb. 2002, pp. 100–107., doi:10.1381/096089202321144667.
- Leahy, Robert L. et al. "Processes Underlying Depression: Risk Aversion, Emotional Schemas, and Psychological Flexibility". *International Journal of Cognitive Therapy*, vol 5, no. 4, 2012, pp. 362-379. Springer Nature, doi:10.1521/ijct.2012.5.4.362.
- Lerner, Jennifer S., et al. "Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions." *Psychological Science*, vol. 15, no. 5, 2004, pp. 337–341., doi:10.1111/j.0956-7976.2004.00679.x.
- Levy, Neil. "Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons." *Ergo*, vol. 6, no. 10, 2019, doi:https://doi.org/10.3998/ergo.12405314.0006.010.
- Linssen, A, et al. "Cognitive Effects of Methylphenidate in Healthy Volunteers: a Review of Single Dose Studies." *International Journal of Neuropsychopharmacology*, vol. 17, no. 6, June 2014, pp. 961–977.
- Litvak, Paul, et al. "Fuel in the Fire: How Anger Impacts Judgment and Decision-Making." In: Potegal, Michael, et al., *International Handbook of Anger: Constituent and Concomitant Biological, Psychological, and Social Processes*, Springer, 2010, pp. 287–310.

- Loewenstein, G, et al. "Risk as Feelings." *Psychological Bulletin*, vol. 127, no. 2, 2001, pp. 267–286.
- Manucia, Gloria K., et al. "Mood Influences on Helping: Direct Effects or Side Effects?" *Journal of Personality and Social Psychology*, vol. 46, no. 2, 1984, pp. 357–364., doi:10.1037/0022-3514.46.2.357.
- Matthews, Gerald, and Adrian Wells. "The Cognitive Science of Attention and Emotion." *Handbook of Cognition and Emotion*, 1999, pp. 171–192.
- "Methadone." *NHS*, United Kingdom National Health Service, 24 Mar. 2021, www.nhs.uk/medicines/methadone/.
- Mills, Chris. "The Heteronomy of Choice Architecture." *Review of Philosophy and Psychology*, vol. 6, no. 3, 2015, pp. 495–509., doi:10.1007/s13164-015-0242-7.
- Mills, Chris. "Why Nudges Matter: A Reply to Goodwin." *Politics*, vol. 33, no. 1, Feb. 2012, pp. 28–36.
- New Zealand Government, Smokefree Environments and Regulated Products Regulations Act 2021. 9(1), 16-19. New Zealand Government, 2021. <https://www.legislation.govt.nz/regulation/public/2021/0204/latest/whole.html#LMS525085>
- New Zealand Government, Smokefree Environments and Regulated Products Regulations Act 1990. 37(1)(b). New Zealand Government, 1990. <https://www.legislation.govt.nz/act/public/1990/0108/latest/DLM223191.html#DLM223190>
- New Zealand Government. "Gambling Act 2003." Section 301(1). 2003. <https://www.legislation.govt.nz/act/public/2003/0051/latest/DLM210246.html#DLM210246>
- Noggle, Robert. "Manipulation, Saliency, and Nudges." *Bioethics*, vol. 32, no. 3, 28 Dec. 2017, pp. 164–170.
- Persson, Ingmar, and Julian Savulescu. *Unfit for the Future: The Need for Moral Enhancement*. Oxford University Press, 2014.
- Ploug, T, and S Holm. "Doctors, Patients, and Nudging in the Clinical Context--Four Views on Nudging and Informed Consent." *The American Journal of Bioethics*, vol. 15, no. 10, 2015, pp. 28–38.
- Pugh, Jonathan. *Autonomy, Rationality, and Contemporary Bioethics*. Oxford University Press, 2020.
- Roberts, Carl, et al. "How Effective Are Pharmaceuticals for Cognitive Enhancement in Healthy Adults? A Series of Meta-Analyses of Cognitive Performance during Acute Administration of Modafinil, Methylphenidate and D-Amphetamine." *European Neuropsychopharmacology*, vol. 38, 21 July 2020, pp. 40–62.

- Rose, S. "Smart Drugs: Do They Work? Are They Ethical? Will They Be Legal?" *Nature Reviews Neuroscience*, vol. 3, 1 Dec. 2002, pp. 975–979.
- Sneddon, Andrew. *Autonomy*. Bloomsbury, 2013.
- Sparrow, R. "Egalitarianism and Moral Bioenhancement." *The American Journal of Bioethics*, vol. 14, no. 4, 14 Apr. 2014, pp. 20–28., www.tandfonline.com/doi/abs/10.1080/15265161.2014.889241.
- Squire, Larry R, et al. "Memory Consolidation." *Coldspring Harbor Perspectives in Biology*, vol. 3, no. 7, ser. 8, 2015. 8, pubmed.ncbi.nlm.nih.gov/26238360/.
- Steffel, Mary, et al. "Ethically Deployed Defaults: Transparency and Consumer Protection through Disclosure and Preference Articulation." *Journal of Marketing Research*, vol. 53, no. 5, 2016, pp. 865–880.
- Stokes, Maranda, and Sara Abdijadid. "Disulfiram." *StatPearls*, 24 Oct. 2022, www.ncbi.nlm.nih.gov/books/NBK459340/.
- Talha, M. "A History of Development in Brainchipsinpresent and Future." *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 2, 2020.
- Thaler, Richard H., and Cass R. Sunstein. *Nudge: The Final Edition*. Penguin Books, 2022.
- Tice, Dianne M., et al. "Emotional Distress Regulation Takes Precedence over Impulse Control: If You Feel Bad, Do It!" *Journal of Personality and Social Psychology*, vol. 80, no. 1, 2001, pp. 53–67., doi:10.1037/0022-3514.80.1.53.
- Vassanelli, Stefano. "Brain-Chip Interfaces: The Present and the Future." *Procedia Computer Science*, vol. 7, 2011, pp. 61–64., doi:10.1016/j.procs.2011.12.020.
- Wachner, Jonas, et al. "The Influence of Nudge Transparency on the Experience of Autonomy." *Comprehensive Results on Social Psychology*, vol. 4, no. 3, 25 Aug. 2020.
- Walker, Rebecca L. "Medical Ethics Needs a New View of Autonomy." *Journal of Medicine and Philosophy*, vol. 33, no. 6, 2008, pp. 594–608., doi:10.1093/jmp/jhn033.
- Weber, E. "Evidence-Based and Description-Based Perceptions of Long-Term Risk: Why Global Warming Does Not Scare Us (Yet)." *Climatic Change*, vol. 77, 2006, pp. 103–120.
- Westra, H., and S. Stewart. "As-Needed Use of Benzodiazepines in Managing Clinical Anxiety: Incidence and Implications." *Current Pharmaceutical Design*, vol. 8, no. 1, 2002, pp. 59–74., doi:10.2174/1381612023396744.
- "What Is ABS and ESC and How Do They Protect Me?" *New Zealand Automobile Association*, 17 Jan. 2022, www.aa.co.nz/cars/motoring-blog/what-is-abs-and-esc-and-how-do-they-protect-me/.
- White, Mark D. *The Manipulation of Choice: Ethics and Libertarian Paternalism*. Palgrave Macmillan, 2013.

- Wilkinson, T.M. "Nudging and Manipulation." *Political Studies*, vol. 61, no. 2, 7 Sept. 2012, pp. 341–355.
- Wilks, Anna. "Kantian Challenges for the Bioenhancement of Moral Autonomy." *Royal Institute of Philosophy Supplement*, vol. 83, 2018, pp. 121–43.
- Wiss, David A., et al. "Sugar Addiction: From Evolution to Revolution." *Frontiers in Psychiatry*, vol. 9, 2018, doi:10.3389/fpsy.2018.00545.
- Yiend, Jenny. "The Effects of Emotion on Attention: A Review of Attentional Processing of Emotional Information." *Cognition & Emotion*, vol. 24, no. 1, 2010, pp. 3–47., doi:10.1080/02699930903205698.