A Critique on Sugden's Use of Self-Control

Abstract:
This thesis' main objective is to investigate Sugden's use of self-control in various parts of his work. Three lines of argumentation are pursued. In the first strand, I argue that Sugden's (2017) method to identify self-control problems is identical to Frankfurt's meta-desire theory. Sugden identifies self-control problems as acknowledging two conflicting desires in a hot state, of which the (second-order) desire should be endorsed in some independently definable circumstances. This definition of self-control problems is inconsistent with the continuing agent in Sugden's opportunity criterion (2004, 2010, 2018B, Sections 5-8), and Sugden's interpretation of Hume's decision theory (2006, 2021). Sugden can only overcome this inconsistency by claiming that self-control problems are uncommonly endorsed. The second strand reviews Sugden's claim that people uncommonly self-acknowledge their self-control problems. This claim is assessed to be inadequately motivated because Sugden only investigates the take-up rate of self-constraint measures and defines self-control problems narrowly. Following these two strands, I propose a new definition of self-control problems rooted in psychology and Pettit (2006). Self-control problems should be interpreted as the failure to finish one's stable medium to long-term goals. Considering that many people suffer from self-control problems, more opportunities are sometimes normatively undesirable. With an adaptation of the capability approach of Sen, the *incapability approach*, I argue that some opportunities can be restricted if people collectively suffer from self-control problems.

Keywords: Sugden; Self-Control; Opportunity Criterion; Hume; Incapability Approach.

Justin G.P. Lever (424999)
Erasmus School of Philosophy (ESPhil)
Master Thesis Philosophy

Supervisor: dr. C. Binder
Advisor: Prof. dr. L. van Bunge

Words: 22.151
Date: 28 July 2021
Study Credits: 15 ECTS

Table of Contents

Acknowledgment iii
List of Abbreviations iv
1. Introduction 1
2. Libertarian Paternalism by Thaler & Sunstein 3
2.1. Introduction to Libertarian Paternalism 3
2.2. The Toolkit of Libertarian Paternalists: Nudges 4
2.3. System 1 and 2 5
2.3.1. Heuristics, biases, and temptation 6
2.3.2. Bounded rationality 6
2.4. The Need for Nudges 6
2.5. Summary of Section 2 7
3. Making Self-Controlled Choices with Economics 7
3.1. Choices in Neoclassical Economics 8
3.1.1. Preferences 8
3.1.2. Utility theory: axioms and rationality 8
3.1.3. Welfare and preferences 8
3.1.4. Take-away 9
3.2. Self-Control Problems in Economics 9
3.2.1. Neoclassical economics 9
3.2.2. Dual-selves 9
3.2.3. Meta-rankings 10
3.2.4. Self-control problems in Thaler and Sunstein. 10
3.3. Summary of Section 3 10
4. Sugden's (In)possibility to Identify Self-Control Problems 11
4.1. Introduction to Sugden 11
4.2. Jane in the Cafeteria 12
4.3. ILS Critique of the Inner Rational-Agent 12
4.4. Sugden's Argumentation on Self-Control 15
4.5. Combining ILS and Sugden (2017) 17
4.6. Summary of Section 4 18
5. The Continuing Self and Self-Control Problems 18
5.1. The Opportunity Criterion 18

5.2. The Continuing Agent 19
5.3. Compatibility of the Continuing Person and Self-Control Problems 20
5.4. Summary of Section 5 21
6. Hume and Self-control 22
6.1. Introduction to Hume 22
6.2. Sugden's Link to Hume 22
6.3. Sugden's Interpretation of Hume's Rationality 22
6.4. Hume on the Continuing Self 24
6.5. Hume's Philosophy and Self-Control Problems 25
6.6. Summary of Sections 4-6 25
7. On the (In)frequency of Self-Control Problems 26
7.1. Sugden's Perceived Frequency of Self-control Problems 26
7.2. A Higher Frequency of Self-control Problems 28
7.3. Summary of Section 7 29
8. A New Definition: Goal-Oriented Self 29
8.1. Self-Control Rooted in Psychology 29
8.2. A Goal-oriented Model in Psychology 30
8.3. The Goal-Oriented Self 31
9. Consequences of Another Self-Control Definition 32
9.1. The Frequency 32
9.2. An Agency Problem for Contractarians 33
9.3. The Capability Approach 33
9.4. The Incapability Approach 34
10. Possible Reaction of Sugden and a Response 35
10.1. An Initial Response 35
10.1.1. Sugden's question 35
10.1.2. A response 36
10.2. On a New Conception of Self-Control 36
10.2.1. Sugden's question 36
10.2.2. A response 37
11. Conclusion 37
12. Bibliography 39

Acknowledgment

This thesis marks the end of being an Erasmus University student. The past six years have taught me to see problems through the lens of behavioral economics and philosophy, and this thesis is a great way to demonstrate that I have learned a great deal in both scientific disciplines. Before continuing with the introduction of this thesis, I should like to thank my supervisor dr. Binder for her inspiration to write my thesis on Robert Sugden's book *The Community of Advantage* and her valuable feedback throughout the process. Furthermore, I would like to thank my advisor Prof. dr. Van Bunge for his time. Also, I would express my gratitude to Bernard Lafontaine, Clement Bezemer, Jeltje van der Haer, Rian Tosserams, and Wei-Jun Shu for having meaningful discussions about this thesis or other essays. Lastly, I would like to thank my family for their encouragement and ideas throughout the past six years.

List of Abbreviations

AJBT – As Judged By Themselves, As Judged By Herself, or As Judged by Himself

BE – Behavioral Economics

LP – Libertarian Paternalism

TS – Thaler & Sunstein

ILS – Infante, Lecouteux, and Sugden

1. Introduction

The phrase "making better decisions" lured me into the field of behavioral economics (BE). Who does not want to make the best, or at least, better, decisions? The ability of behavioral economists to make better decisions than standard economists originates from the additional ingredients that flavor the standard rules. Behavioral economists consider psychological insights and measure the in-field behavior of people, which led to the discovery of predictable human errors. We overweigh the probability of winning the lottery, sell our stocks too late if they decrease in value, and eat less food if the restaurant's plates are smaller. Keeping these structural errors in mind, Thaler and Sunstein (TS) explored what the consequences for (behavioral) welfare economics were given these findings, and they dedicated several articles and books like *Nudge* (2008) to this question. A concise answer is encompassed by the term: libertarian paternalism (LP). Libertarian entails that people have the freedom to choose, and paternalism means that the government should nudge to make people better off, as judged by themselves (AJBT). A nudge is a benevolent manipulation in a choice architecture, like switching the default to 'enroll'. For example, experiments with opt-out systems for pension savings and organ donation led, respectively, to more savings and more donations (Thaler & Benartzi, 2004). LP's goal is to make people better off without limiting their choice options. This goal may be praiseworthy, firm criticism arose on the justification of LP, of which the recent book *The Community of Advantage* by Robert Sugden (2018B) is prominent.

In *The Community of Advantage*, two strands are of main interest in this thesis. The first strand is that preferences are often context-dependent and there is no theoretically correct preference purification method. In a nutshell, latent preferences are one's *true* preferences, which are one's preferences if one were not prone to any reasoning imperfections. Sugden claims that the term AJBT is philosophically unattainable without an identification method to assert a correct or incorrect label for context-dependent preferences. The second strand is that creating more opportunities for individuals (i.e., through well-functioning institutions like markets) is normatively desirable. My engagement in this discussion is rooted in the special role that the self and self-control play in Sugden. This interest originates from my M.Sc. thesis for BE, where I investigated the correlation between a mathematical model of dynamic time preferences (the DI-index) and self-control problems (Lever, 2021).

In TS's LP, self-control is defined as a combination of temptation and intertemporal preferences. Nudges, then, are a desirable way to combat these self-control problems and make people better off, AJBT. Despite his reservations about nudging, Sugden (2017) acknowledges that people with self-control problems can be nudged to make them better off, AJBT. Sugden identifies self-control problems as having two conflicting desires in a hot state, of which the *true* desire, the desire one fails to act on, is developed during some *independently definable circumstances*. However, Sugden implies that self-acknowledged self-control problems occur infrequently or uncommonly (2017, p. 122, 2018A, pp. 11-12, 2018B, p. 81), and calls the need for self-regulation "exceptional" or "not very common" (2018B, p. 88, 154). Uncommonly and infrequently will be used interchangeably. I found this stance peculiar for various reasons. On the one hand, Sugden's opportunity criterion and interpretation of Hume seem irreconcilable with self-control problems. The continuing agent, which is the foundational agency of the opportunity criterion, is described as a "continuous locus of responsibility" (Sugden, 2004, 2010, 2018B). Suffering from self-control problems does not strike me as being a continuous locus of responsibility. Also, Sugden likes Hume's philosophy and psychology with a special emphasis

on his decision theory (Sugden, 2018, p. 96, 2020, p. 69). However, Hume called the self a "fiction of the imagination (1740/2009, p. 320) ", and his decision theory without a stable self is a poor candidate to identify self-control problems. On the other hand, the statement that people uncommonly endorse their self-control problems is at odds with other findings. For instance, Sunstein (2018) found in a survey that 70% agrees with the statement that they experience self-control problems. The discrepancy between Sunstein's survey results and Sugden's statement could be explained through a definition of self-control problems. Unfortunately, Sugden does not investigate the robustness of his arguments if a different definition of self-control is used, nor does he weigh the (dis)advantages of committing to a theory of self-control. This thesis will explore this gap in the literature. Therefore, the research question is: *To what extent does an alternative concept of self-control refute Sugden's defense of an opportunity-based market and Hume's decision theory?*

To answer this research question, I will present three main strands in this thesis. Firstly, Sugden agrees that self-control is a valid criterion to make people better off, AJBT, albeit self-control problems occur infrequently. I argue that Sugden's (2017) identification method with two conflicting desires repostulates the meta-desire theory of Frankfurt (1971). Subsequently, I argue that this meta-desire strategy of Sugden conflicts with Sugden's continuing agent in the opportunity criterion (2004, 2018B, Section 5.5), and Hume's decision theory (2006, 2021). One strategy to overcome this problem is that Sugden assumes that self-control problems occur infrequently in order to make this philosophical incoherency less relevant.

My second line of reasoning is that Sugden's recognition of the uncommon endorsement of self-control problems is incoherent. Sugden's approach is that, intuitively to him, it is uncommon that people would say their behavior stems from self-control problems (2017, p. 121). His methodology to determine how often self-control problems occur is to review the take-up rate of self-constraint measures. With this method, he forgets two other categories of self-control problems. Namely, the category of people that do not change their behavior when confronted with self-control problems, or people that try to exert self-command and fail. I argue that Sugden's methodology is insufficient to claim the infrequency of self-control problems. Furthermore, I contend that his definition of self-control of being unable to resist temptations limits the expected frequency of self-control problems. Other disciplines, like psychology, define self-control problems in a broader way.

In the third part, a new conception of self-control is explored, where self-control can be defined as the ability to prevent encountering or resist temptations that undermine one's long-term goals. This definition is inspired by a psychological approach promoted by Marange and Baumeister (2016). Coming back to Sugden's theory, this definition has two consequences. (1) A significantly large group can be classified as having self-control problems, AJBT. (2) A consequence of this definition is that the continuing self and consumer sovereignty are put under pressure. Since the contractarian agency model is unsuitable to deal with collective self-control problems, a new solution has to be found. Inspired by Sen, I argue with the *incapability approach* that some opportunities diminish the effective freedom of people. For instance, the widespread availability of unhealthy food diminishes the effective freedom to stay on a desirable weight level.

This thesis is divided into eleven sections. The next section, Section 2, will provide the background of the LP project by TS. To understand TS' approach and the conception of self-control better, an introduction to choices in neoclassical economics and self-control problems in BE is given in Section

3. Section 4 is dedicated to the identification of self-control problems according to Infante, Lecouteux, and Sugden (ILS or Infante et al.) and Sugden's (2017) method. Section 5 will examine the irreconcilable relation between self-control problems and the continuing self. The latter is the basis of Sugden's opportunity criterion. The relation between self-control problems and Hume is assessed in Section 6. Afterward, Sugden's thesis that self-control problems are uncommonly endorsed is examined and criticized in Section 7. Section 8 will introduce self-control from a psychological perspective. In Section 9, I argue that some opportunities should be limited in line with the incapability approach. A possible response from Sugden will follow in Section 10. In Section 11, this paper ends with a conclusion that less opportunities are normatively desirable in areas where people collectively suffer from self-control problems.

2. Libertarian Paternalism by Thaler & Sunstein

The term LP has been mentioned in the introduction of this thesis but needs further clarification, which will be given in Section 2.1. Afterward, Section 2.2 is dedicated to defining nudges. How nudges work according to TS, through dual-process theories, will be discussed in Section 2.3. When nudges should be applied, is discussed in Section 2.4. In Section 2.5, a conclusion follows.

2.1. Introduction to Libertarian Paternalism

LP is first introduced by TS (2003) and subsequently defended in several papers (Sunstein and Thaler, 2003) and their bestseller book *Nudge* (TS, 2008). The book starts pretentiously, where TS introduce LP as a "newfound power" that BE has developed over the last decades (2008, p. 2). BE has found nudges, that are subtle changes in a choice architecture, systematically influence how people behave. The goal of TS is to develop a coherent theory on what *any* choice architect should do with this power to influence her customers or citizens. The term choice architect usually takes the form of a policy advisor or a social planner, but the term is broadly defined. For instance, restaurant owners are also choice architects because they can influence the customers' choice with the display of food or the menu card. They coin the term LP, after which they define both words separately:

> The libertarian aspect of our strategies lies in the straightforward instance that, in general, people should be free to do what they like – and to opt out of undesirable arrangements if they want to do so (2008, p. 5).

They continue with the term paternalism:

> The paternalistic aspect lies in the claim that it is legitimate for choice architects to try to influence people's behavior in order to make their lives longer, healthier, and better. … In our understanding, a policy is 'paternalistic' if it tries to influence choices in a way that will make choosers better off, *as judged by themselves* (Idem, emphasis in original).

Thaler and Sunstein summarize the core of LP as follows:

> [1] That is, we emphasize the possibility that in some cases individuals make inferior choices, choices that they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower. [2] Once it is understood that some organizational decisions are inevitable, that a form of paternalism cannot be avoided, and [3] that the alternatives to paternalism (such as choosing options to make people sick, obese, or generally worse off) are

> unattractive, we can abandon the less interesting question of whether to be paternalistic or not and turn to the more constructive question of how to choose among paternalistic options (TS, 2003, p. 175, numbers added).

In the last quote, TS explain three of the four premises of their conclusion that there are only unattractive alternatives to LP. The first premise is that individuals make inferior choices, which contrasts with standard economists, who often assume that individuals make choices that make them best off. In their book *Nudge*, TS consider that a large proportion, currently 39%, of the *world* population is overweight (World Health Organization, 2020). Although being overweight itself is not per definition irrational, it is hard to argue that these aggregate choices are the result of people pursuing their best interests (TS, 2008, pp. 7-8). TS argue that if people had complete information, unlimited cognitive abilities, and no lack of willpower, obesity would become infrequent to non-existing. The second claim is that many people are choice architects, some of those unaware, who have "the responsibility for organizing the context in which people make decisions (TS, 2008, p. 3)." Returning to the example of a restaurant, the menu card can be ordered in several ways. For instance, alphabetically, profit-enhancing by putting the most expensive meals on top, maliciously by highlighting calorie rich meals, or animal-friendly by creating a vegetarian page. Among those options, TS plea that the welfare-enhancing way is the best way, which is the third premise. The fourth premise, which was mentioned at the beginning of this section, is that BE have the newly found power to influence people's decisions. Thus, the design of the menu card of a restaurant will affect the choices that people make, implying context-dependency. The last remark is that LP does not alter the existing conditions in other ways. For instance, LP does not involve any reduction of the number of options in the restaurant, nor does it involve changing monetary incentives, like lowering the price.

So far, the four premises of the LP discussed are: (1) People do not always choose what is in their best interest; (2) People must make choices in a particular choice environment; (3) Welfare-enhancement is the best form of choice architecture; (4) People's choices are often context-dependent. With these four premises, TS draw the conclusion that there is no viable alternative to LP. Any choice architect should present choices in a welfare-enhancing way. The term for influencing this decision is called nudge. In the next subsection, more details will be given on the definition of nudges.

2.2. The Toolkit of Libertarian Paternalists: Nudges

The term nudge was summarized by the terms "subtle changes" or "influencing decisions", but both terms do not capture the essence fully and this section is devoted to defining nudges.

TS's emphasis on changes in behavior returns in their working definition of nudges. They define nudges as: "Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives (2008, p. 6)." This definition captures the two premises that people must make choices and that these choices are context-dependent. However, this definition misses another essence of the LP, captured by premise three. LP was meant to improve the welfare of people, AJBT. In the example of a school cafeteria, it may seem obvious that the director does not want her students to become fat. Here, a nudge is, obviously, an aspect of the choice architect to make the students eat healthier. However, extending nudges to the private sector becomes troublesome. Does LP entail that private restaurants or grocery stores should nudge people towards buying the best (i.e., healthiest) products, or are they allowed to nudge people

towards profitable products? If a supermarket sells candy next to the cashier's counter, it certainly does not encompass LP. The counter candy is relatively expensive and sugar-rich, which are both undesirable features for many customers. For this reason, there are two options to define nudges. Either, one accepts that nudging is disentangled from LP, which results in nudges being defined as any changes in the choice architecture, or one defines nudges only for measures targeted at making one better off, AJBT. Since this thesis will discuss nudging in relation to LP, the latter is the preferred option. This position is also favored by Hansen (2016), who also investigates more (dis)advantages of what a nudge should encompass. As he comes to a similar conclusion, he defines nudge as follows:

> A nudge is a function of (I) any attempt at influencing people's judgment, choice or behavior in a predictable way (1) that is made possible because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally *in their own declared self-interests* and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts (Hansen, 2016, p. 158, emphasis added).

In this thesis, I will use the word nudge only in the context of LP like the aforementioned definition. The terms evil, malicious, or profit-enhancing nudge is *contradictio in terminis*, because it indicates that other goals are prioritized over well-being, and that contrasts LP. A minor addition to Hansen's definition is the emphasis on *their own declared self-interests*. As will be discussed in Section 4, Sugden's criticism aims at the method to determine *their own declared self-interests*, or as TS call it, the criterion "that will make choosers better off, AJBT (2008, p. 5)." Other distinctions can also be made in nudging, like if the nudges are aimed at rational persuasion or psychological foibles (Hausman & Welch, 2010), but this line of reasoning is not of direct interest to the AJBT criterion.

All in all, according to LP, choice architects should nudge humans to make their decisions better. Hansen's definition implies that nudges make use of our boundaries, biases, routines, and habits, and are aimed to make people better off, AJBT. In subsection 2.3, these boundaries, biases, routines, and habits in relation to brain processes will be further investigated.

2.3. System 1 and 2

The roots of why humans make suboptimal decisions lie, according to TS, in our brain processes (2008, pp. 21-24). To be more precise, TS follow a dual-process theory that has been described by Kahneman's book (2011). The title of the book summarizes their main idea accurately: *Thinking, Fast and Slow*.

On the one hand, humans can think fast (or rather: think little), and thinking fast is correlated with biases and temptation. On the other hand, when people think slow (i.e., deliberately take time to think), decisions are generally better, but still prone to mistakes due to bounded rationality. Before elaborating on biases and bounded rationality, we will first examine the contours of the dual-process theory. This binary thinking coincides with the dual-process theory. In dual-process theory, the first system is an automatic system that is intuitive, fast, typically biased, independent of cognitive abilities, and is associated with heuristics and biases. The automatic system is responsible for the nervous feeling when an airplane suddenly has turbulence. The anxiety of crashing during a turbulent flight is a natural, automatic reaction, but upon deliberation, this anxiety is statistically quite strange. According to an estimation in the Economist (2015) called *a crash course in probability*, the chance is 1 in 5.4 million,

meaning that one could fly, on average, 14.716 years before crashing. The second system is the reflective system and is associated with (instrumental) rationality and deliberation. It costs time and effort, of which we have limited capacity. Furthermore, it can be associated with cognitive abilities. For instance, the process of writing a philosophy master thesis can be associated with system 2 thinking. Since not everyone has the cognitive abilities of Einstein, people make errors due to bounded rationality.

It is not in the scope of this thesis to assess the validity of this division between systems 1 and 2, and the corresponding link with biases and rationality. However, it may be worthwhile to mention that some scholars are committed to different theories, like the single-process theory (Osman, 2004; Kruglanski & Gigerenzer, 2011). Also, it might be problematic to only associate one system with rationality.[1] Lastly, some argue that system 2 is an addition to system 1 and that system 2 does not act independently (Infante et al., 2016A, p. 15). Evans and Stanovic (2013) provide further analysis of this topic.

2.3.1. Heuristics, biases, and temptation

Heuristics and biases are associated with system 1 thinking. In this paragraph, a few biases are described but this list is certainly not exhaustive. In a famous paper, Tversky and Kahneman (1974) set out several heuristics like anchoring, availability, and representative heuristic and biases like loss-aversion, status quo bias, and overconfidence. One bias is loss-aversion, which became later part of the (cumulative) prospect theory. This descriptive theory by Tversky and Kahneman (1992) is designed for decisions under risk. Their central insight is that people dislike losing more than they like winning. If Maddy would lose 10 euros and win 10 euros in a casino subsequently, the chances are high that she would be sad instead of having a neutral feeling. Simultaneously, the perception of risk is also not conceived linearly. Small probabilities tend to be overweighted, and large probabilities are underweighted, which may explain why Maddy is in the casino in the first place. Sugden will criticize the normative refusal of this prospect theory in his critique on preference purification, which will be discussed in Section 4.3.

2.3.2. Bounded rationality

After examining the contours of system 1, we can explore the contours of system 2. System 2 is associated with being boundedly rational. Boundedly rational means that, while having thought deliberately about a decision, one still makes a reasoning error relative to neoclassical economics. Thaler and Sunstein identify five circumstances in which individuals tend to be boundedly rational, which depend on the following: benefits and costs are separated over time, the degree of difficulty, the frequency of decisions, the frequency and level of feedback, and unfamiliarity of outcomes.

2.4. The Need for Nudges

After TS's analysis of when people make mistakes due to biases or bounded rationality, TS try to provide guidelines on when nudges are needed to make people better off, AJBT. TS agree with the solution of asymmetric paternalism. Asymmetric paternalism is defined as "A regulation … [that]

[1] The idea that rationality corresponds to one process of thinking can be seen as an oversimplification. Evans and Stanovic write: "Rationality is an organismic-level concept and should never be used to label a subpersonal process (i.e., a type of processing). As an example, people's face recognition systems are neither rational nor irrational. They are, instead, either efficient or inefficient. Subprocesses of the brain do not display rational or irrational properties per se, although they may contribute in one way or another to personal decisions or beliefs that could be characterized as such (2013, p. 229)."

creates large benefits for those who make errors, while imposing little or no harm on those who are fully rational (Camerer et al., 2003, p. 1212)." Asymmetric paternalism is measured as a sum, where the benefits of boundedly rational agents should be larger than the total costs for rational agents, the implementation cost, and firms' change in profits. Boundedly rational is here loosely interpreted as all biases, temptation, and bounded rationality together. Coming back to the statistics of airplane crashes, an asymmetrical paternalist might prohibit the selling of airplane insurance at the airport. Only irrational agents opt for flight insurance, because one has to extremely overweigh the likelihood of a crash to financially benefit from this deal. The aggregate benefit of prohibiting this insurance outweighs the total costs. Technically, TS's LP cannot accommodate any reductions in the choice set, but they often flirt with these ideas in their book in addition to their LP approach (Sugden, 2009, p. 369).

Rather, the core of LP is that nudges are required when humans do not act like 'econs'. TS refer to an econ as an imaginary person as smart as Albert Einstein, with the memory of a supercomputer, and the willpower of Gandhi (2008, pp. 7-9). The econ is, therefore, a rational textbook-like individual who has no trouble with difficult problems. Humans do not possess these characteristics and suffer from behavioral biases, bounded rationality, and self-control problems as discussed in Section 2.3. The assumption is that individuals would make econ-like choices (i.e., error-free) "if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control (TS, 2008, p. 5)". Sugden calls these conditions "reasoning imperfections" (2018B, p. 55). TS offer examples where people can be made better off, AJBT, throughout their book. For instance, they mention self-control problems, mortgages, retirement savings, credit markets, road signs to watch one's left in London, organ donations, public urinals, energy-saving, et cetera. These examples are convincing if one reads the book as a manager in one's leisure time, but the lack of formal criteria is "frustratingly vague" for philosophers (Sugden, 2009, p. 370).

2.5. Summary of Section 2

The goal of Section 2 was to summarize the line of argument of TS. TS maintain that LP is an inevitable new way of thinking. By tinkering with the choice architecture, planners and policymakers may influence the behavior of their citizens or customers by their self-declared interests through nudging. If we "had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control", nudging would have been unnecessary. However, humans make mistakes and nudging uses those mistakes, that are identified by BE, to influence behavior in a welfare-improving manner. The goal of these nudges is per definition to make people better off, AJBT. One of these criteria is that people suffer from self-control problems. In Section 3, choices in neoclassical economics and self-control problems will be further analyzed.

3. Making Self-Controlled Choices with Economics

The aim of *Nudge* by TS (2008) is to nudge the error-prone human to let him become more like an econ. In Section 3.1, a concise introduction to making choices in Neoclassical economics and BE is given in order to understand the concept econ better. Subsequently, self-control problems from four economic perspectives are defined in Section 3.2.

3.1. Choices in Neoclassical Economics

3.1.1. Preferences

One could say that a choice consists out of two or more options, and that one ought to choose the option that he or she likes best. Neoclassical economists capture the relation between two or more alternatives in a similar manner with the concept preference. In this thesis, Hausman's way of preferences will be used, since this concept is discussed by Sugden later (Infante et al., 2016A, pp. 10-11). If Maddy prefers an apple (A) to a Snicker bar (S), an economist would say that Maddy has a preference where A is preferred over S, all things considered. A preference is subjective, meaning that other people can have different preferences. Preferences are not limited to post-lunch snacks but can be about everything, ranging from money, time, hospital treatments, and ice cream cones to cinema vouchers. If Maddy says that she classifies a snicker as tasty without comparing it to another food item, it is not considered as a preference, because a preference consists of at least two options. Neoclassical economists assume that people maximize their utility in their choice (i.e., people choose what they like most), and take all relevant factors into account. Since people choose what they like, economists infer that people's preference equals what they choose. This is what economists refer to as revealed preferences.

3.1.2. Utility theory: axioms and rationality

Neoclassical economics has axioms that are the starting points for preferences to hold. If these axioms hold, also the expected utility theory of Neumann and Morgenstern holds. The axioms generally are:

(1) Stability, meaning that preferences do not change immensely in a short period (i.e., no dynamic inconsistency);
(2) Context-independency, meaning that preferences are not affected by irrelevant contextual clues;
(3) Internally-consistent, meaning that preferences satisfy the properties of completeness and transitivity;
(4) Monotonicity, meaning that more is better than less;
(5) Well-informed beliefs and updating according to Bayes' law, meaning that if the options include uncertainty, individuals are assumed to have well-informed beliefs about this uncertainty. If an event unfolds itself, individuals update their beliefs according to Bayes' law.

If these axioms hold, the agent is per definition rational. Other terms for rationality are coherency or consistency, which are somewhat similar and capture the essence intuitively better. However, not all philosophers agree on those terms. Sugden, for instance, calls these preferences 'integrated' and dislikes the terms consistency and coherency because they commit to a single system of preferences (2018B, p. 7). Sugden commits to a non-rational decision theory by Hume, which will be explored in Section 6. In this thesis, the term rationality is interpreted as formal rationality instead of instrumental rationality.

3.1.3. Welfare and preferences

A heated discussion in the philosophical literature is how preference satisfaction is related to well-being. This discussion depends on the definition of well-being (i.e., does well-being equate to happiness or other components), and the definition of preference satisfaction (i.e., what is a preference). A prominent position in this debate is Hausman's relation between preferences and well-being. He says that "Satisfying actual preferences is nevertheless typically a good way to make people

better off, since people's preferences often are self-interested and their beliefs often reliable (2010, p. 341)." A note is that these preferences should be "self-interested, rational and well-informed preferences (Idem)." As will be discussed in Section 4, the discussion about making people better off, AJBT, will revolve around their preferences. Throughout this thesis, preference satisfaction, and in particular latent preferences, are assumed to make people better off, AJBT, which is in line with TS. However, for many economists like Sugden, preference satisfaction does not equal well-being, and other economists like Hausman consider it as a valuable indicator. This discussion is not in the scope of this thesis. In the next part, an introduction to self-control problems in economics is given.

3.1.4. Take-away

This subsection introduced Hausman's concept of preference that will be used by ILS. Also, among the axioms of rational choice theory is the axiom context-independency. This axiom will play a large role in the recognition of self-control problems in Section 4. Lastly, the assumption was made that the satisfaction of latent preferences makes individuals better off. Having considered how choices are made in neoclassical economics, we should review how economics looks at self-control problems. A review of their methods will be beneficial to understand how Sugden defines self-control problems in Section 4.

3.2. Self-Control Problems in Economics

In the economic literature on self-control problems, many consider the example of Ulysses (i.e., Strotz, 1955). Ulysses asked his crew to tie him to the ship's mast, put wax in their ears, and disobey his orders at all costs when they would sail past the magical Sirens. The Sirens could allure anyone with their beautiful songs to stop and kill anyone who was attracted to stop. What would have happened if Ulysses did not pre-commit himself according to different economic theories?

3.2.1. Neoclassical economics

In neoclassical economics, self-control problems are considered intertemporal inconsistencies. Since econs, the perfectly rational beings in neoclassical economics, have well-defined preferences over time, technically, no self-control problems should exist. Hence, Ulysses has no problem refusing the Sirens.

3.2.2. Dual-selves

One method to explain self-control problems for theoretical economists is to assume that an agent consist out dual, or multiple, selves. The dual-self model is used by economists to differentiate between a planner and doer. The self who lives now is the doer, and the self who plans the future is the planner. Here, the planner suffers from the present bias. O'Donoghue and Rabin define this bias as follows: "When considering trade-offs between two future moments, present-biased preferences give stronger relative weight to the earlier moment as it gets closer (1999, p. 103)." In the self-control literature, the long run-self is usually seen as the superior self, because most long-term goals require effort and dedication. For a detailed overview of the use of discount rates and changes in impatience, I refer to Lever (2021, pp. 3-6, 10-12).

The multiple-selves model interprets unstable or context-dependent preferences as integrated preferences through a dynamically inconsistent discount function. Furthermore, BE has to assume that the agent is naïve (i.e., unaware of or fails to act on this change). Thus, if Ulysses is decreasingly impatient and unaware of this change, he suffers from a self-control problem. In this case, Ulysses

would say that he will sail past the Sirens, but at the moment when they arrive at the Sirens, he changes his mind and stops at the Sirens.

3.2.3. Meta-rankings

A form of meta-ranking or meta-desires, which will be used interchangeably in this thesis, is advocated by Frankfurt (1971), who used this theory to defend his compatibilist position in the free will debate. Frankfurt argued that the essential difference between persons and other creatures lies in the capacity to form second-order desires. He writes:

> Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are (Frankfurt, 1971, p. 7, emphasis in original).

He distinguishes two kinds of desires. The first-order desire aims at objects or generally things that are not desires. Furthermore, it is more than a mere inclination towards a particular activity, but it is "an *effective* desire – one that moves (or will or would move) a person all the way to action (Frankfurt, 1971, p. 8)." The word desire is throughout my thesis defined as an *effective* desire, or a Frankfurtian desire, which is distinct from the way Hume uses desire in Section 6 of this thesis. Coming back to the story, Ulysses has a first-order desire to be lured by the Sirens. However, persons have second-order desires, which are desires about first-order desires. Second-order desires wish that the first-order desire would be different. When confronted with the Sirens, Ulysses (in the second-order) wants that Ulysses (in the first-order) wants that his desire to resist is stronger than his desire to be attracted. In the original story, Ulysses acts upon his second-order desire and forms a second-order volition. This volition means that Ulysses acts upon this second-order desire, and pre-commits himself by demanding his crew to tie him to the ship's mast. The second-order desire is not equal to the first-order desire, and serves as a normative higher-ranked self. For instance, Frankfurt calls people that fail to act upon second-order desires "wantons" (1971, p. 11). I will argue that Sugden supports a form of meta-ranking, which will be discussed in Section 4.3.

3.2.4. Self-control problems in Thaler and Sunstein.

According to TS, self-control problems are a mixture of temptation and mindlessness (2008, pp. 44-47). TS write: "Ulysses successfully solved his [self-control] problem. For the most of us, however, self-control issues arise because we underestimate the effect of arousal (2008, p. 45)." The effect of arousal is a reference to Loewenstein, who uses the 'hot' and 'cold' state (1996). The cold state is described as the stage when one is not aroused, which means that Ulysses pre-emptively binds himself to the mast. The hot state occurs when the Sirens sing for Ulysses, and he commands his disobeying crew to stop the ship. The implicit premise is that the hot state leads to self-control problems because arousal impairs our original preferences of the cold state. Thaler and Sunstein use a combination of dual-selves and meta-ranking. The planner, who plans during the cold state, is considered as the higher-self in contrast to the doer, who acts during the hot state and is considered as the lower-self. Economic theorists only need to assume the dual-self or the meta-ranking, but a combination of both theories is also possible, as demonstrated by Thaler and Sunstein.

3.3. Summary of Section 3

This section has provided some background on how self-control problems work in economics. Especially the concepts of preference, and the axioms stability and context-independency play a large

role in the ILS's conception of the inner rational-agent, which will be discussed in Section 4.3. Furthermore, the conceptions of dual-selves and meta-rankings will be used to recognize Sugden's form of meta-rankings in Section 4.4, and assess the implications of the continuing agent on dual-selves and meta-rankings in Section 5. After the introductory Sections 2 and 3 that helped to sketch the context of Sugden's philosophy, the first strand of this thesis will be investigated in Sections 4-6.

4. Sugden's (In)possibility to Identify Self-Control Problems

As noted in the introduction of this thesis, the first strand of this thesis was that Sugden is incoherent in his use of self-control. To discuss the incoherency, a look is needed into Sugden's definition of self-control problems. This section is devoted to providing Sugden's method to identify self-acknowledged self-control problems with Sugden (2017, 2018A) and the impossibility to identify self-control problems through an external observer with Infante et al. (2016A, 2016B). Before going into depth, an introduction to Sugden and his book *The Community of Advantage* will be given in Section 4.1. In Section 4.2, the example of Jane in the cafeteria will be presented. The question will be raised how the cafeteria owner can nudge Jane to make her better off, AJBT. The answer of ILS is that an outsider cannot say that Jane is better off, AJBT, without invoking a theoretically unattainable theory. This unattainable theory, also known as preference purification, will be discussed in Section 4.3. Sugden's method to identify self-acknowledged self-control problems is the central theme in Section 4.4. I argue that this method is based on a meta-desire theory of Frankfurt (1971). In Section 4.5, I argue that Sugden fails to correctly apply his identification method (2017) to an example in his book. The conclusion that Sugden commits to meta-desires is drawn in Section 4.6.

4.1. Introduction to Sugden

Robert Sugden has dedicated his career to developing philosophical perspectives on findings of experimental economics, Sugden's preferred name for BE. The book *The Community of Advantage* (2018B) is Sugden's answer to reconcile the insights of the mainstream BE with a liberal tradition in economics. Among other philosophers, he builds upon the liberal tradition of Smith, Hume, Mill, and Hayek, as well as contractarians such as Buchanan, Hobbes, and Rawls. His philosophical position is in that sense unique, because he combines contractarianism and his refusal of rational choice theory in normative analysis. The sub-title of his book, "a behavioral economist's defense of the market", summarizes his approach accurately, because it contains implicit criticism to other behavioral economists, who commit to paternalistic measures, like TS do with LP. His book is clustered around four points. The first point is to convince economists that their current model of policy recommendations is unrealistic, because they often offer their advice as if they were writing to a benevolent autocrat. Nowhere in the process of recommending policies do economists consider the individuals as respected citizens or as the directors of their life. Instead, they should aim to address their policies to individuals themselves. This point will be briefly touched upon in Section 9 of this thesis. Sugden's second strand is related to deciding if and how context-dependent preferences can be transformed to *true* or latent preferences. Besides this book, Sugden wrote frequently on this subject, and he has a fervent argument against preference purification (Infante et al., 2016A, 2016B; Sugden, 2009, 2017, 2018A). This section is dedicated to this topic. The third point Sugden makes is related to the 'opportunity criterion' as a criterion for normative economics. The opportunity criterion is "a coherent approach to the problem of reconciling normative and behavioral economics (Sugden, 2018B, p. x)." This criterion will be discussed in Section 5. Sugden's last strand is concerned with the

moral status of market relationships. These topics embody altruism, team reasoning, and the ethics of markets. This last point is beyond the scope of this thesis and will not be discussed. Besides this book, Sugden has worked extensively on other projects, especially in the world of experimental economics. Furthermore, he wrote several articles on Hume of which two will be discussed in this thesis (Sugden, 2006, 2021). In this thesis, I will contrast Sugden's (2017) work to identify self-control problems with his work on the opportunity criterion and Hume. Section 4 is dedicated to further analyze Sugden's critique on latent preferences and the role of self-control problems.

4.2. Jane in the Cafeteria

The start of the debate brings us back to the cafeteria again, which is visited by a customer called Jane (Sugden, 2018B, pp. 46-48). Jane likes to buy afternoon snacks in the restaurant, and she has two options: a cream cake or an apple. An additional fact is that Jane is overweight and should, as most doctors would agree, lose weight to live a healthier life. Nevertheless, Jane likes to eat the cream cake in the current scenario. The cafeteria owner, a fervent fan of the book *Nudge*, sees Jane eating the cake every day, and believes that it is in the interest of Jane to eat healthier. Therefore, the owner puts the cake in the back and prominently displays a basket of apples at the beginning of the counter, which is the optimal place for selling cafeteria goods. Imagine that, for whatever reason, Jane takes the apple instead of the cake now every day. Can we say that the cafeteria owner interpreted the book *Nudge* right, and he made Jane better off, AJBT?

ILS (2016A, 2016B) argue that the only way to find what made Jane better off is to find what she *truly* wants. To find what Jane *truly* wants, one needs to reconstruct her *true* preference, which would be her *true* preference if she was not prone to reasoning imperfections. These *true* preferences are also known as latent, underlying, laundered, or purified preferences, and the reconstruction process is called preference purification or preference laundering. The task of the cafeteria owner is to show that Jane's behavior is subject to reasoning imperfections that led to the mistake of eating cream cake. To conclude that she is better off, the cafeteria owner needs to prove that Jane truly wants to eat apples (i.e., she has a context-independent preference for apples). First, ILS's reply that the cafeteria owner cannot help Jane, AJBT, will be further discussed in subsection 4.3. Sugden's solution to help the cafeteria owner through self-acknowledged self-control problems is discussed in subsection 4.4.

4.3. ILS Critique of the Inner Rational-Agent

The subsection is dedicated to ILS's answer to the question in Section 4.2: the cafeteria owner cannot know if Jane is better off, AJBT. ILS contend that behavioral welfare economists implicitly assume the existence of an inner rational-agent. This model of the inner rational-agent is used if four criteria are satisfied, which will be specified in the upcoming paragraph. The main problem with the rational-agent is that this agency is unrealistic, and that there is neither philosophical nor psychological foundation for this agency. If this inner rational-agent assumption turns out to be necessary, it will crack the foundations of the LP. Namely, if no latent preferences exist, the criterion AJBT cannot be used. Following the argumentation structure of ILS (2016B), the example of Jane in the cafeteria will be reexamined.

ILS argue that the cafeteria owner cannot say that Jane is better off, AJBT, without invoking an inner rational-agent model. This model holds if it satisfies four properties (2016B, p. 33). (1) The approach to declare that Jane is better off, is based on a violation of the context-dependency property. This property means that some contextual clues, irrelevant for one's interest or well-being, influence the

choice of an agent. In the current example, Jane's preference for cream cake or apples depends on an irrelevant contextual clue, namely, the place of both objects in the cafeteria. (2) Latent or *true* preferences are taken as the relevant normative criterion, meaning that the cafeteria owner should adhere to Jane's latent preferences. These latent preferences are the preferences that Jane would have without acting on incomplete information, limited cognitive abilities, and a lack of self-control; or put differently, a reasoning imperfection occurs if humans do not act as econs. (3) Latent preferences are considered as "individuals' *subjective judgments* about their interests or well-being (Idem)", and they do not encompass any objective statement like how many apples one has. This thesis has already discussed this concept of preference[2] in Section 3.1.1. (4) "These latent preferences are assumed to be context-independent (Idem)." That means Jane should have a *true* preference for cake or apples, regardless of the irrelevant contextual clues in the cafeteria.

The mutual satisfaction of these four properties is what ILS call the inner rational-agent model. This model, used by behavioral welfare economists, regards the decision process as a black box (Infante et al., 2016A, p. 10). TS and other behavioral economists can only deduce that people made a mistake relative to the econ by assuming people have an inner rational-agent. However, the inner rational-agent fails to have rational preferences due to some reasoning imperfections, which can be explained by psychological mechanisms, like the difference in salience. ILS call these imperfections the outer irrational-shell of the inner rational-agent. BE is not concerned with the right theoretical justification for the use of latent preferences or latent reasoning capacities. Let us examine an example of Bleichrodt, Pinto-Prades, and Wakker (2001) to illustrate the critique.

For instance, Bleichrodt et al. (2001) give the example of a puzzled doctor who is unsure if an unconscious patient with inconsistent preferences benefits from treatment, AJBT. Bleichrodt et al. adjust the inconsistent preferences by transforming the preferences using the prospect theory of Tversky and Kahneman (1992). The probability weighting function and loss-aversion parameter of the prospect theory are used to convert the preferences to fit expected utility. Bleichrodt et al. are normatively committed to this theory. Subsequently, the expected utility adjusted values are considered as latent preferences of the unconscious patient. Sugden writes that "What Bleichrodt et al.'s purification methodology reveals is that, *relative to the benchmark of expected utility theory*, the person who makes … choices has behaved *as if* he held false beliefs about probabilities of the relevant events (ILS, 2016B, p. 20, emphasis in original)". The crux is that Bleichrodt et al. assume that the unconscious patient has failed to accurately reason according to expected utility theory, and followed the prospect theory instead. However, we only know that the outcome of the decisions approaches expected utility theory and closely approximates prospect theory, but there is no information on the reasoning process behind the patient's thinking. There is neither a psychological theory supporting the actual existence of latent context-independent preferences nor a correct latent reasoning method that enables us to form these preferences.

Let us now review an example in which the subject is awake. Another possible strategy to identify latent preferences, advocated by Bleichrodt et al. (2001) as well as Li, Li, and Wakker (2014), is to design a bias-free decision environment, in which an interviewer confronts the subject with its inconsistent choices. As an interpretation of this solution to Jane, imagine that an interviewer puts the

[2] Rational preferences are here defined as complete, transitive and context-independent total comparative evaluations (ILS, 2016A, p. 14).

cream cake and the apple on the same plate. If these items are put next to each other, there are no saliently distorting effects, and the interviewer can mitigate Jane's answers into integrated preferences and interpret what she *truly* prefers. However, ILS are discontented with the solution of a bias-free environment. First of all, many elicitation methods do not yield integrated preferences.[3] A bias-free environment infers that, first, a distribution of attention between several objects should be made, after which a correct or false label has to be put on every choice following that distribution. However, ILS suggest that behavioral welfare economics failed to come up with any operational definition of an error, which results in there being no information on which array of attention is correct or false. For the sake of their argumentation, we can imagine that some behavioral economists could arbitrarily define a (un)biased environment as follows: If two objects are within one centimeter of each other, it is a bias-free environment. If two objects are more than one meter away from each other, the decision is biased. With this imagined arbitrary definition, TS can say that the preference for cream cake results from a reasoning error due to a difference in salience. In principle, that is also what the econometric salience model of Bordalo, Gennaioli, and Shleifer (2013) does, it creates a bias-free environment after correcting for salient differences by implying a linear utility function. However, this model fails to show how there is any "latent [reasoning] capacity to work through this process [weighing various distances with expected utility] correctly (Infante et al., 2016B, p. 36)." In the perspective of ILS, there is no psychological theory that explains why there is any reasoning capacity that is superior in the one-centimeter frame in comparison to the one-meter frame. Returning to the apple and cream cake, at which distance would the true reasoning occur and when are the reasoning faculties distorted due to a particular distance? Subsequently, what is the theoretical justification for using expected utility theory or prospect theory to normatively purify these preferences? ILS's answers are: there is no correct distance, and BE provides neither philosophical nor psychological justification.

The explored path of ILS is different. They welcome non-rational theories for normative purposes, like Sugden's opportunity criterion discussed in Section 5 of this thesis. Furthermore, ILS claim that that context-(in)dependent preferences are not per definition related to (in)correct reasoning. For context-dependent preferences, like Jane's for apples and cream cakes, there is no justification given for why these non-integrated preferences are a mistake. Even with sound reasoning, people could fail to order their preferences context-independently. For instance, Jane is more spontaneous on Fridays, her cheat day, and that made her opt for a cream cake. Analogously, a context-dependent choice is not per definition an error. The context-independent smoker, Norman, could still reflect on his couch that he needs to stop smoking to improve his health (Sugden, 2018B, pp. 87-88). Albeit, in the supermarket, he cannot resist the temptation of buying cigarettes and smoking a pack in one day. When he sits on the couch again, Norman, reflectively thinking about his future, does not endorse his decisions. Thus, context-(in)dependent choices do not say that a reasoning error has been made.

To strengthen this example, ILS discuss an example of Joe in the cafeteria, and he is akin to Jane. In a nutshell, Joe in the cafeteria took the cafeteria's cake and is happy with his choice while he is still in the cafeteria. However, when Joe looks at his waistline in the mirror at home, he regrets the decision of eating the cake. ILS assert that Joe, equally, would regret the decision if he had not taken enough money with him to buy the cake in the cafeteria. Thus, "Joe-in-front-of-the-mirror weights the relevant

[3] Choosing the correct elicitation method is another problem that ILS (2016A, 2016B) point at. With some methods, preference rehearsal occurs often. This problem is only supportive and not necessary in their theoretical disapproval of latent preferences.

aspects of well-being in a way that supports a preference for fruit; Joe-in-the-cafeteria weights them in a way that supports a preference for cake (ILS, 2016B, p. 37)." Even if the cafeteria owner would see a disappointed Joe in front of the mirror, the owner cannot assert that Joe has self-control problems. Joe's decisions depend on the context, and no context can be labelled as correct or incorrect. Thus, the cafeteria owner cannot deduce any latent preferences of Joe.

The conclusion of ILS is even stronger. Any external observer, like a doctor, a cafeteria owner, a choice architect, or a behavioral welfare economist does not have a theoretically correct method to construct latent preferences, which, subsequently, can be used for normative analysis. As ILS argued, one needs to illegitimately assume the existence of an inner rational-agent that makes reasoning imperfections due to a psychological shell. Instead, ILS are open to the idea that non-integrated preferences can also be realistic. Thus, the cafeteria owner cannot determine if Jane can be nudged, AJBT. However, Sugden offers an exemption how Jane can be made better off, which will be presented in Section 4.4.

4.4. Sugden's Argumentation on Self-Control

In this subsection, Sugden's paper (2017) with the title *Do people really want to be nudged towards healthy lifestyles?* is examined. In this paper, as well as in other places (2018A, 2018B, Section 4.8), Sugden proposes a viable method for the cafeteria owner to find Jane's true preferences through self-acknowledged self-control problems. Albeit self-control problems are uncommonly endorsed (Sugden, 2017, pp. 121-122), it is in principle possible.

Referring to self-control problems, Sugden writes that "latent preferences are not a hypothetical construct (2017, p. 118)", but:

> They are the preferences that the individual *actually endorses* in some independently definable circumstances (perhaps in a cool emotional state, reflecting about her welfare), and which in some sense she continues to acknowledge even when she fails to act on them. In other words, Thaler and Sunstein may be using a model of self-acknowledged akrasia (that is, failure of self-control) (Idem, emphasis in original).

To rephrase the former quotation, Sugden is not committed to latent preferences that are constructed by outsiders; he is committed to latent preferences if the individual endorses or self-acknowledges that she fails to act on these latent preferences that are constructed in *independently definable circumstances*. Notice that for self-control problems, one needs to feel two inclinations simultaneously. This method is, effectively, a meta-desire strategy from Frankfurt (1971). To illustrate the point further, let us review two examples that Sugden's (2017) paper examines from TS.

The first story concerns a bowl of cashew nuts. This story takes place during a party hosted by Thaler where his guests were eating cashew nuts right before dinner. Thaler, worrying that no one would have any appetite left for dinner, put the nuts out of their sight in the kitchen. Afterward, his guests were happy that he put the bowl away and, the guests, unanimously, self-acknowledged that they were suffering from the temptation of eating the nuts and told Thaler they were glad that he put the bowl away. Sugden agrees that this nudge, moving the nut bowl to the kitchen, made the people better off, AJBT, because:

> On the most natural interpretation, the guests were aware that the sight of the nuts induced an urge to eat they found hard to resist but which, even *at the moment of eating*, they felt a desire not to act on. As soon as the visual stimulus was removed, they endorsed that desire as their true preference (Sugden, 2017, p. 118, emphasis added).

The second example goes as follows:

> It is easy to imagine a Tom who, *even while accepting a share of the second bottle and choosing a crême brulée*, acknowledges that he is acting against his better judgment. Such a Tom might reasonably be said to want to be nudged towards keeping his resolutions—at least if, when accepting the dinner invitation, he was not 100% confident of his willpower (Idem, emphasis added).

From both examples, one could derive the rule that Sugden describes. In order to classify a feeling as a self-control problem, the guests self-acknowledged the temptation "at the moment of eating" and Tom self-acknowledged the failure to act "even while accepting [the desert]". The actors felt two different inclinations. Only one inclination, derived from some *independently definable circumstances*, perhaps generated in some reflective method in a cold state, can make someone better off, AJBT.

I would argue that Sugden's solution is the same strategy as Frankfurt's meta-desires (1971). Following a Frankfurtian language, the explanation is as follows. Thaler's guests had two inclinations: to eat cashew nuts or to resist them, and the former inclination was stronger than the latter. The guests had the first-order desire, or an *effective* desire, to eat cashew nuts when the bowl was on the table. After the bowl was moved out of sight by Thaler, the guests thanked him and were happy, after which they endorsed their second-order desire. The guests wanted to eat cashew nuts, but they wanted to want to resist the temptation, but they failed to act upon this second-order desire. During some *independently definable circumstances*, probably when the guests' parents taught them not to snack before dinner, the guests failed to act on the second-order desire. Right after the bowl was replaced, the guests endorsed that they had a second-order desire of not snacking. Again, in Frankfurtian language, Thaler's guests are self-acknowledged wantons. They lacked the second-order volition to act and could not prevent themselves from acting on their first-order desires. The example of Tom can be explained according to a similar Frankfurtian scheme. Tom had a first-order desire to eat an unhealthy dessert. However, while accepting a share of the second bottle and choosing the desert, he acknowledged a second-order desire of dieting. Thus, Tom wanted to eat a dessert, but Tom wanted to want to keep his resolutions. Furthermore, even the concept of willpower fits well in a Frankfurtian scheme. A lack of confidence in his willpower impaired his second-order volition. Willpower can be defined as the ability to turn second-order desires into second-order volitions.

ILS's example of Joe fits well in this meta-desire theory. Joe does not feel any second-order desires *at the moment of choosing*. Joe solely experiences first-order desires and no second-order desires in the cafeteria. Joe only regrets his decision in front of the mirror, but a second-order desire in retrospect (i.e., in front of the mirror) cannot be classified as a self-control problem. ILS assert that this second-order desire comes due to different weighing processes in different situations, but no latent preferences can be constructed with retrospective second-order desires. Thus, retrospective second-order desires are irrelevant for self-control problems.

Coming back to the question posed in Section 4.2, how can the cafeteria owner establish that Jane suffers from self-control problems? Following Sugden (2017), the cafeteria owner needs to ask Jane

if she experienced the feeling of temptation during the process of buying the cake. Moreover, her preference for buying the cake should also conflict with some other preference that is endorsed in some *independently definable circumstances*. In other words, the cafeteria owner can ask Jane if she has experienced a second-order desire of not eating the cake when she ordered it.

So far, we established that self-acknowledged self-control problems form a legitimate reason for Sugden to nudge people, AJBT. For self-acknowledged self-control problems, one needs two conditions. (1) Two conflicting inclinations and (2) of which one inclination is developed in a reflective moment in some *independently definable circumstances*. I argued that Sugden implicitly repostulates a meta-desire theory. Both Thaler's guests and Tom felt two conflicting inclinations, of which the normatively higher-ranked desire is formed in some reflective state. In subsection 4.5, an example of Sugden's book will be investigated.

4.5. Combining ILS and Sugden (2017)

So far, we have examined ILS's general objections against preference purification and Sugden's approach to recognizing self-control problems. At this point, I argue that Sugden fails to incorporate his approach (2017) to an example of Jane in *The Community of Advantage* (2018B, pp. 81-82).

In Section 4.3, the conclusion was drawn that an outsider cannot know what the latent preferences are from other people. One method to purify the preferences was through self-control problems. Effectively, ILS claim that an external expert does not have the theoretical tools to judge whether cafeteria-Jane suffers from self-control problems. Jane could, just like Joe, weigh the different attributes of particular decisions differently on different occasions. Following Sugden (2017), only individuals can self-acknowledge their self-control problems.

Interestingly Sugden applies his criticism uncarefully in his only example in the Section "Akrasia" in *The Community of Advantage* (Sugden, 2018B, pp. 79-82). Everyone knows the vows of friends in which they pledge to study more, work harder, eat healthier, jog further, etc. after New Year's Eve. Imagine that Jane makes the resolution to stop drinking. She has intrinsic motivation to stop because she wants to lose weight. Three weeks later, she goes to a restaurant to celebrate a friend's birthday and drinks a few glasses of wine. Is this the self-control problem that TS are talking about? Sugden maintains that there is no reasoning error in Jane's alcohol intake. New Year's Eve is associated with creating new habits, and the atmosphere of a restaurant is associated with spontaneity and taste. Sugden says: "This is not a self-control problem; it is a change of mind (2018B, p. 82)." To explain Sugden's way of thinking, he cites a maxim of Kahneman on another place that captures the essence well: "Nothing in life is as important as you think it is when you are thinking about it (Sugden, 2018B, p. 153; Kahneman, 2011, p. 402)." The impossible task of the choice architect is to show that, on one moment, Jane has *true* context-independent preferences (i.e., at New Year's Eve), and the other moment a context-dependent preference (i.e., in the restaurant) due to reasoning imperfections. ILS (2016A, 2016B) and Sugden (2018B) find that this mission is doomed to be impossible, because they deem that the step of identifying an error mechanism, the lack of self-control, implicitly relies on an inner rational agent, who is hassled by the psychological shell. Instead, Sugden's most logical explanation is that Jane changed her mind. However, strictly speaking, Sugden (2018B) cannot know if she changed her mind or if she suffers from self-control problems. There is no method for external observers to judge if she changed her mind or if she is suffering from self-control problems. Following the method of Sugden (2017), Sugden needs to provide Jane's self-acknowledged feeling of resisting temptation. Without any

introspection of Jane, it is impossible to declare that her personality changed. One could object to this reasoning by stating that Sugden did a mere guess that she changed her mind. However, this reasoning requires that self-control problems are infrequently endorsed, and that line of reasoning will be explored in Section 7.

4.6. Summary of Section 4

Up to this point, TS plea for LP has been examined and they established that self-control issues were a legitimate reason to nudge, AJBT. ILS concluded that outsiders, including behavioral welfare economists, cannot assess what the latent preferences are of other people. Sugden agreed that self-control issues were a valid criterion to make people better off, AJBT. I argued that Sugden's method to identify self-control problems is identical to the meta-desire theory of Frankfurt (1971). Furthermore, I opted for a small correction in Sugden's book. He concluded that Jane's personality has changed, but it is impossible to know if she did without listening to herself.

In the coming two sections, I will argue that Sugden's analysis of self-control cannot be reconciled with another role it plays in other parts of his works. In Section 5, the relation between the opportunity criterion and self-control problems will be investigated, and Section 6 will be devoted to the relation between Sugden's interpretation of Hume and self-control problems. In both sections, I will argue that the existence of meta-desires cannot be combined with both theories.

5. The Continuing Self and Self-Control Problems

In this section, we will review the relationship between the continuing self and self-control problems. The continuing self is an agent that plays a crucial role in the opportunity criterion, and this criterion is a normative approach that respects individuals' choices without invoking preferences. It states that more opportunities, generated in a broad market economy, are better than fewer opportunities. After a summary of this opportunity criterion in Section 5.1, the implicit view of the self, the continuing person, is explored in Section 5.2. Both the existence of self-control problems and the existence of meta-desires conflict with the continuing person, as I will argue in Section 5.3. Section 5.4 concludes.

5.1. The Opportunity Criterion

The opportunity criterion is rooted in Sugden's contractarian background. Sugden proposes a new line of normative economic thought that is compatible with neoclassical integrated preferences, but that does not require these assumptions to be true. Behavioral economics has shown that neoclassical assumptions are often violated, which is the reason why a new normative criterion cannot start from the satisfaction of preferences. Following his contractarian background, this criterion should be specified to respect the individual's sovereignty, and since it concerns economics, the criterion should respect consumer sovereignty. Sugden's suggestion for this criterion is opportunity because opportunity satisfies several desirable properties, like being operational, transparent, objective, and in general respectful to applications (Sugden, 2018B, pp. 83-85). The simplest explanation of the opportunity criterion is that individuals collectively benefit if there are a multitude of opportunities to undertake any voluntary transaction they desire. Sugden recognizes that the size and richness of opportunity sets are valuable in themselves, and are built on other philosophical traditions, like Sen and Arrow (Sugden, 2004, p. 1016). Subsequently, Sugden argues that competitive markets provide ample opportunities that are beneficial, even for agents with non-integrated preferences.

5.2. The Continuing Agent

Departing from the idea that an agent does not have to satisfy neoclassical preferences, Sugden has set himself up for the task of portraying a realistic individual who makes non-integrated choices at different moments in different contexts. Perhaps some agents act like those normative agents, but that theory would not be suitable for all. Sugden introduces the continuing agent who can make non-integrated choices over time. The special quality that Sugden assigns to this agent is her continuing identity, which entails that she agrees with all her choices, ranging from past, present to future. Sugden formally demonstrates that, given this agent, more opportunities are beneficial. I want to bracket the formalities and focus on these specifications of this continuing agent.

As an example, Jane who lives in four time periods is taken (Sugden, 2018B, pp. 102-106). She is endowed with money at $t = 0$, and buys a ticket at $t = 1$ for a concert that takes place at $t = 3$. At $t = 2$, Jane has the possibility to sell her ticket at the current market price, but the current market price is lower than the original ticket price. Jane regrets the decision of buying a ticket and decides to sell the ticket at $t = 2$. The subsequent question is how economists theorists should interpret this story. These theorists assume a discontinued identity, in which the self at $t = 2$ is distinct from the self at $t = 1$. These theorists use an internal decreasing impatience function (i.e., the dual-self) or appoint a superior desire (i.e., the meta-ranking) in order to philosophically glue the broken identity of an agent. Sugden starts from a different point than some theorists. Sugden finds both strategies are unattractive for people that choose according to different principles than rational choice theory (2018B, p. 105). Instead, he proposes a "radically" new method, taking the opposite stance of these theorists; Sugden assumes that the identity of agents is coherent. I will argue later that Sugden's rejection of meta-ranking is incoherent, but for now, let us continue with Sugden's argumentation. As an example, he takes Jane:

> We should think of the continuing Jane – let us call her Jane* - as the *composition* of the selves which perform the various parts of whatever sequence of actions performed (Idem).

If Jane reviewed her choice to buy and sell the ticket at $t = 3$, she realizes that she lost money compared to the situation where she would not have bought the ticket. Sugden continues that:

> Jane can concede this, yet still see both buying and selling as *her* autonomously chosen actions: she wanted to buy, and she bought; she wanted to sell, and she sold. She does not have to disown either of those actions as the work of an alien self, or as the result of weakness of will (Idem).

Sugden proposes that:

> The continuing agency of a person across time and across contexts should be understood as the continuing existence of a self-acknowledged *locus of responsibility*. The intuitive idea is that a person is a continuing locus of responsibility – for short, a responsible agent – to the extent that, at each moment, she identifies with her own actions, past, present, and future (2018B, p. 106, emphasis in original).

There are three consequences of perceiving Jane as a self-acknowledged *locus of responsibility* in contrast to the dual-agency and meta-rankings. This method effectively lays off the behavior interpreting economist, because there is no interpretation needed of Jane's behavior. Jane could have acted upon

different principles than rational choice theory, which makes a theorist's interpretation go astray. Now, decisions are a self-acknowledged *locus of responsibility*, and therefore, there is no need to further open the black box of decision-making. In other words, assuming this level of responsibility shuts the door for behavioral welfare economists who feel the need to interpret non-integrated preferences as an incongruent identity. Secondly, the self-acknowledged *locus of responsibility* fits perfectly in Sugden's contractarianism and liberalism. A contractarian wants to respect the individual's decision-making, and by asserting that an individual is a *locus of responsibility*, decision-makers have to respect what these responsible agents do. Moreover, it fits in a liberal tradition. Sugden associates some parts of this theory with John Stuart Mill (Sugden, 2010, pp. 48-49, 58-60), and it is not hard to recognize Mill's optimism of the individual in the continuing person. Mill wrote that: "while with respect to his own feelings and circumstances, the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by any one else [including behavioral welfare economists] (Mill, 1859/2001, p. 70)." If the phrase "including behavioral welfare economists" is included, Mill's famous quotation summarizes Sugden's philosophy well. The third consequence is that Sugden can promote markets with the opportunity criterion. Opportunities that restrict the future self (i.e., self-constraint) have a neutral value in Sugden's opportunity criterion (2004), because they form an additional opportunity (to restrict), but simultaneously, they restrict future opportunity.

5.3. Compatibility of the Continuing Person and Self-Control Problems

In this subsection, we will review to what extent the continuing person is compatible with self-control problems. Let us recite Sugden's quotes to illustrate how problematic self-control is for the continuing person. He says: Jane "does not have to disown either of those [buying or selling] actions … as the result of weakness of will (2018B, p. 105)." Assuming the reverse holds, Sugden says that an individual disowns his or her actions if he or she suffers from a weakness of will.[4] Also, Sugden writes: "a responsible agent … identifies with her own actions … (2018B, p. 106)", which implies that a lack of identification disqualifies an agent from being responsible for her actions.

The example of Jane buying a concert ticket is not suitable to demonstrate this passage, therefore, let us consider Jane in the cafeteria again. Imagine that Jane acted upon self-acknowledged self-control problems and was tempted to buy the cream cake. Whilst eating the cream cake, she immediately feels regret about the urge that led her to eat the cream cake. This urge contradicts her second-order desire to lose weight, and that desire has been formed during some *independently definable circumstances*. If Jane endorses these self-control problems according to Sugden (2017), two conclusions can be drawn. Firstly, Jane disowns her action when she eats the cake taken during a moment of self-acknowledged temptation, therefore, she fails to be a *locus of responsibility* during these self-control problems. If these self-control problems occur often, the plausibility of the continuing agent is impaired. Secondly, if Jane has self-acknowledged self-control problems, it means that some opportunities have a negative value. Sugden recognizes this issue:

> For example, in problems involving self-control, an earlier (or higher) self may approve of restrictions on the opportunities of a later (or lower) self. Similarly, a later self may wish that an earlier self had had less opportunities to act imprudently. But if an individual is understood as a continuing locus of responsibility, any increase in that individual's lifetime opportunity is good

[4] Sugden does not define a weakness of will. However, having a weakness of will in this context fits well as impairing a second-order volition, which is in line with the argumentation in Section 4.4 of this thesis.

> for her in an unambiguous sense. The larger her opportunity set is, the more she - construed as responsible agent with a continuing existence through time - is free to do. (Sugden, 2004, p. 1018).

Here, Sugden contradicts self-control problems explicitly with the continuing agent. A theorist following Sugden's responsible agent cannot say that responsible Jane suffers from self-control problems. However, a meta-ranking theorist can say that the option of cream cake has a negative value. The option of cream cake at the cafeteria has a negative value for *ex post* Jane, and *ex post* Jane wished that *ex ante* Jane would have acted prudently. Extending this line of argumentation to a societal level, opportunities in areas where many people suffer from self-control problems, have a *negative* value overall. This line of argument will be explored further in Section 9.

Sugden, recognizing the incompatibility of self-control problems and the opportunity criterion, found two ways to avoid the two conclusions. The first strategy is to assert that people uncommonly endorse their self-control problems. Given this infrequent occurrence, the implausibility of the continuing agent or the existence of negative value options is not threatening his philosophy. The validity of this argumentation will be explored in Section 7 of this thesis. Secondly, Sugden is not convinced that market opportunities can be restricted. Sugden writes:

> This prompts the question of whether markets tend to provide individuals with opportunities to constrain their later choices, if and when they want to do so. A rough and ready answer is that, in a competitive market, self-constraint technologies tend to be made available to those people who are willing to pay for them, but so too are the counter-technologies that allow people to escape from constraints they no longer wish to be bound by (Sugden, 2018B, p. 149).

This counter-argument is hard to refute. For supermarket products, like alcohol and cigarettes, Sugden's argumentation is right to claim that individual self-regulation is difficult to achieve. An individual can drive to another city to anonymously buy these goods. However, the government can impose restrictions on the number of selling points. I will spend more time on this argumentation in Section 10.

5.4. Summary of Section 5

In the previous subsections, the responsible or continuing person was presented. The continuing person cannot experience self-control problems, because the *locus of responsibility* is what an agent is missing during self-control problems. Furthermore, the opportunity criterion fundamentally excludes the consequences of self-control problems. Opportunities that restrict the future self have a neutral value in Sugden's opportunity criterion (2004), but these self-constraining options have a positive value in the case of self-control problems. Instead of facing these incongruencies, Sugden maintains that self-control problems occur infrequently (2017, pp. 121-122), and adds that self-regulating counter-measures are readily available in the market. The conclusion that self-control problems are uncommonly endorsed will be explored in Section 7, and a word on the self-constraining counter-measures will be discussed in Sections 9 and 10. Despite these reservations, I conclude that the existence of self-control problems is, in principle, irreconcilable with the responsible self. Section 6 of this thesis will be dedicated to investigating the philosophical relation between David Hume and self-control problems. Sugden is inspired by Hume's psychology and philosophy, and Hume's decision-theory could shine light on self-control problems.

6. Hume and Self-control

6.1. Introduction to Hume

The philosophy of David Hume should require no introduction, but, in a nutshell, he was a major Enlightenment philosopher. He is well-known for his empiricism, skepticism, and naturalism, and was active in many scientific branches, including economics and philosophy. For this thesis, his work on the self will be reviewed and I shall look beyond his famous quote: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them (Hume, 1740/2009, p. 636)."

Before investigating the link between Hume and self-control, Sugden's link with Hume is explored in Section 6.2. I will argue that Hume cannot support a meta-desire theory of Frankfurt, and would disagree with Sugden's solution with regard to some *independently definable circumstances* in Section 6.5. Before doing so, Sugden's interpretation of Hume will be described in Section 6.3, and Hume's stance on identity will be described in Section 6.4. A recapitulation of Section 4 to 6 is made in Section 6.6.

6.2. Sugden's Link to Hume

As a philosopher that criticizes the use of preferences in normative economics, Sugden searches for a new manner to understand how decisions are made. In *The Community of Advantage*, he reviews the example of Jane through the lens of a decision theory based on the psychology of attention and desire. As Jane has often been to the cafeteria, the information she has about the two products is equal: she knows what she will get. The real reason for her choice is the attention she gives to the two items. "As experienced by Jane, this inclination may be simply a *feeling*, not a *proposition* to which she assents. This might be all there is (Sugden, 2018B, p. 73)." Although Sugden does not mention any psychologist or philosopher to support this example, I interpret that Sugden implicitly proposes that rational choice theory should be replaced with a decision theory inspired by Hume.

Sugden often refers in his work to Hume, but it remains unclear how prominent the role of Hume is in Sugden's philosophy. Nevertheless, I interpret that Sugden's decision theory is, at least, strongly influenced by Hume. He admits on several occasions that he finds Hume's theory of decision-making plausible (2018, p. 96, 2020, p. 69), even saying: "I conclude that Hume's theory of mind is consistent and psychologically well-grounded (2021, p. 836)." Lastly, he writes that "Hume's [decision] theory … [is] compatible with current developments in experimental psychology and behavioral economics (2006, p. 365)." I choose not to further investigate the role of other philosophers, like his mentor Buchanan, since other authors have already explored the role of Buchanan in welfare analysis in more detail (i.e., Dold, 2018).

I will continue to explore Sugden's interpretation of Hume's rationality. If Hume cannot commit to rational preferences or latent preferences, self-control problems cannot be identified. Self-control problems are irrational for rational choice theorists because self-control problems violate the stability or context-independency property. However, if Hume also declares that the underlying desires cannot be classified as right or wrong, economic theorists cannot identify self-control problems.

6.3. Sugden's Interpretation of Hume's Rationality

In this subsection, Sugden's stance on Hume's decision theory will be explored based on two of Sugden's papers (2006, 2021). In Sugden's former paper, he presents two aims. The first aim is to

show that, in Hume's theory, rationality cannot be assigned to any actions. Secondly, some authors perceive the disconnection between rationality and actions as a mistake of Hume, which it is not. Sugden argues that Hume's philosophy is coherent without a connection between rationality and actions, and that other economic philosophers have trouble reading Hume due to their presupposed validity of rational choice theory. One of those philosophers is Dreier, who proposed a thought experiment where a person has conflicting desires about sun-bathing (1996). A tan looks great but increases the chance of skin cancer. Dreier explores a person with the following three preferences:

> There would be real trouble if all things considered I preferred staying out of the sun to basking in it, and basking in it to a short exposure, and a short exposure to staying out of the sun. If Humeanism is committed to saying that the combination of those three preferences is perfectly rational[5], then Humeanism is certainly not worth defending (Dreier, 1996, p. 250; Sugden, 2006, p. 370).

However, Sugden argues that Hume's decision theory is worth defending, because the concepts of desire and volition are representative of our decision theory. He writes:

> Desire is an 'emotion of propensity' which 'unites us to' the idea of some object. That is, it is a passion which focuses on the idea of some object and induces us to approach, possess or consume it. Volition is the felt experience of intentional action, 'the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind' (Sugden, 2006, p. 382).

The association of ideas governs these desires. Sugden concludes that "In Hume's theory, when two mental items are associated with one another, consciousness of one tends to increase the vivacity of the other (2006, p. 284)." Considering the example of the cake and apples in a cafeteria, Hume would find it perfectly logical to get more appetite if both meals are beautifully presented. Sugden argues that Hume, hereby, successfully recognizes and explains the existence of context-dependent behavior. Coming back to the tanning example, Hume would not classify the mutual existence of these three preferences as an error. Instead, Hume's theory of mind is dynamic. One temporary mental state may cause the temporary existence of the subsequent mental state, and, therefore, it is not strange that this person has conflicting desires at several moments. Thus, our desires can be called inconsistent, thereby failing to comply with rational choice theory. Set theory usually demands, at least, an acyclic preference ordering. This acyclic preference ordering is a false representation of human decision making because humans often act on inconsistent desires, and these desires can neither be called true nor false.

Moreover, a latent preference in Hume's philosophy does not exist, and that is what Sugden's (2021) demonstrates in the second paper. Although Sugden provides a much deeper understanding of Hume throughout his paper, there is a superficial shortcut that provides Sugden's interpretation already:

> If preferences are comparative desires, they are original facts and realities, incapable of being pronounced true or false. Having arrived at empirical explanations of people's actual feelings of desire, there seems to be nothing left for a Humean theorist to do: there is no space in the theory for a concept of error (Sugden, 2021, p. 840).

[5] This example would violate the axiom completeness.

To put theory into practice, Sugden offers an example akin to Jane. Imagine that Jane usually prefers apples as her afternoon snack in the cafeteria. However, on an afternoon, the smell of the cream cake may activate her memory with some sweetness associations, also called momentary appearances by Hume, and those appearances lured her to choose the cream cake. Is this a mistake, Sugden asks? No! "The existence of context-independent standards of correctness in judgments about desirability does not imply that there are corresponding standards of correctness in desires (Sugden, 2021, p. 843)." In a similar fashion, Sugden discusses time-preference reversals, which Hume already explored in his philosophy. Hume would interpret this phenomenon as "a change of will, not a failure of will (2021, p. 846)." Hence, no error label can be applied to that phenomenon.

Given the last two examples, it is straightforward to conclude that Hume, in Sugden's representation, cannot categorize an action as correct or incorrect. Therefore, rational choice theorists cannot identify any self-control problems, because a self-control problem requires an error label. Hume's self is dynamic, ever-changing, and his theory of action depends on contextual clues, on changing desires, or a change of will. In all these cases, there is no foundation of latent preferences to establish self-control problems. This line of reasoning will be further explored in subsection 6.4.

6.4. Hume on the Continuing Self

In this subsection, I would like to argue that Sugden's self-control problem identification method is unattainable for Hume. Hume's original work on identity, as discussed in *Treatise*, will be used here. Before continuing to what Hume wrote on personal identity, It is important to acknowledge that Hume's opinion was never fully formed. He perceived his former work as a "labyrinth", but it remains a guess as to what aspects Hume finds to be a labyrinth. Garrett provides several details on why Hume calls it a labyrinth (2016, pp. 237-242), but none of those reasons are crucial to discuss. If these reservations are kept in mind, a fruitful analysis on the relation of Hume and self-control problems is possible.

In the book *Treatise*, Hume goes to show that identity is a fiction. The argumentation is simple. According to Hume, identity is the idea of an object that "remains invariable and uninterrupted thro (1740/2009, 397; Noonan, 1999, p. 191)." Humans, as well as plants and non-human animals, change dynamically and continuously, and therefore, attributing identity to them is fiction and relies on a mysterious unattainable notion of the self. This line of reasoning also comes back in a famous passage of Hume:

> For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception (1740/2009, p. 395; Noonan, 1999, p. 193).

In Hume's work, there self is continuously influenced by perceptions. The idea of identity comes from the faculty imagination, which associates uninterrupted progress of thought and successions of resembling thought with identity and memory. As commentator Noonan writes:

> Hume is able to regard memory not merely as providing us with access to our past selves, but also as contributing to the bundles of perceptions which we can survey, elements which represent and thus resemble earlier elements; and so—since resemblance is a relation which

> enables the mind to slide smoothly along a succession of perceptions—as strengthening our propensity to believe in the fiction of *a continuing self* (Noonan, 1999, p. 203, emphasis added).

To recapitulate, Hume would argue that there is no continuing self with regard to identity because the bundle of desires is in a constant flux interacting with its environment. Noonan elaborates, among other critiques, that Hume walks in a philosophical pitfall. An identity does not require to be stable, but a person or object can be identified by its constant change. For instance, a person could be defined by an ever changing-self. Keeping that critique in mind, we can reassess Sugden's continuing self. I find that Sugden's term *continuing self* fails to grasp the core of Hume's thought, namely, it fails to capture the changing part. If Sugden would be solely inspired by Hume, he could change this term to the *continuously-changing self*. This term corresponds better with the essence of Sugden's philosophy and helps to explain the incoherency between Hume and self-control problems.

6.5. Hume's Philosophy and Self-Control Problems

At this point, it is time to draw a conclusion regarding the relationship between Hume's philosophy and self-control problems. In Section 6.3, Sugden shows that there is no preference purification possible with Hume's philosophy. No desire can be labeled as correct or wrong, and behavioral welfare economists can only assert that someone's will has changed, but they cannot prove an error in behavior. Following the critique of ILS, no outsider can construct that one suffers from self-control problems. Moreover, Sugden's interpretation of Hume also has two consequences for the personal identification of self-control problems. Firstly, Hume cannot classify desires as right or wrong and, therefore, Hume's decision theory cannot be reconciled with Frankfurt's theory of desire. In Hume, there is no hierarchy in feelings, desires, inclinations, or volitions. In Frankfurt's analysis, there is a hierarchy in the first-order and second-order desires, whereas the latter is normatively important. Secondly, in Hume's part on identity, Hume deemed it impossible to decouple the self from its perception. One temporary mental state may cause the next temporary existence, and our selves are in a dynamic flux. This flux forms a problem for Sugden's identification method. Sugden proposes to base self-control problems on some reflective thoughts developed in some *independently definable circumstances*. However, in Hume, the self is continuously changing and cannot be perceived in isolation of its perception. Hume would assert that identifying some *independently definable circumstances* is in itself impossible, because the term *independently* would imply that there exists a moment in which one can observe a self in some stable isolation. Thus, Sugden's self-control problem identification method cannot be reconciled with Hume's philosophy.

6.6. Summary of Sections 4-6

In Sections 4 to 6, I have examined the coherent use of self-control in Sugden's philosophy. In Section 4, Sugden's (2017) method to identify self-control problems is discussed. These self-control problems are self-acknowledged mistakes developed during some *independently definable circumstances*, and the individual recognizes these mistakes at the moment of temptation. The goal of Sections 5 and 6 was to see if the existence of self-control problems can be reconciled with his opportunity criterion and Hume's decision theory. The relation between self-control problems and the opportunity criterion and the continuing or responsible self was examined in Section 5. The conclusion was that self-control problems are a counter-example of a responsible agent, and that self-control issues can have a negative effect on the value opportunity. The opportunity criterion does not account for these problems. Hume's decision theory, which is supported by Sugden, was reviewed in this section. I argued that it

is not possible to identify self-control problems in Hume's philosophy, because there is no correct method to find *independently definable circumstances* because the self cannot be perceived in isolation.

The conclusion drawn is, if Sugden wants to be consistent in his philosophical works, he has to admit that self-control problems cannot co-exist with the opportunity criterion and Hume's decision theory. As a result, Sugden has to reassess the merits of his approach to identify self-control problems (2017, 2018A) or his opportunity criterion and Hume's decision theory. The only 'way out' for Sugden is to downplay the size of self-control problems, and, thereby, acknowledging that his philosophy is incoherent, but only in minor respects. In Section 7, the claim that people uncommonly self-acknowledge their self-control problems will be assessed.

7. On the (In)frequency of Self-Control Problems

How often do people endorse their self-control problems? Sugden's answer is best summarized by: "My guess is that most people prefer not to explain their own behavior as the result of error (2017, p. 121)." As this section will show, his argumentation for this answer is, to say the least, minimal. In Section 7.1, Sugden's analysis on the frequency of self-control problems is reviewed. He suggests looking at the take-up rate of self-constraint measures. In Section 7.2, a different explanation on the frequency of self-control problems is given. In an empirical paper, Hofmann, Vohs and Baumeister (2012) use a definition of self-control akin to Sugden. Furthermore, Sugden's definition of self-control problems is considered to be narrow. In Section 7.3, a conclusion will be drawn that a new definition of self-control problems should be explored.

7.1. Sugden's Perceived Frequency of Self-control Problems

Sugden's analysis on how often self-control problems occur is somewhat ambiguous. In his debate with TS, Sugden conveyed the message that "self-acknowledged self-control problems are a lot less common than many behavioral economists seem to think (2017, pp. 122)." With ILS's critique discussed in Section 4.3, an external observer cannot determine if someone suffers from self-control problems. Self-control problems are a valid criterion to make someone better off if an individual self-acknowledges these self-control problems. A solution to investigate how often self-control problems occur is to ask people how often they experience self-acknowledged self-control problems. In Sugden's chapter *Why a contractarian cannot be a paternalist* (2018B, Section 3.4), Sugden discusses the scenario in which the cafeteria owner asks Jane directly why she picks the cream cake, despite being advised by doctors to take the apple. Sugden comes up with seven reasons Jane could answer in a survey, of which three will be highlighted:

> (b) When I am a few years older I will adopt a healthier diet, so my current eating habits are not a problem.
>
> (e) All my grandparents were thin but died relatively young. It is quite likely that I will die young too, whatever I eat.
>
> (g) I always go into the cafeteria having resolved not to choose cake, but when I see the cake at the front of the counter, I can't resist the temptation (Sugden, 2018B, pp. 47-48).

Reason (b) and (e), and the other four reasons that are omitted, are based on false assumptions or behavioral biases, like procrastination (b) or overweighing personal experience (e). Only reason (g) is

based on a self-control problem. Although reasons (a) to (f) are poor reasons to choose a cream cake, Sugden argues that these responses (a) to (f) are "are much more likely" than option (g) (2018B, p. 48). At this point, Sugden's claim is a guess or a guesstimation based on personal experience at best.

A subsequent step would be to verify Sugden's guess with some empirical data. For instance, Sunstein (2018) asked in a preliminary survey how many people suffered from self-control problems, whether small or large. The majority (70%) agreed or strongly agreed with this statement, which would, at least, indicate that self-control problems are endorsed frequently. Sugden objects to this method saying that "it sets the bar very low (2018A, p. 12)." Instead, Sugden proposes a different research method:

> I think a better test of the prevalence of self-acknowledged self-control problems would be to investigate the take-up of options that are explicitly presented as self-constraint mechanisms, such as 'panic buttons' that allow online gamblers to lock themselves out of their accounts for fixed periods (Idem).

He refers to a British betting firm called William Hill that installed a panic button where users could lock themselves out of their accounts for a fixed period. Only 3000 accounts per week used this button compared to the total 2.7 million users. Self-constraint is defined as "the imposition of external restrictions on one's future choices, such as locking away spirits in a cupboard and sending the key back to oneself (Schubert, 2015, p. 280)." I agree that this ratio is indeed low. In a similar way, Sugden analyzes two papers by DellaVigna and Malmendier (2004, 2006). The essence of the two papers is that companies, like fitness clubs, enrich themselves due to naïve customers with biased (mainly over-optimistic) preferences. In the latter study called "Paying not to go to the gym", the average gym-goer has a monthly subscription that costs 17$ per gym visit, in comparison to a single ticket of 10$. DellaVigna and Malmendier write that: "Leading explanations for our findings are overconfidence about future self-control or about future efficiency. Overconfident agents overestimate attendance as well as the cancellation probability of automatically renewed contracts (2006, p. 694)." Sugden interprets the papers in line with Jane's changing personality. His conclusion is that this overestimation is related to Kahneman's maxim: the desire to visit the gym at the moment of contract signing is more important than two months later when people's personality has changed. In other words, Sugden concludes that new gym members do not pay to self-constraint themselves.

However, is Sugden's methodology correct to assert that only these people would suffer from self-control problems? I would argue that this argumentation is odd for various reasons, because Sugden fails to differentiate strategies to mitigate self-control problems. These three strategies are (1) doing nothing, (2) self-command, and (3) self-constraint, of which the latter two terms are inspired by Schubert (2015, p 280).

To examine strategy (1), let us reexamine an example of Sugden discussed in Section 4.4. Thaler's guests had a self-acknowledged problem of eating cashew nuts, but they failed to act upon this self-control problem until Thaler moved the bowl to the kitchen. Thus, Thaler's guests used no self-constraint measures, but Sugden agreed that this group suffered from self-acknowledged self-control problems. Hence, by only investigating self-constraint measures, Sugden omits the group that suffers from self-control problems but does not act upon those. In a Frankfurtian language, he forgets the wantons.

Secondly, self-command is the strategy to exert willpower for people experiencing self-control problems. For instance, one of Thaler's guests is on a diet and has a voracious appetite. The act of resisting the temptation of the cashew nuts is called self-command. Imagine that this guest fails to resist the cashew nuts after 30 minutes of glaring, and eats the cashew nuts while feeling guilty. Thus, this guest has not enough self-command and has a self-acknowledged self-control problem. By only looking at self-constraint measures, Sugden would omit the group that fails at exercising self-command.

Sugden evaluates only the take-up of self-constraint measures to determine that self-control problems occur infrequently, like the take-up rate of panic buttons of William Hill or the number of gym-goers. Without invoking the failure of self-command or the failure to start working on one's self-control problems altogether, Sugden underestimates the severity of self-control problems. In the next subsection, an alternative explanation on the frequency of self-control problems will be given.

7.2. A Higher Frequency of Self-control Problems

Another way to derive how often self-acknowledged self-control problems occur is to conduct an empirical analysis. Interestingly, Hofmann et al. (2012) empirically measure self-control problems with a definition akin to Sugden. The study researched the intensity and frequency of felt desires, to what extent the desires conflict with other goals, and how likely the subjects are to resist temptation induced desires. For Sugden's self-control problems, a study needs subjects who self-acknowledge a desire induced by a temptation that one does not want to act on. To paraphrase Sugden, x induced an urge to do x they found hard to resist but which, even *at the moment of* x, they felt a desire not to act on. This method is, effectively, a psychology based desire-control method. That is exactly what Hofmann et al. (2012) investigate. One staggering result is that "Self-control failure rates were highest for desires to engage in media activities, with 42% of those desires enacted even when people had attempted to resist (Hofmann et al, 2012, p. 585). In total, the authors estimate that "the average adult spends approximately 8 hours per day feeling desires, 3 hours resisting them, and half an hour yielding to previously resisted ones (Hofmann et al., 2012, p. 587)." Following these numbers of Hofmann et al., the statement that people uncommonly endorse their self-control problems is odd if, on average, people feel that they experience 30 minutes of irresistible desires a day. Thus, with Sugden's definition of self-control problems, the statement that self-control problems are uncommonly endorsed is inadequately explained, and the empirical evidence of Hofmann et al. points to the opposite conclusion; people commonly endorse their self-control problems.

However, there is another explanation for the difference between Sugden and Sunstein's survey results. This difference arises due to Sugden's desire-control or meta-desire method, whereas self-control problems can be much broader defined. Combing back to the example of Joe in-front-of-the-mirror of Section 4.3, it can also be the case that Joe endorses his self-control problem right after finishing the cream cake. In this case, his behavior does not classify as a self-control problem according to Sugden's identification method, because Joe did not have second-order desire *at the moment* he was eating. Sugden would argue that Joe did not have any latent preference, but he chooses differently in different contexts (i.e., the eating context or right after eating context) (Infante et al., 2016B, pp. 35-36). However, Joe can still say that he suffers from a self-acknowledged self-control problem if he commits to a different definition of self-control problems. Other disciplines, like economics or

psychology, define self-control problems much broader.[6] The difference in expected frequency of self-control problems can be explained by Sugden's narrow definition of self-control problems. A new definition will be further explored in Section 8.

7.3. Summary of Section 7

To sum up, Sugden's claim that self-control problems occur infrequently was investigated. Firstly, Sugden's methodology to examine the frequency of self-control problems was investigated. Sugden explored the take-up rate of self-constraint measures. By only looking at self-constraints, Sugden is right to state that the take-up rate is small, especially in the case of William Hill. However, Sugden forgets two groups: 'wantons' that do not act upon their self-control problems, and people that fail to exercise self-command. It remains dubious why Sugden considers self-constraining mechanisms and ignores the other two groups. Secondly, by looking at Sugden's (2017) identification method, an empirical paper with a similar self-control problem definition can be consulted. A study by Hofmann et al. (2012) found that self-acknowledged self-control problems occur, on average, 30 minutes per day. Their findings justify that self-control problems are endorsed commonly. Lastly, Sugden fails to account for self-acknowledge self-control problems that are formed with a different identification method. Sugden's definition is narrow in comparison to the other definitions. All in all, Sugden's claim that self-control problems occur uncommonly is ungrounded.

Looking back at the past sections, I argued that Sugden's identification method of self-control problems is irreconcilable with the continuing self, and Hume's decision theory. Furthermore, I suggested that infrequency of self-control problems could be an escape for Sugden, but after this section, the conclusion is that this escape is invalid. To further demonstrate that Sugden's escape is invalid, a new conception of self-control is considered in Section 8. This conception of self-control is rooted in psychology and Pettit (2006), and uses a different model of the self: a goal-oriented self. Self-control problems are considered as a failure to reach one's goals. This definition of self-control problems is broader defined than Sugden's definition. Section 9 will explore the consequences of this new model.

8. A New Definition: Goal-Oriented Self

This section will set forth another definition of the self, and thereby self-control. Hume called the related philosophical branch of identity a 'labyrinth', and I argued that Sugden's self-control problems are narrowly defined. I will defend that a definition of self-control should be rooted in psychology in Section 8.1. In Section 8.2, a goal-oriented model in psychology will be presented. Subsequently, this model of agency will be fitted with Pettit (2006), and the goal-oriented self will be discussed in Section 8.3.

8.1. Self-Control Rooted in Psychology

A new framework of self-control problems should be rooted in psychology. Sugden and I agree that "we need a normative economics that does not presuppose a kind of rational human agency for which

[6] An interesting question, outside the scope of this thesis, would be if Sugden agrees that people can be made better off, AJBT, if they suffer from self-acknowledged self-control problems, but these self-acknowledged self-control problems do not follow his self-control problem identification method. Self-control problems can also be defined more broadly. For instance, Joe in the mirror could endorse his self-control problems, whereas he did not have a desire to resist the cream cake *at the moment of eating*, but only endorsed these problems after eating.

there is no known psychological foundation (2018B, p. 82)." Contrary to Sugden's admiration for Hume's philosophy, I would propose to look at contemporary psychology. With Hume, there is little explanatory power over the actions of people due to his dynamic conception of the self – the continuously changing bundle of desires. That is different in the study of personalities, which is a study that recognizes patterns and coherence in one's life despite the uniqueness of every person. A broadly accepted definition comes from Pervin, who defined personality as follows:

> Personality is the complex organization of cognitions, affects, and behaviors that gives direction and pattern (coherence) to the person's life. Like the body, personality consists of both structures and processes and reflects both nature (genes) and nurture (experience). In addition, personality includes the effects of the past, including memories of the past, as well as constructions of the present and future (1996, p. 414).

From this definition, one could recognize that despite the difference between people and the variations in behavior, several personality traits and behavioral patterns can be established. At this point, there is a wedge between Sugden and psychology. Sugden argued that the simplest explanation is that Jane had no self-control problems because her personality changed. Fortunately, there is more to say about Jane, because psychology has set up similar experiments and developed models in which they describe processes of willpower, self-regulation, and self-control. These experiments do not concern one Jane, but thousands of persons like Jane, making the explanatory power larger than Hume's experiments, which are based on a form of introspection. One experiment, akin to the example of Jane, let two groups of consumers choose between a fruit salad and a cake (Shiv & Fedorikhin, 1999). The group that had to conduct a high-processing-effort task was more likely to opt for the cake compared to the group that had a low-processing-effort task. Thus, a definition of self-control should be rooted in psychology due to their explanatory power due to their empirical experiments.

8.2. A Goal-oriented Model in Psychology

As inspiration for a self-control definition, the *Handbook of Self-Regulation* edited by Vohs and Baumeister is used. In a chapter on *Self-Control and Ego Depletion*, the authors propose the following definition:

> Self-control, defined as the ability to alter one's thoughts, emotions, and behaviors or to override impulses and habits, allows one to monitor and regulate oneself to meet expectations. These expectations can be imposed by society or by oneself, and include laws, norms, ideals, goals and other standards. … Self-control is important not only for following specific rules and meeting specific standards, but also for succeeding in myriad broad domains: academic excellence, occupational accomplishments, stable and satisfying relationships, good adjustment, mental and physical health, overcoming prejudice, resisting addiction, regulation of criminal and violent acts, positive emotionality, and longevity (Maranges & Baumeister, 2016, p. 42).

The essential element of self-control becomes "to meet expectations", which can be roughly equated to achieving goals. The term self-regulation, often interchangeably used for the term self-control, exemplifies the need for regulation in regard to the goal. Baumeister and Tierney write: "Self-control without goals and other standards would be nothing more than aimless change, like trying to diet without any idea of which foods are fattening (2001, p. 63)."

At this point, there is a difference between Sugden's notion of self-control, and a broader psychological definition of self-control. Sugden's notion can be equated with synchronic self-control (problems), which is (the failure of) resisting the temptation in the heat of the moment, because the temptation is in sharp contrast to a goal. The issue is that this analysis severely limits the broader understanding of self-control. For instance, high-scoring individuals on self-control scales would be expected to exert self-control the most. However, De Ridder, Kroese, and Gillebaart (2018) found that individuals with a high score on self-control did *not* exercise this power frequently, instead, they formed habits in which there was less need to exercise this self-control. Self-control should be interpreted in a much broader way. Diachronic self-control, literally the self-control over time, is the ability to prevent coming into the situation in which one has to exercise self-control. Self-control becomes an indicator of success in many disciplines: getting higher grades at school, more interpersonal skills, and less binge eating or alcohol abuse (Tangney, Baumeister, Boone, 2004). Hence, a combination of synchronic and diachronic self-control is needed to form a comprehensive view on self-control problems.

8.3. The Goal-Oriented Self

Both synchronic and diachronic self-control have in common that they aim to achieve a particular goal. After noting that goal-setting is a crucial element of self-control, it is time to reexamine the self, and through this self, self-control problems. I want to introduce a philosophical companion that uses a similar form of agency. I concur that Philip Pettit's (2006) has a valuable position that combines folk psychology and decision theory. Pettit's writes:

> The fundamental tenet of our common sense psychology of human agents is that agency involves acting to realize various goals in a way that is sensible in light of the apparent facts. … Agents seek goals, construe facts, and choose an action that will achieve their goals … . (Pettit, 2006, p. 138).

The benefit of Pettit's position is that goal setting is a process that is similar to preference formation. Imagine that Jane has the robust goal to lose weight. This goal to lose weight will be revealed in her preferences by choosing healthy foods, and one could assign a high level of utility for healthy food *as if* Jane was a utility maximizer. Furthermore, the construction of facts can be interpreted as belief formation, in which Jane can assert a high probability that eating healthy foods will lead to a weight reduction. This agency overlaps roughly speaking with the axioms of rational choice theory.[7]

Following Pettit's agency, I argue that every agent has particular goals in life that are relatively stable over time. These goals can be either positive (i.e., I want to achieve x) or negative (i.e., I do not want to lose x). For instance, these goals can range from not losing one's job, finishing a philosophy master's degree, to becoming a famous football player. The goals relevant for this agency are medium to long-term, which can be roughly translated into goals ranging from months to years. For instance, the goal 'going to the supermarket' is not relevant on agency level. It is important to note that people have different goals in different domains, depending on interest and skill levels. The origin of these goals

[7] An informed preference consequentialist can now impose any form of (libertarian) paternalism, assuming that the goals of the individuals are based on all relevant information. An informed preference consequentialist is committed to preferences based on full information and a form of utilitarianism. Since Sugden and I are not committed to informed preference consequentialism, another solution to tackle self-control problems has to be found. I refer to Qizilbash (2021) for a detailed discussion.

can be defined as a black box process. The goals can be formed due to desires, (moral) feelings, religious beliefs, unconscious processes, or other decision mechanics. In line with liberal thinkers, I find that the determination of the ends (i.e., goals) belongs to the people themselves as long as those ends do not harm other people.

Perceiving goals in the medium to long-term enables us to identify what self-control actions are. Self-control problems can be defined as follows: *repetitive actions that systematically undermine stable medium to long-term goals people set*. With this definition, self-control problems are, as Sugden already suggested, actions that individuals self-acknowledge to be self-control problems. The term stable refers to the fact that agents should have their medium to long-term goals for a considerable period. The term stability overcomes the context-dependency problem. As Kahneman had put it: "Nothing in life is as important as you think it is when you are thinking about it (Kahneman, 2011, p. 402)." For instance, becoming an international volleyball player might be a dream of a young student when she watches an international game, but her desire might vanish over time. It would be inappropriate to classify her actions as self-control problems. Therefore, it is preferable to look at stable goals in the medium to long-term that fluctuate less than short-term goals. Albeit there are some empirical reservations about how to measure goals (Bürger, 2015), there is empirical evidence that personality traits and major life goals remain moderately to highly stable, even over an extent of 20 years (Atherton et al., 2020). In a longitudinal study over one year, the motivation of students remained substantially stable; and well above the majority of students exhibited a stable motivational profile over time (Tuominen-Soini, Salmela-Aro & Niemivirta, 2011). This evidence aligns with the perception that people form habits to work towards those stable goals, and the depiction of the goal-oriented self becomes a viable alternative.

After establishing a new definition of self-control problems, I want to go one step further. In the next section, I will demonstrate that, if many people suffer from self-control problems, less opportunity of markets may be desirable, which I will defend in the subsequent incapability approach.

9. Consequences of Another Self-Control Definition

In this section, the consequences of another definition of self-control will be presented. The high frequency of self-control problems will be discussed in Section 9.1. In Section 9.2, I argue that if self-control problems occur frequently, the contractarian model agency becomes unrealistic. One way to overcome Sugden's incoherency and the unrealistic contractarian model is to consider that some opportunities have a negative value. In order to argue that there is a negative value to opportunity, I want to introduce the *incapability approach* in Section 9.4. This approach is an adaptation of the capability approach developed by Sen, which will be introduced in Section 9.3.

9.1. The Frequency

With this new definition of self-control of failing at one's stable long-term goals, many goal-failing groups can be added to the self-control problem list. For instance, this list can be associated with traditional self-control problems of losing weight and stop smoking. According to the national center for US health statistics, 49.1% of all adults tried to lose weight in the last 12 months (Martin et al., 2018), but many of those fail to meet or stay on their desired weight level. Furthermore, one study estimated that the average smoker needs 30 attempts to stop smoking (Chaiton et al., 2016). Also,

approximately 70% of the current adult smokers in the United States want to stop, 55% attempt to do so accordingly, and a marginal 7% beat the odds and stop smoking (CDC, 2017).

Keeping these numbers in mind, Sunstein's (2018) survey findings seem like an accurate description of self-control problems. These numbers do not form indisputable empirical proof that self-control problems are widespread, and further research is required to answer the question in which domains people fail to meet goals. For the coming argumentation, I need to assume that in some domains, there is a large group of people who systematically fail their goals, which is in line with Sunstein's results.

9.2. An Agency Problem for Contractarians

If people fail to reach their goals in regard to dieting and social media consumption, the question remains: Why would that be a problem? As discussed in ILS's critique in Section 4.3, Sugden's refusal of rational choice theory as a normative theory cannot be reconciled with latent preferences constructed by outsiders. There is no correct theoretical method to construct the individuals' latent preferences. This purification would systematically exclude people that do not choose according to rational choice theory. Furthermore, Sugden's contractarianism refuses paternalism on the grounds that it lacks a valid addressee. In Sugden's language, every attempt to attach a non-individual predicate of self-control problem originates from 'a view from nowhere', which is borrowed from Nagel (1986). In this process, a platform has to be created at which an expert, the behavioral welfare expert, the societal consensus after reasoning, or a social planner can call particular actions a self-control problem. For various reasons that are not of interested in this thesis, Sugden finds these models unrealistic (2018B, p. 24).

However, how realistic is the contractarian model? Sugden's contractarianism wants to address recommendations "to individuals as the directors of their lives (Sugden, 2018B, p. 49)". Only, Sugden does not consider how realistic it is to address individuals as directors of their lives if they, *en masse*, fail at reaching their goals. I would argue that economists find Sugden's contractarian model "fanatical" (Sugden, 2018B, p. 50), because it is based on a too optimistic view of the agency capacity of people. What is the added benefit of addressing a recommendation to eat differently to each individual citizen who already wants to lose weight, but fails to adjust their lifestyles accordingly? Self-control problems that concern many individuals transcend the individual level and a collective action is needed. Without a reorganization of society, these problems cannot be solved.[8] A solution to this issue can be the *incapability approach*. This approach will be presented in Section 9.4, which is derived from the capability approach developed by Sen (1985, 1989).

9.3. The Capability Approach

The capability approach is Sen's evaluation method that looks at "how a person's interests may be judged and his or her personal 'state' assessed (Sen, 1985, p. 1)," in other words, this approach is a method to evaluate individual and societal well-being. Sen complained that economists have focused too much on increasing the gross domestic product (GDP) as an end in itself, but Sen deems that a GDP is an impractical means to an unidentified end. He proposes in his capability approach that the

[8] A contractarian model could also incorporate societal change, but this route will not be explored in this thesis. One reason is that if a policy harms an individual's interest, this individual should receive compensation and that can lead to excessive conservatism (Sugden, 2018B, p. 41).

quality of life should be an end worth pursuing, where this end can be chosen by authentic self-direction. The capability approach should be understood by two core concepts: functionings and capabilities. Functionings are a set of "doing and beings" that refer to states a person has reason to value, and they are not related to commodities themselves. For instance, being literate is a functioning rather than the possession of books. These functionings can vary from "escaping morbidity and mortality, being adequately nourished, undertaking usual movements et cetera, to many complex functionings such as achieving self-respect, taking part in the life of the community, and appearing in public without shame (Sen, 1989, p. 44)." Capability, in its turn, is the set of valuable vectors of functionings that are available to the individual. Taking the capability 'health', one should look at the functionings (i.e., the general availability) of being nourished, doing enough physical exercise, et cetera, and the effective freedom to choose from the set of functioning vectors. Every person has a different conversion factor, which implies an inter-individual variation to convert commodities into functionings (Sen, 1985, p. 16). The goal of the capability approach is to increase subjective well-being by expanding the option set, and make people capable of choosing what they value and have reason to value. Sen's work is originally specified at underdeveloped countries, facing, for instance, famines. The incapability approach is aimed at developed western countries.

9.4. The Incapability Approach

A crucial insight of the capability approach is that Sen makes a difference between substantive freedoms, also called the capabilities, and formal freedoms. In my analysis, I want to argue, especially in the western developed countries, there is a high level of formal freedom, but in some societal areas, there is a low level of *effective* freedom. I want to use an example of the self-acknowledged failure to lose weight.

There are many people that want to lose weight but fail to do so. In Sen's language, they do not have the functioning of being adequately nourished (i.e., not too nourished). From this perspective, too much availability of commodities (i.e., the availability of unhealthy food) may decrease the capability of people to choose appropriate nutrition and live with a body mass that is closer to the individuals' wishes. To speak in Sen's terms, there are different conversion factors in which individuals need a different intake of goods, say food, in order to get a certain functioning, say staying on a desirable weight-level, because individuals have different metabolic rates, body size, gender, medical conditions, et cetera (Sen, 1985, p. 16). There is certainly a threshold where too many calories lead to too many fat gains, AJBT. However, many individuals do not have the functioning of being well-nourished, because they consume too many and unhealthy commodities (i.e., nutrition) with many disadvantages, like high-sugar, high-fat, and high-salt. For a given person, like Jane, having more unhealthy food decreases, after too much intake, her ability to function in society. Being overweight diminishes the *effective* freedom of people, because they live shorter, have a shorter life without impairments, and are more constrained by their body to move. Furthermore, overweight limits the possibility of living a life that people have reason to value. Thus, I would argue that a widespread availability of unhealthy nutrition leads to less *effective* freedom (i.e., of being at a desirable weight level). Therefore, some opportunities to buy unhealthy commodities should be limited to make people more capable to lead the lives they have reason to value.

In a language without Sen's key terms (i.e., without functionings), I argue that the obesity crisis in the developed western world is a problem because people themselves want to lose weight, but fail

consistently, even after multiple attempts. The widespread availability of unhealthy food, like fast food, candy, or chocolate bars, causes people to, influenced by their self-control problems, choose an inadequate diet for losing weight. In the Frankfurtian language, they have a second-order desire to lose weight, but they have no volition to keep acting upon it. The availability of unhealthy food plus self-control problems makes people incapable of losing weight. The inability to lose weight makes people incapable of leading the lives they have reason to value. For instance, they live less long, have more health complications, and can have trouble walking. Self-control problems cannot be taken away, but other luring elements of society, like unhealthy food, can be reduced.

This analysis can be extended to other functionings that are impaired due to self-control problems. As a solution to this impairment of effective freedom, a participatory valuation and evaluation in terms of capability enhancement can be performed by a method of Alkire (2005). In the first stage, philosophers discuss several categories of values, whereas local citizens participate in the subsequent phase. In this phase, they will discuss what values people commit to and what goals they want to reach. The details of this evaluation method are not of primary concern here, but the main point is that, through public reasoning, some of the opportunities might be restricted.

10. Possible Reaction of Sugden and a Response

In this section, I want to anticipate Sugden's reaction to this analysis of self-control problems. Before Sugden's book *The Community of Advantage* was published, Schubert (2015) published some similar critique on the continuing person notwithstanding that his critique had the goal to indirectly promote a different normative criterion: the 'opportunity to learn'. In his opportunity to learn, Schubert promotes some paternalistic measures in case some agents do not learn their future preferences. According to Schubert, the opportunities themselves are not what matters most, but the chance to develop one's preferences. I will indirectly defend some of Schubert's critique on self-control problems by applying Sugden's (2015) response to this paper in Section 10.1. Furthermore, I will add additional criticisms that Sugden could have on a new conception of self-control in Section 10.2.

10.1. An Initial Response

10.1.1. Sugden's question

I would estimate that Sugden's initial reaction to this thesis would encompass a combination of being flattered and disappointed. On the one hand, I presume he would be flattered. Many BE have trouble or fundamentally disagree with his work on preference purification. More attention to his work would be greatly appreciated by Sugden. Furthermore, Sugden acknowledged that his work may not be fully consistent on every detail, and that counter-arguments help his progress (2015, pp. 297-298). On the other hand, I presume that he would be disappointed that I write my whole thesis about Sugden's relationship to self-control problems, whereas, standing in Sugden's shoes, it is a relatively small element of his whole work. The goal of Sugden's philosophical project is to criticize normative economics based on preferences and to build an alternative based on opportunities. He would argue that discussion of self-control is not of crucial importance for the opportunity criterion. Sugden writes:

> In a competitive market, self-constraint technologies will tend to be supplied to those people who are willing to pay for them, but so too will be the counter-technologies that allow people to escape from constraints they no longer wish to be bound by (2015, p. 301).

To say it bluntly, Sugden could ask: why should we bother with self-control problems if markets provide self-regulating technologies and counter-technologies?

10.1.2. A response

The main answer is that our starting positions are different. If the goal is to defend consumer sovereignty, at all costs, explained by "the tendency of the market to supply each consumer with what he or she wants (2015, p. 298)", any exercise of looking at self-control is unfruitful. However, my aim differs significantly from Sugden's defense of consumer sovereignty.

Firstly, it was my goal to show Sugden's inconsistencies regarding self-control problems. I argued that his self-control problem identification method was compatible with Frankfurt's meta-desire theory. This theory conflicts with the continuing self of the opportunity criterion. The continuing self is a *locus of responsibility*, but self-control problems cause the responsible agent to disidentify with herself. Furthermore, I contended that Sugden's identification method for self-control problems only looked at self-constraint measures to conclude self-control problems occur uncommonly. I argued that, within Sugden's definition, the group also consists of people that take no action against their self-control problems, and a group that experiences a failure of willpower. With a psychology rooted definition of self-control problems, I argued that the group of people suffering from self-control problems is even larger. The frequent existence of self-control problems undermines the agency level of a contractarian. A contractarian cannot confidently claim that people are the director of their life when they fail to achieve their goals, AJBT. For that reason, I opted for the *incapability approach*. With the incapability approach, public reasoning, through a democratic process, could decide that the availability of counter-technologies should be reduced. Sugden mentioned an example of a smoker that has a particular disposable unit to throw its cigarettes away (2018B, p. 149). However, Sugden mentions that this smoker, after throwing his cigarettes away, can still go to the grocery store to buy a new package. One solution from public reasoning could be to limit the opportunity to buy cigarettes if the majority of smokers want but cannot quit smoking.

Thus, I fundamentally disagree that consumer sovereignty should be defended on every occasion. I share a great sympathy with the opportunity criterion, but my conclusion is that some opportunities in the market should be limited based on self-control problems, especially when many individuals suffer from self-control problems. I think it is reasonable to defend that some collective self-control problems weigh heavier than the normative costs of decreasing the availability of some (unhealthy) commodities.

10.2. On a New Conception of Self-Control

10.2.1. Sugden's question

Sugden could have a question regarding my proposed definition of self-control problems. I have argued that a self-control problem can be identified as undermining some stable medium- and long-term goals that an individual has. Sugden could argue that this strategy is almost similar to Bernheim's and Rangel's (2007, 2009) method. In a nutshell, Bernheim and Rangel suggest that welfare economists should only take coherent preferences into consideration for a welfare assessment. Their main strategy, through a generalized choice situation, is to eliminate that choices were affected by ancillary conditions (i.e., reasoning imperfections). A necessary step in their identification is that all inconsistent choices have to be removed. In a similar way, all conflicting goals of subjects have to be disregarded. Following

ILS (2016B), there are many people that have multiple conflicting goals, like losing weight and being spontaneous. How would you resolve this tension?

10.2.2. A response

I acknowledge that this tension is hard to resolve. If the goals do conflict, I have to admit that those goals cannot be used to determine what makes people better off, AJBT. If an interviewer would be present, this inconsistency *could* be resolved by confronting the subject in a similar manner as Bleichrodt et al. (2001) do, but an individual is free to have conflicting goals. These conflicting goals cannot be used in any welfare analysis or public debate. My conjecture is that these goals do not conflict too much. In some empirical papers, having conflicting goals are associated with depressive and anxious symptoms (Moberly & Dickson, 2018).

11. Conclusion

Sugden's philosophy aims to challenge behavioral welfare economists to derive their normative implications without drawing on rational preferences, but instead, to propose a normative economics based on opportunity. Sugden critiqued mainstream welfare economics, which was in this thesis represented by TSs book *Nudge* (2008), in which they proposed LP policies. In reply to TS, Sugden (2017, 2018A) argued that self-control problems are a valid criterion to make people better off, but that people uncommonly endorse their self-control problems (2017, pp. 121-122). Following that claim, I analyzed his concepts of the self and self-control and assessed to what extent his philosophy needs to be adjusted if a new conception of self-control is used. In this thesis, I challenged that this identification method can be reconciled with the opportunity criterion and Sugden's interpretation of Hume's decision theory. Furthermore, his claim that people uncommonly self-acknowledge their self-control problems was criticized. Lastly, a new definition of self-control problems was introduced.

To provide a thorough background, an introduction was given to TS's book *Nudge*. TS proposed to nudge people to make them better off, AJBT. One criterion to nudge people was based on self-control problems. Since Thaler provided a behavioral economics background, an additional sketch was given on what axioms of neoclassical economics are violated by BE, as well as an introduction on self-control problems in economics. One particular way to identify self-control problems in behavioral economics is to use a model of meta-desires by Frankfurt (1971).

Afterward, Sugden's solution to identify self-control problems was examined. He argues that latent preferences exist if an individual can self-acknowledge that she suffers from a self-control problem if two conditions hold. She needs to have two conflicting feelings in a state of temptation, of which one feeling is her self-acknowledged true preference that is developed during some *independently definable circumstances*. I argued that, implicitly, Sugden commits to meta-rankings, because the two conflicting feelings represent a first-order and second-order desire, of which the latter is recognized during these *independently definable circumstances*. Consequently, the relation between the continuing or responsible agent and self-control problems is investigated. The continuing agent represents a realistic agent that does not necessarily act according to rational choice theory, and is a cornerstone of Sugden's opportunity criterion. Sugden defined this agent as a *locus of responsibility*, who acknowledged the responsibility for her actions of the past, present, and future. However, I argued that this agent is unreconcilable with self-control problems. Self-control actions make the agent irresponsible with regard to her second-order desire. In Sugden's language, an agent disidentifies with her actions caused

by self-control problems. Moreover, the relation between self-control problems and Sugden's interpretation of Hume was examined, because Sugden expressed his admiration to Hume's decision theory on several occasions. Sugden's perception of Hume does neither allow for any identification of errors nor latent preferences. That means no latent preferences can be identified, including latent self-control problems. With some reservations about Hume's later doubts about identity, Hume cannot identify the self without a perception. Thus, Sugden's method to identify a self-acknowledged preference during some *independently definable circumstances* is impossible because any perception influences the self.

After concluding this incoherency in Sugden's philosophy, Sugden's only escape was to declare that self-control problems are infrequently endorsed. I argued that his method to identify the infrequency of self-control problems was incoherent. By relying on the concepts of doing nothing, self-command and self-constraint from Schubert (2015), I asserted that Sugden only considered self-constraint activities as evidence for self-control problems, but he forgot about the applications of the two former categories. Also, a study by Hofmann et al. (2012), that has a similar definition of Sugden's self-control problem identification method, found that, on average, subjects could not resist their unwanted desires for 30 minutes per day. I would also not call that 'uncommon'. Another issue was that Sugden's identification method is narrowly defined. Other disciplines have other definitions, and more people could self-acknowledge that they have self-control problems. All in all, the claim on infrequently endorsed self-control problems is ungrounded.

Furthermore, I explored a new conception of self-control based on psychology and Pettit (2006). Leaning on their perspective, I proposed that self-control problems can be perceived as the undermining of one's stable medium to long-term goals. From this angle, many people suffer from self-control problems, and that impairs the contractarian agency. With self-control problems, people are not the directors of their live anymore. As a solution, the *incapability approach* was proposed. Keeping the capability approach of Sen in mind, I argued that the effective freedom is diminished if people collectively suffer from self-control problems. In Sen's terms, many individuals do not have the functioning of being well nourished. In this approach, one could think about limiting the availability of high-sugar foods in the light of public reasoning.

Further research could look further into two different directions. An empirical route could look at how often people endorse their actions as self-control problems. This research could investigate the prevalence of self-control problems according to a particular definition. Multiple groups could be constructed based on several definitions of self-control problems, such as one's personal interpretation, Sugden's definition, and a goal-oriented definition. A philosophical research idea could be to investigate the relation between collective self-control problems and contractarianism. The study of Hofmann et al. (2012) already pointed to the fact that 42% of the people failed to resist social media temptation. With the rise of new electronic devices, this number has probably increased over the last ten years. Is a contractarian position philosophically defensible in the light of collective self-control problems? To answer that question, one has to investigate what the merits are of considering citizens as 'directors of their lives' if they suffer from self-control problems. Furthermore, one should research how compensation should be given to individuals that do not suffer from self-control problems.

12. Bibliography

Alkire, S. (2005). *Valuing Freedoms*. New York: Oxford University Press.

Atherton, O. E., Grijalva, E., Roberts, B. W., & Robins, R. W. (2020). Stability and Change in Personality Traits and Major Life Goals From College to Midlife. *Personality and Social Psychology Bulletin*, 47(5), 841–858. doi: 10.1177/0146167220949362.

Baumeister, R. F., & Tierney, J. (2011). *Willpower: Rediscovering the greatest human strength*. New York: Penguin Press.

Baumeister, R. F., & Vohs, K. D. (2016). *Handbook of Self-Regulation: Research, Theory, and Applications* (3rd ed.). New York: The Guilford Press.

Bernheim, D. B., & Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review: Papers and Proceedings*, 97(2), 464-470. doi: 10.1257/aer.97.2.464.

Bernheim, D. B., & Rangel, A. (2009). Beyond revealed preference: choice theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, 124(1), 51-104. doi: 10.1162/qjec.2009.124.1.51.

Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, 47(11), 1498-1514.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and Consumer Choice. *Journal of Political Economy*, 121(5), 803-843. doi: 0022-3808/2013/12105-0004$10.00.

Bürger, K. (2015) *The Stability and Variability of Goals in Learning Contexts: A Systematic Literature Review and a Quantitative Investigation*. In: Schnotz W., Kauertz A., Ludwig H., Müller A., Pretsch J. (eds) Multidisciplinary Research on Teaching and Learning. Palgrave Macmillan, London. doi: 10.1057/9781137467744_2.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism". *University of Pennsylvania Law Review*, 151(3), 1211-1254. doi: 10.2307/3312889.

Centers for Disease Control and Prevention (CDC). (2017). Quitting smoking among adults – United States, 2000-2015. *Morbidity and Mortality Weekly Report.*, 65(52), 1457-1464.

Chaiton, M., Diemert, L., Cohen, J. E., Bondy, S. J., Selby, P., Philipneri, A., & Schwartz, R. (2016). Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. *BMJ*, 6(6). doi: 10.1136/bmjopen-2016-011045.

DellaVigna, S., & Malmendier, U. (2004). Contract Design and Self-Control: Theory and Evidence. *The Quarterly Journal of Economics*, 119(2), 353–402. doi: 10.1162/0033553041382111.

DellaVigna, S., & Malmendier, U. (2006). Paying Not to Go to the Gym. *American Economic Review*, 96(3), 694–719. doi: 10.1257/aer.96.3.694.

Dold, M. (2018). Back to Buchanan? Explorations of welfare and subjectivism in behavioral economics. *Journal of Economic Methodology*, 25(2), 160-178. doi: 10.1080/1350178X.2017.1421770.

Dreier, J. (1996). Rational preference: decision theory as a theory of practical rationality. *Theory and Decision*, 40, 249–76. doi: 10.1007/BF00134210.

The Economist. (2015, January 29). *A crash course in probability*. Retrieved on 07 January 2021 from https://www.economist.com/gulliver/2015/01/29/a-crash-course-in-probability.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. doi: 10.1177/1745691612460685.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5. doi: 10.2307/2024717.

Hansen, P. (2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation*, 7(1), 155-174. doi: 10.1017/S1867299X00005468.

Hausman, D. M. (2010). Hedonism and Welfare Economics. *Economics and Philosophy*, 26(3), 321–344. doi: 10.1017/s0266267110000398.

Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy*, 18(1), 123–136. doi: 10.1111/j.1467-9760.2009.00351.x.

Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What People Desire, Feel Conflicted About, and Try to Resist in Everyday Life. *Psychological Science*, 23(6), 582–588. doi: 10.1177/0956797612437426.

Hume, D. (1740/2009). *A Treatise of Human Nature*. The Floating Press.

Infante, G., Lecouteux, G. & Sugden, R. (2016A). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, 23(1), 1-25. doi: 10.1080/1350178X.2015.1070527.

Infante, G., Lecouteux, G. & Sugden, R. (2016B). 'On the Econ within': a reply to Daniel Hausman. *Journal of Economic Methodology*, 23(1), 33-37. doi: 10.1080/1350178X.2015.1070526.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review*, 118, 97–109.

Lever, J. G. P. (2021). On Time Inconsistencies: Is the DI-index a game-changer for identifying self-control problems?. *Business Economics*. Retrieved from: http://hdl.handle.net/2105/55641.

Li, C., Li, Z., & Wakker, P. P. (2014). If nudge cannot be applied: a litmus test of the readers' stance on paternalism. *Theory and Decision*, 76(3), 297-315. doi: 10.1007/s11238-013-9375-2.

Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, 65(3), 272–292. doi: 10.1006/obhd.1996.0028.

Marange, H. M., & Baumeister, R. F. (2016). Self-Control and Ego Depletion. In R. F. Baumeister & K. D. Vohs (Ed.), *Handbook of Self-Regulation: Research, Theory, and Applications* (3rd ed., pp. 42–61). New York: The Guilford Press.

Martin C. B., Herrick K. A., Sarafrazi, N., & Ogden, C. L. (2018). *Attempts to lose weight among adults in the United States, 2013–2016*. NCHS Data Brief, 313. Hyattsville, MD: National Center for Health Statistics.

Moberly, N. J., & Dickson, J. M. (2018). Goal conflict, ambivalence and psychological distress: Concurrent and longitudinal relationships. *Personality and Individual Differences*, 129, 38–42. doi: 10.1016/j.paid.2018.03.008.

Nagel, T. (1986). *The View From Nowhere*. Oxford: Oxford University Press.

Noonan, H. W. (1999). *Routledge philosophy guidebook to Hume on knowledge*. New York: Routledge.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin Review*, 11, 988–1010. doi: 10.3758/BF03196730.

Pervin, L. A. (1996). *The science of personality*. New York: Wiley.

Pettit, P. 2006. Preference, deliberation and satisfaction. *Philosophy*, 81(Suppl. 59), 131–153.

Qizilbash, M. (2021). Informed preference consequentialism, contractarianism and libertarian paternalism: on Harsanyi, Rawls and Robert Sugden's The Community of Advantage. *International Review of Economics*, 68(1), 67–88. doi: 10.1007/s12232-020-00361-x.

Ridder, D., de, Kroese, F., & Gillebaart, M. (2018). Whatever happened to self-control? A proposal for integrating notions from trait self-control studies into state self-control research. *Motivation Science*, 4(1), 39–49. doi: 10.1037/mot0000062.

Schubert, C. (2015). Opportunity and preference learning. *Economics and Philosophy*, 31(2), 275–295. doi: 10.1017/S0266267115000139.

Sen, A. (1985). *Commodities and Capabilities*. Amsterdam: Oxford University Press.

Sen, A. (1989). *Development As Capability Expansion*. Retrieved on 2 June 2021 from https://livelihoods.net.in/wp-content/uploads/2020/05/DEVELOPMENT-AS-CAPABILITY-EXPANSION.pdf.

Shiv, B., & Fedorikhin, A. (1999). Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *Journal of Consumer Research*, 26, 278-292. doi: 0093·53011200012603·0005$03.00.

Strotz, R. H. (1955-1956). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, 23(3), 165-180. doi: 10.2307/2295722.

Sugden, R. (2004). The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences. *American Economic Review*, 94(4), 1014–1033. doi: 10.1257/0002828042002714.

Sugden, R. (2006). Hume's Non-instrumental And Non-propositional Decision Theory. *Economics and Philosophy*, 22(3), 365–391. doi: 10.1017/s0266267106001027.

Sugden, R. (2009). On Nudging: A Review of Nudge: Improving Decisions About Health, Wealth and Happiness by Richard H. Thaler and Cass R. Sunstein. *International Journal of the Economics of Business*, 16(3), 365–373. doi: 10.1080/13571510903227064.

Sugden, R. (2010). Opportunity as Mutual Advantage. *Economics and Philosophy*, 26(1), 47–68. doi: 10.1017/s0266267110000052.

Sugden, R. (2015). Opportunity and Preference Learning: A Reply to Christian Schubert. *Economics and Philosophy*, 31(2), 297–303. doi: 10.1017/s0266267115000140.

Sugden, R. (2017). Do people really want to be nudged towards healthy lifestyles? *International Review of Economics,* 64(2), 113–123. doi: 10.1007/s12232-016-0264-1.

Sugden, R. (2018A). 'Better off, as judged by themselves': a reply to Cass Sunstein. *International Review of Economics*, 65, 9–13. doi: 10.1007/s12232-017-0281-8.

Sugden, R. (2018B). *The Community of Advantage: A Behavioral Economist's Defence of the Market*. New York: Oxford University Press.

Sugden, R. (2020). The Community of Advantage: An Interview With Robert Sugden. *Erasmus Journal for Philosophy and Economics*, 13(1), 61–78. doi: 10.23941/ejpe.v13i1.483.

Sugden, R. (2021). Hume's experimental psychology and the idea of erroneous preferences. *Journal of Economic Behavior & Organization*, 183, 836–848. doi: 10.1016/j.jebo.2020.11.017.

Sunstein, C. R. (2018). "Better off, as judged by themselves": a comment on evaluating nudges. *International Review of Economics*, 65, 1–8. doi: 10.1007/s12232-017-0280-9.

Sunstein, C., & Thaler, R. (2003). Libertarian Paternalism Is Not an Oxymoron. *The University of Chicago Law Review*, 70(4), 1159-1202. doi: 10.2307/16005.73.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, 72(2), 271–324. doi: 10.1111/j.0022-3506.2004.00263.x.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee savings. *Journal of Political Economy*, 112, 5164-5187. doi: 10.1086/380085.

Thaler, R. H., & Sunstein, C. (2003). Libertarian Paternalism. *American Economic Review*, 93 (2), 175-179. doi: 10.1257/000282803321947001.

Thaler, R. H., & Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven: Yale University Press.

Tuominen-Soini, H., Salmela-Aro, K., & Niemivirta, M. (2011). Stability and change in achievement goal orientations: A person-centered approach. *Contemporary Educational Psychology*, 36(2), 82–100. doi: 10.1016/j.cedpsych.2010.08.002.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131, doi: 10.1126/science.185.4157.1124.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323. doi: 10.1007/bf00122574.

World Health Organization. (2020, April 1). *Obesity and overweight.* Retrieved on 4 January 2021 from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.