

Weakness of will. The limitations of revealed preference theory

ALEKSANDER OSTAPIUK* 

Wrocław University of Economics and Business, ul. Komandorska 118/120, 53-345, Wrocław, Poland

Received: April 7, 2020 • Revised manuscript received: November 4, 2020 • Accepted: November 27, 2020

© 2022 Akadémiai Kiadó, Budapest



ABSTRACT

The phenomenon of weakness of will – not doing what we perceive as the best action – is not recognized by neoclassical economics due to the axiomatic assumptions of the revealed preference theory (RPT) that people do what is best for them. However, present bias shows that people have different preferences over time. As they cannot be compared by the utility measurements, economists need to normatively decide between selves (short- versus long-term preferences). A problem is that neoclassical economists perceive RPT as value-free and incorporate present bias within the economic framework. The axiomatic assumption that people do what is best for them leads to theoretical and practical dilemmas. This work examines weakness of will to resolve some shortcomings of RPT. The concept of intention is used to provide multiple self conception with the framework to decide between selves, which had not been done before. The paper concludes that individuals should not always follow their revealed preferences (desires) but the intentions (reason) because the latter indicates what people really want.

KEYWORDS

weakness of will, revealed preference theory, choice over time, present bias, multiple selves

JEL CLASSIFICATION INDICES

B20, B40, D15

1. INTRODUCTION

Neoclassical economists cannot accept the existence of weakness of will, which leads to many problems. To know why they cannot accept this phenomenon, a brief history of economic thoughts is presented in this paper and the axiomatic assumptions of the revealed preference

* Corresponding author. E-mail: aleksander.ostapiuk@gmail.com

theory (RPT) are indicated. One of them – preference consistency – is especially important for weakness of will because it occurs mainly when people choose between the present and the future (Ainslie 1975, 2001). Neoclassical economists argued that individuals compare their choices over time by utility measurements and choose the best option (Samuelson 1937). These axiomatic assumptions mean that people never experience weakness of will because they stick to their preferences over time and choose what they want.

These assumptions are not realistic due to hyperbolic discounting, which shows that preferences are not consistent over time. Moreover, the comparison of intertemporal choices by utility measurements, as done in neoclassical economics, seems very problematic as utility is relative. Therefore, some authors argue that we do not have one self with well-defined preferences but multiple selves with different preferences that clash. This was illustrated by the precommitment strategies (Schelling 1980, 1984).

Although present bias demonstrates that the discounted utility model used by neoclassical economists is unrealistic, it is not sufficient to dismiss RPT because hyperbolic discounting can be incorporated into economic models. However, this maneuver does not resolve the problems faced in economics, which is proven by the limitations of RPT. These are presented in the case of the axiomatic assumption that agents choose what is best for them, which explicitly dismisses weakness of will. It is argued that this assumption leads to two crucial problems for economics: first, instrumental rationality, where means and not ends are important, leads to a tautology; second, revealed preferences are not always a good indicator of well-being. This paper analyzes weakness of will to check if it can help economics to address the above-mentioned issues.

This article does not aim to resolve the long and complicated discussion on weakness of will, but to provide economists with a framework to decide which self should be in control, “long-term” or “short-term”? This had not been done successfully by the previous supporters of multiple self-conception, and hence, it did not gain popularity in economics (exception: Davis 2003, 2010). Holton’s method of differentiating intentions from desires is used to propose this framework (Holton 1999, 2003). It is argued that intentions point out what the real preferences of an individual are contrary to the revealed choices that frequently result from desires. The conception of intention is also used to provide precommitment strategies (Schelling 1980, 1984) with a conceptual framework, and to support “long-term human” who is guided by reason and long-term preferences. In addition, it shows that neoclassical economists should distance themselves from their conception of utility which is always maximized and can mean anything. Moreover, values should be distinguished from utility because the distinction made within the utility framework does not work in the case of choices over time. This is demonstrated by Mill’s (1863) unsuccessful attempt to save utilitarianism.

Although we used a method to distinguish intentions from desires and precommitment strategies had identified what the real preferences of individuals are, it would not be enough to justify why individuals should follow intentions and long-term preferences. The discussion between Kant and Hume, and the insights from the contemporary philosophers like Dworkin (1988) and Frankfurt (1971, 1988) indicate that individuals should follow their intentions/commitments and not emotions/desires because only the former supports autonomy and freedom.

Before starting our analysis, it is necessary to clarify the following. This paper focuses on traditional RPT (Samuelson 1938) which is still used by the neoclassical economists. According to RPT, choices are synonymous with preferences and economists are not interested in why



people choose something. Furthermore, goals are not assessed. It is assumed that people do what is best for them, and rationality is defined by technical criteria (consistency and completeness of preferences). In general, RPT is perceived as value-free by the neoclassical economists (Gul – Pesendorfer 2008). Of course, RPT and rational choice theory have been criticized for many years from various perspectives (Sen 1973, 1977, 1993; Hausman 2008, 2012; Herfeld 2020). Moreover, currently, there is no single and universal conception of RPT, and many different brands are available (Hands 2013). Here, we focus on the traditional RPT (Samuelson 1938). Despite its age and being an object of constant criticism, RPT seems to be the main approach used by the neoclassical economists which shows their methodological stance (Hands 2013, 2015; Angner 2018). Moreover, we discuss Samuelson’s RPT because our goal is to propose the enriched version of multiple self (Schelling 1980, 1984; Elster 1987), which was designed in the context of traditional RPT.

Last but not least, although the paper underlines some limitations of RPT, it does not indicate that the theory should be thrown away. Economists have good reasons to use it and have been using it successfully over the years. This paper aims to show the limitations of RPT in the case of weakness of will and proposes a framework that can address some problems concerning choices over time. It can be argued that the contemporary RPT addresses some of the theory’s shortcomings concerning the behavioural foundations (Ross 2005; Hands 2013; Hedoin 2016). However, the normative context of multiple self and agency, which is the main theme of this paper, has not been fully addressed by the contemporary versions of RPT (Hedoin 2016).

2. THE ORIGINS OF REVEALED PREFERENCE THEORY

Neoclassical economics emerged at the beginning of the 20th century. Economists wanted it to be as objective and clear as the hard sciences such as mathematics and physics, and therefore, dismissed psychology and perceived normative approaches as unscientific. The path toward objectivity ended with RPT. Thanks to its axiomatic assumptions, RPT always works. We briefly analyze this process to explain why weakness of will does not have a place in neoclassical economics.

The most important person in this process was Lionel Robbins. From our perspective, the primary conclusion from his essay (Robbins 1932) is that economics should not deal with ethics or human goals because normative theories are unscientific. Robbins was influenced by logical positivism which assumes that ethical statements do not express logical propositions but emotional attitudes. He indicated the difference between ethics and economics as follows: “Economics deals with ascertainable facts; ethics with valuations and obligations. The two fields of inquiry are not on the same plane of discourse” (Robbins 1932: 148). Due to logical positivism, all value judgments were put into one bag of nonsenses which cannot be rationally discussed. Therefore, Robbins wanted to distinguish economics from ethics and argued that “economic analysis is *wertfrei* (value-free)” (Robbins 1932: 91).

The next reason behind formulating neoclassical economics as a value-free science is removing psychology from economics, a process which took place at the beginning of the 20th century and is called “Pareto turn.” Many economists, including Bruni – Sugden (2007), comprehensively explained how it was done. Hence, only the main factors of this process are described here. Initially, marginalists believed in *cardinal utility* which allows assessing and



comparing utility. However, they could not find a convincing method to measure utility which attracted criticism (Robbins 1932). Psychological hedonism quickly lost intellectual credibility, and it was evident to marginalists that they had to reform its theoretical foundations. Cardinal utility with the belief that it is possible to assess utility was untenable. Therefore, economists started to use *ordinal utility* in which consumers only rank their preferences. The inability to assess utility resulted not only in a new ordinal utility approach, but also, in escaping from a debate about human goals. It was thought that happiness is too elusive, and economists should not discuss it.

This escape from psychology and hedonistic foundation ended with the development of RPT by Samuelson. Logical positivism was again the main inspiration for economics. Samuelson (1938) tried to base the consumer choice theory on strictly observable foundations. The purpose of his approach was not to “reveal” preferences but to create a strictly operational theory of consumer choice in which preferences or utility do not matter at all. He promised to remove the last “vestigial traces of the utility concept” (Samuelson 1938: 61) from the consumer choice theory.¹ Moreover, Samuelson was not interested in explaining (why people choose something) but instead wanted to describe the world, for which he needed a method of measurement, and therefore, based his model on the observable terms and consequences.

The main purpose of RPT was practical. We need to have observable and measurable data for building a theory. We cannot observe preferences directly and can only look at the act of choice. Samuelson based his theory on this assumption. According to his theory, choices are consistent if they satisfy the “weak axiom of revealed preference” (WARP), which means that if x is revealed to be preferred to y , then y must not be revealed to be preferred to x . If choices satisfy the consistency conditions, we can construct a complete, transitive and continuous revealed-preference ordering (Sen 1973).

One of the most important features of RPT is its particular perception of rationality. It states that agents are rational when they act in accordance with WARP. Therefore, rationality is instrumental because economists do not want to know about people’s motivations. Neoclassical economists are solely interested in the results but not in the causes of a behaviour. They stopped being interested in human motivations and assumed *a priori* that people are always rational and maximize their utility. However, their utility is different from Bentham’s or marginalists’. Sometimes, economists speak about individuals aiming to maximize utility. However, this does not mean that utility is perceived as an ultimately good thing. To be a utility maximizer is to choose only what one prefers. To say that agents maximize their utility implies nothing about the nature of their preferences; rather, it only connects preference and choice. Rational individuals rank available alternatives and choose what they prefer the most (Hausman 1992).

Samuelson greatly influenced neoclassical economics, which instrumentally perceives rationality, has no interest in human motivations, and axiomatically assumes that choice “reveals” preferences. Thanks to the axiomatic assumptions of RPT, economics does not need to be entangled with a normative investigation of the goals (Gul – Pesendorfer 2008). Due to this reason, the conception of weakness of will, which assumes that people sometimes do not do what they want, cannot be accepted by neoclassical economists who use RPT to this day.

¹In the 1950s, however, it was obvious that this theory was not different from the ordinal approach (Wong 2006: 73–74, or Hands 2014).



3. CHOICES OVER TIME. HYPERBOLIC DISCOUNTING AND REVEALED PREFERENCE THEORY

As was indicated earlier, RPT and rational choice theory have many limitations and have been widely criticized ([Herfeld 2020](#)). One issue that is essential from the perspective of this article is that the assumption of preference consistency is problematic when it comes to choices over time. In contrast to the neoclassical assumptions, individuals' behaviours do not always remain consistent. Therefore, neoclassical economists have limited ability to analyze and explain intertemporal choices that do not adhere to WARP. They are indeed focused on the moment where preferences are revealed through choice. However, the neoclassical economists knew that choices at present and in the future differ from each other significantly. Stanelly Jevons wrote:

“To secure a maximum of benefit in life, all future events, all future pleasures or pains, should act upon us with the same force as if they were present [...] But no human mind is constituted in this perfect way: a future feeling is always less influential than a present one.” ([Jevons 1911](#): 72–73)

Economists have tried to resolve the problem of differences in choices over time. For instance, they have acknowledged that the present choices are valued more than the later ones, and hence, they discount future utility. [Samuelson \(1937\)](#) introduced the discounted utility model. In neoclassical economics, individuals are presumed to assess utility that would be received in the future. In his paper, Samuelson pointed out some possible problems related to his model that he calls “serious limitations” (1937: 159). One of his concerns was whether people discount the future at rates that vary over time; for instance, they may change their minds as time moves forward. It bothered Samuelson that people might display time-inconsistent preferences. He was aware that such behaviours exist, e.g., people purchase total life insurance as a compulsory savings measure. However, he did not investigate the problems regarding inconsistency further. Economists followed his suit, and the discounted utility model became a standard of intertemporal choice ([Thaler 2015](#): 101). It is not surprising that with time, economics dealt with an increasing number of cases, which showed that individuals cannot discount future utility according to an exponential rate resulting from the present bias. In fact, people try to discount future utility but frequently fail to predict what will improve their well-being ([Gilbert 2007](#)). Moreover, the complications with discounting utility are not only ordinary anomalies, but also, indicate a systematic problem which suggests that the foundations of economic human – stable preference – cannot be fulfilled.

One of the problems with evaluating the future is present bias, which is one of the main reasons for hyperbolic discounting ([Ainslie 1975, 2001](#); [Loewenstein – Elster 1992](#); [O'Donoghue – Rabin 1999](#)). Present bias can be illustrated with the following example. If people are asked to choose 100\$ in 30 days or 105\$ in 31 days, the majority will choose the second option. However, if they choose between 100\$ today and 105\$ tomorrow, almost everyone will pick the first option. The difference between the choices is not a problem for economics because the discounted utility model assumes that the future is less important than the present ([Samuelson 1937](#)). Therefore, waiting one day after 30 days to get 5\$ is easier than waiting one day from today. However, as it turns out, people do not discount the future according to an exponential rate. They opt for 105\$ in 31 days, but when the 30th day passes and they have to choose between 100\$ now and 105\$ tomorrow, they will switch their preference to 100\$. This behaviour



is inconsistent with the long-term preferences and means that individuals over-value their present choices. All along, they want some choice more than the other but when selection arrives their preferences change. For neoclassical economics, present bias is not a problem because at the end individuals make choices and their cognitive biases are the reason why rationality is not perfect (people optimize). The issue here is that neoclassical economists axiomatically assume that people do what is best for them. We will discuss the problems resulting from this approach later.

Economists were not interested in time inconsistency until the 1980s, except for [Strotz \(1955\)](#). George Ainslie, a psychologist, introduced the idea of hyperbolic discounting ([1975](#)). [Thaler – Shefrin \(1981\)](#) implemented this idea in economics. It gained credibility with time, and economists had to admit that hyperbolic discounting is a common feature of human behaviour. However, economists dealt with inconsistent time preferences in a specific way. They created models where choices over time are compared by utility measurements considering hyperbolic discounting among other manifestations of bounded rationality such as, self-control, temptation, dual-self, and so on ([Laibson 1997](#); [O'Donoghue – Rabin 1999, 2001](#); [Gul – Pesendorfer 2001, 2004](#); [Loewenstein – O'Donoghue 2004](#); [Fudenberg – Levine 2006, 2012](#); [Andreoni – Sprenger 2012](#); [Cohen et al. 2020](#); for an overview: [Ericson – Laibson 2019](#)). The process by which these models work will not be elaborated because they encounter similar problems as the static models (discounted utility model) which will be analyzed in the next section.

However, it is necessary to provide a few words of explanation. Most of the above-mentioned models incorporate choices over time within the utilitarian framework which allows comparing them. The next section objects to this idea and indicates that economists need to normatively distinguish different selves with separate preferences. Moreover, the example of [Becker – Murphy \(1988\)](#) of the rational addict is presented to show that the utilitarian framework can be used to support the assumption of RPT that people always choose what is best for them. Of course, the models taking hyperbolic discounting into consideration improve economics. However, most of them do not deal with RPT's foundations (relativity of utility and weakness of will). The next sections argue that RPT is not value-free and has implications for people's well-being. Therefore, whether they want or not, economists decide what is good for people. This position is strongly contested by many who are reluctant toward behavioural welfare economics where consumers' autonomy is questioned ([Gul – Pesendorfer 2008](#); [Sugden 2018](#)). It is thus necessary to present why weakness of will is real and how one can know people's real preferences if one does not rely on people's choices.

4. CHOICES OVER TIME AND UTILITY MEASUREMENTS

Hyperbolic discounting indicates that the assumption of consistent preferences over time is unrealistic. This argument was used to introduce the conception of multiple self into economics, which was done by [Schelling \(1980, 1984\)](#) and [Elster \(1987\)](#). They argued that the hyperbolic discounting makes it impossible to compare choices over time by utility measurements. Therefore, we cannot indicate a simple, enduring self but should distinguish multiple selves with different preferences. Moreover, the conflict between long- and short-term preferences is observed frequently. Schelling addresses this with the example of a precommitment strategy in which people deliberately restrict their freedom because they do not want to succumb to a short-



time temptation. This was illustrated by Ulysses, who ordered his crew to tie him to the mast so that he could listen to the sirens' song without falling under their spell (Elster 1987). Schelling considers this behaviour as really disturbing when he refers it to neoclassical economics and RPT. He indicates that people frequently choose between long-term goals and short-term pleasures, and sometimes, curb their freedom to achieve the former. For example, we do not want anyone to give us a car's keys when we are drunk or a cigarette when we are trying to quit smoking (Schelling 1984).

According to Schelling, the problem here is not that these situations exist, but a question of which self should be in control: the self that has long-term preferences or that which falls into the temptation? He uses the example of Captain Ahab from *Moby Dick*, who lost his leg because of a bad wound:

"The blacksmith enters with a hot iron to cauterize the stump. Ahab begs not to be burned. The crewmen hold him down as he spews out the apple in a scream, and steam rises where the iron is tormenting his leg. The movie resumes with Ahab out of pain and apparently glad to be alive." (Schelling 1984: 9)

At first, Schelling treats this example as an easy situation. We get much utility in exchange for our limited freedom – life in exchange for a short but very intensified moment of pain. Later, however, he notes that even if Ahab asked the smith to cauterize his wound and even if he felt grateful for the operation was performed upon him, it would not mean that this was the best decision that gave him the greatest utility. Schelling writes, "If you burn me so that I may live I'll thank you, afterward, but that is because I'll be feeling no pain and not anticipating any when I thank you" (1984: 9). He concludes by saying, "How do we know whether an hour of extreme pain is more than life is worth? [...] The conclusion that I reach is that I do not know, not for you and not for me" (1984: 9).

Schelling stated that it is impossible to assess which choice is better if only utility is considered because every moment is different. Thus, he reached a similar conclusion to neoclassical economics. It is not possible to compare choices in terms of cardinal utility as utility is relative. As far as neoclassical economics is concerned, it is thought that if we cannot assess whether some action gives more utility than others, we axiomatically assume that people are rational and always maximize their utility. If we did not assume that, RPT would not work. Schelling comes to another conclusion that it is impossible to compare utility over time because "Each self is a set of values [...] on the particular issues on which they differ fundamentally there doesn't seem to be any way to compare their utility increments and to determine which behaviour maximizes their collective utility." (Schelling 1984: 8).

We cannot use the utility measurements because human beings have two different selves with different preferences. Thus, these selves should be handled separately by economics. Schelling did not want to decide which of these two are more important as their utilities cannot be compared. However, this relativity is not a problem because "Sometimes, but not always, it is easy to know which is Dr. Jekyll and which is Mr. Hyde" (Schelling 1980: 98). We do not want to be a person who drinks and becomes vicious or someone who goes to a casino to have fun and loses all her money (Schelling 1980).

Schelling was aware that he did not have a rigid framework to decide which self is more important: "Anyone who is happily addicted to nicotine, benzedrine, valium, chocolate, heroin, or horse racing, and anyone unhappily addicted who would not elect the pains and deprivations



of withdrawal, are not my subject” (Schelling 1984: 4). However, he made a subtle distinction which helped him to justify a choice “I am not concerned with whether cigarettes or rich desserts are bad for you, only with the fact that there are people who wish so badly to avoid them that, if they could, they would put those commodities beyond their own reach” (Schelling 1984: 4–5).

Even though Schelling did not want to choose, it can be seen from the last quotation that in the end, he chooses reason and the long-term human over a person focused at the present moment. His slight emphasis on people who succumb to temptations against their better judgment is known as *akrasia* or weakness of will.² Neoclassical economists do not recognize *akrasia* because they assume that people do what is best for them (following RPT). Thaler (1980) discussed why neoclassical economics resists the existence of *akrasia*. In the 1970s, he organized a supper for his co-workers from Chicago, known for their orthodox approach in which people are assumed to be fully rational. Thaler prepared a bowl of cashews for the guests to whet their appetite before supper. In a short time, it appeared that the guests too often reached for the snack, so Thaler took away the bowl as he did not want the guests to overeat before supper. His decision was met with the guests’ approval. They wanted to eat only a few cashews, but they gave into temptation and ate more. Their approval does not make sense from the perspective of neoclassical economics because people are rational and know what is best for them. Therefore, having more choices is better because individuals can find a more preferable one. Thaler presents a hypothetical conversation between *homo oeconomicus* (ECON) and a human being about this topic:

ECON: Why did you remove the cashews?

HUMAN: Because I did not want to eat any more of them.

ECON: If you did not want to eat any more nuts, then why go to the trouble of removing them? You could have simply acted on your preferences and stopped eating.

HUMAN: I removed the bowl because if the nuts were still available, I would have eaten more.

ECON: In that case, you prefer to eat more cashews, so removing them was stupid.

(Thaler 2015: 94–5)

This discussion leads nowhere. Neoclassical economics cannot admit that weakness of will exists because it would mean the dismissal of RPT where choice is synonymous with preference.

5. PROBLEMS WITH REVEALED PREFERENCE THEORY

The most substantial problem from the perspective of this article is not that RPT is descriptively inaccurate. For neoclassical economics, it is not an issue (Gul – Pesendorfer 2008), and the more important problem is the tautologicality of utility and rationality in which no available data can refute the theory that people maximize their utility and behave rationally (Sen 1977; Boland 1981; Hodgson 2012; Ostapiuk 2020). The same problem is indicated in this paper, but the main

²For now, both terms are used interchangeably, even though some philosophers make distinction between them (e.g., Holton 1999). The reason for this simplification is to ensure clarity for the readers who do not need to be entangled in philosophical discussion at this point.



focus is on choices over time and weakness of will. This is shown using the *Theory of Rational Addiction*, in which Becker – Murphy (1988: 675) claim: “that addictions, even strong ones, are usually rational in the sense of involving forward-looking maximization with stable preferences.”

Drugs is a topic where they push the rational choice theory to its logical extreme. The article indicates that addicted people are not irrational, and their behaviours are not anomalous at all. Moreover, Becker and Murphy argued that the use of drugs maximizes an addict’s utility. This is possible because instrumental rationality is used: “This paper relies on a weak concept of rationality that does not rule out strong discounts of future events. The consumers in our model become more and more myopic as time preference for the present (a) gets larger” (Becker – Murphy 1988: 683). This means that taking drugs is rational because the pleasure obtained from heroin now is so immense that it overshadows the future costs. Therefore, the use of drugs can maximize people’s utility. It looks strange that heroin addiction is rational if one looks at its results. However, it is not a problem because Becker did not assume what is good or bad for people. We just assume that people maximize their utility in the moment of choice. In the case of drugs, people make decisions between present and future utility. Addicted people understand that they are addicted. However, they choose to take drugs because the gains from taking them exceed the costs of not taking them (e.g., fighting with their addiction). Becker and Murphy assumed that weakness of will (*akrasia*) does not occur and people compare their long-term goals (e.g., health) with short-term pleasures (e.g., smoking). Economists do not know how people do it (they are not interested in motivations). They just assume that people perform this cost-benefit analysis. Hence, economists can know about hyperbolic discounting and incorporate it into RPT.

The methodological belief in utility maximization is strong in the economic approach. Even if people regret some decisions made at the spur of the moment, it does not mean that they do not choose what they want.

“The claims of some heavy drinkers and smokers that they want to but cannot end their addictions seem to us no different from the claims of single persons that they want to but are unable to marry or from the claims of disorganized persons that they want to become better organized. What these claims mean is that a person will make certain changes – for example, marry or stop smoking – when he finds a way to raise long-term benefits sufficiently above the short-term costs of adjustment.” (Becker – Murphy 1988: 693)

According to Becker and Murphy, people’s declarations do not mean anything. If abstinence from smoking gives one more utility than smoking, the person will quit cigarettes. Ultimately, in economics, the deeds, not the words, matter the most.

The problem with this understanding of rationality and utility is that many examples indicate when it is strange to assume that people do something because they really want it. Following this logic, when we see how much pleasure one can obtain from using drugs, we are forced to admit that it maximizes utility. However, the question is whether this understanding of rationality and utility makes sense. In this approach, an ex-alcoholic who does not want to drink because she knows where it might lead to is rational when she comes back to drinking. She does not want to drink, and after going on a binge, she feels awful about her behaviour. Is it possible to decide if the temporary need to relapse to alcohol is more important than long-term goals? Becker and Murphy assume that people can integrate these conflicting motivations and bring them down to a common utility denominator. This understanding of rationality gives rise to some absurd conclusions:



“Imagine the situation where, owing to the shipwreck a man lies on a small boat in the middle of the ocean, and he thinks about drinking water from the sea. He knows that he should not drink this water and it can only make him feel bad. Despite this knowledge, he succumbs to the temptation and drinks it. He gets sick and dies afterwards. Is it reasonable to say that this person behaves rationally? Is impulse, second utility enough to call some behavior rational?” (Ostapiuk 2019: 643)

The above examples raise the question if we should not distinguish between reason and biological needs, as well as long-term goals and trivial pleasures. Instrumental rationality does not differentiate between behaviours and can mean anything that leads to a tautology. Obviously, the previous examples are extreme; however, the resulting questions concern most human actions. Individuals frequently choose between now and later. Consume now or save for later; have fun now or work hard now and have more fun later. The neoclassical economists assume that people always choose what is best for them, but it is hard to agree with this assumption when we look at hyperbolic discounting and behavioural economics’ findings (Frederick et al. 2002; Thaler 2015).

Neoclassical economics which axiomatically assumes that what people do is the best for them leads not only to theoretical but also to real problems. The biggest among them is that many economists treat RPT as a positive, not normative theory (Thaler 1980, 2015).³ As a result, many perceive well-being in economics as a descriptive theory (Gul – Pesendorfer 2008). This is wrong because there is no descriptive theory of well-being (Reiss 2013). It is impossible to discuss well-being without making some value judgments. Many economists did not see that because they used the formal and not the substantive theory of well-being which does not tell what is good ultimately but tells how to find what is good for people (Hausman – McPherson 2006). We should just wait and see what people choose (revealed preferences). Reiss concludes that “economists should not have substantive views about the conception of the good. But this is of course to make a moral judgment and to subscribe to a particular theory of well-being: namely, wellbeing is what people desire” (Reiss 2013: 214).

The concept of well-being based on revealed preferences is very problematic and has been criticized over the last years. Behavioural economics and economics of happiness demonstrate that people do not have stable preferences, and often, preferences are created “on the fly” (Hausman 2012: 134). Preferences are context-dependent (Kahneman – Tversky 1979), change with time (Ainslie 1975), and are malleable (Brickman et al. 1978). Moreover, people have cognitive biases that are not anomalies but common to human reasoning and judgmental processes. Because of these findings, RPT, in the current form, seems not to be a good indicator of well-being (Frey – Stutzer 2002; Bruni – Porta 2007; Frey 2008).

In order to make their model work and justify the assumptions that people choose what is best for them, neoclassical economists had to idealize *homo economicus* who was perceived as fully rational. Given the limitations of RPT, one might wonder why economists stick to this descriptively inaccurate theory. Nowadays, many of them argue that neoclassical economics always knew that people are not fully rational and that *homo economicus* is a necessary idealized model of the economic agent, one which is psychologically unrealistic. However, it is not a complete truth. Many economists created models based on the assumptions of full rationality

³However, it is changing and more economists started to see RPT as a normative theory (Hands 2013).



and used them in real life (Rodrik 2015; Thaler 2015). The article's goal is not to blame the economists. Its aim is to underline that the models perform reality, and sometimes, scientists can perceive their models as reality (Callon 2006; Morgan 2012; Boldyrev – Svetlova 2016) which was the case with *homo oeconomicus*.

This is illustrated by Modigliani (1966) who used the assumptions of neoclassical economics (the discounted utility model) to deal with the choices over time. He based his model on an individual's total income. In his theory, 'the life-cycle hypothesis, people are rational and are presumed to make a plan when they are young about how to smooth their consumption over their lifetime. The hypothesis assumes that people not only can make all calculations (with rational expectations) about how long they will live, how much they will make, but also, possess self-control to achieve the optimal plan. We can find a similar idealization concerning choices over time in the works of Friedman (1957) and Barro (1974).

The influence of RPT on reality is even greater. It is possible to make an analogy between contemporary capitalism and RPT. Both assume that people are rational and always know what is best for them. It is enough to give them options, and they will choose in accordance with their autonomous preferences. This means that nobody should interfere with people's choices (e.g., state) and negative freedom is enough (Friedman – Friedman 1980). Moreover, many economists believe that efficiency is value-free, which allows them to "conclude that, *ceteris paribus*, perfectly competitive equilibria are morally desirable and market imperfections that interfere with the achievement of competitive equilibria are morally undesirable" (Hausman – McPherson 2006: 66)⁴.

Neoclassical economists can believe that they are objective scientists who do not decide what is best for people. However, due to the present bias, neoclassical economics based on axiomatic RPT unconsciously supports "short-term human" who frequently falls into temptation and chooses simple pleasures and regrets it later. RPT provides economics with false objectivity (assuming that what people do is the best for them is a normative claim) and has a negative impact on the well-being of individuals.

6. POSSIBILITY OF WEAKNESS OF WILL

The situation, when somebody does something but wants to do something else, is common. After making such a choice, we hear "I know it's irrational, but. . .". As was indicated before, it poses a problem for neoclassical economics which assumes that people have consistent and well-defined preferences and choose what they want. However, it looks more like individuals often have contradictory preferences. This violates the axioms of transitivity and completeness of preferences (WARP). However, in real life, it is not always a problem. Sen writes, "There is nothing particularly schizophrenic in saying: 'I wish I had a vegetarian's tastes, for I disapprove of the killing of animals, but I find vegetarian food so revolting that I can't bear to eat it, so I eat meat'" (1974: 378–379).

Sometimes, discussion about weakness of will relies on stories and anecdotes. In order to properly answer the question whether weakness of will (*akrasia*) exists, we need to dig into the

⁴Due to the paper's limits, the analysis of hidden normativity in economics ends here, but it should be mentioned that this topic has been attracting much attention in the last year (see Ostapiuk 2020).



philosophical discussion. In general, *akrasia* is perceived as doing something against one's better judgment. So, we do something, and we know that it is not the best course of action for us to take. Not only economists were perplexed about this possibility. Socrates denied the existence of weakness of will in *Protagoras*. "No one," he declared, "who either knows or believes that there is another possible course of action, better than the one he is following, will ever continue on his present course" (Plato 1961: 348). In the 20th century, Hare agreed with Socrates and argued that if we evaluate something to be the best action for us then we must do it. This is because evaluative judgments are connected with a prescription (Hare 1952, 1963). Therefore, when one judges that, the best action is to quit smoking, then she must quit smoking.

Often, we hear people say that they know that quitting smoking would be the best action, but they cannot do it. For Hare, as well as for Socrates, it does not mean that weakness of will exists. In this case, one cannot quit smoking because deep down one still wants to smoke and this is one's strongest desire. We can find an analogy between this line of reasoning and neoclassical economics. As was indicated before, for Becker, declarations do not mean anything; they are only cheap talks. If one really wants to quit smoking, one will do it. Thanks to this line of reasoning, present bias is not a problem in RPT. It is not treated as an irrational feature of behaviour which proves that people do not do what they really want. They are just overpowered by the present desires, and deep down, they want them. The tautological (axiomatic) construction of neoclassical economics, in which rationality is perceived instrumentally, enables economists to deny the existence of weakness of will and still be consistent. Despite the approaches where weakness of will is impossible by definition, for many people it is a real feeling. If we always do what we want deep down, then how so many of us regret after choice? The contemporary debate about this phenomenon was started anew by Davidson (1980).

Davidson thinks that weakness of will is possible and he is mostly known for defining akratic action. In doing x , an agent acts akratic if and only if: (1) she does x intentionally; (2) she believes that there is an alternative action y open to him, and this guarantees that the action is free; (3) she judges that, all things considered, it would be better to do y than to do x (Davidson 1980). It is important to underline the phrase "all things considered." It does not mean that if one thinks that an apple is healthier than chips, it is one's evaluation. Here, one only sticks to one set of values (health) versus another (taste). "All things considered" means comparing a different set of values.

Nowadays, most philosophers who are interested in the weakness of will agree with Davidson that akratic action is not only possible but also actual, and a common feature of people's lives. Moreover, most philosophers admit that *akrasia*, if possible, must be irrational (Stroud – Tappolet 2003). Here, it is crucial to underline the distinction between economics and philosophy. As was indicated before, neoclassical economists treat rationality instrumentally, whereas many philosophers treat it as something worth striving for. Moreover, neoclassical economists focus on choices, whereas philosophers distinguish between belief and choice where belief is beyond the revealed choice. The differences in the perception of rationality result in the situation where philosophers attempt to define what rational behaviour might consist in, while economists do not.⁵

⁵Some economists define rationality (e.g., ecological rationality by V. Smith, procedural rationality by H. Simon). However, instrumental rationality dominates in economics (Berg-Gigerenzer 2010).



With this short introduction, we can analyze whether *akrasia* is rational or not. As was stated before, it refers to the behaviour where one chooses to do something that one believes not to be the best option. The belief that a certain option is not a good thing to do is either rational for one to hold or not. When it is not a rational belief to hold, then one is irrational in one's belief. For example, Martha is dissatisfied with her vacation in Spain and believes that she should have gone to Italy instead. However, she is mistaken and the vacation in Italy would be much worse. Then, her belief that Italy is a better place for a vacation is false and it is irrational to hold it. On the other hand, if her belief that Italy is a better place for a vacation than Spain is true, then she has a rational belief to hold. However, in this case, it cannot be rational for her to go to Spain. Therefore, *akrasia* necessarily involves irrationality, either in her beliefs or in her choices (Wedgwood 2003).

Although many philosophers agree that *akrasia* must be irrational, they argue whether it is possible intentionally and freely to act against one's better judgment. For example, Watson (1977) argues that akratic action does not exist as acting against one's better judgment is not free. It is so because there is no clear difference between *akrasia* and compulsion in situations when an agent is motivated by emotion or desires that one cannot resist. Watson considers two possible explanations for why people fail to resist a desire that they do not want to have, and he rejects both. People's failure cannot be explained in terms of what they choose to do because a choice must follow better judgment (Hare 1952, 1963). In addition, it cannot be explained by an insufficient attempt to resist. If the action is free, we must assume that an agent can control oneself, and here arises the question of why one did not try hard enough. Watson concludes that the agent was unable to resist the rebellious desire, and in both cases, one's action was unfree. However, many philosophers argue that it is possible to distinguish between weak-willed and compelled actions and Watson's conclusion is controversial.⁶

After setting the stage for weakness of will, we move to the most important goal of this paper – providing a framework to decide between selves. Holton (1999, 2003) argues for a position similar to Schelling's. According to Holton, weakness of will means that a certain kind of intention – a resolution – was violated. The term “resolution” refers to a plan or intention one has to hold in the face of potential temptations. Thus, we can identify resolution with pre-commitment strategies. Holton thinks that “central to the idea of weakness of will is an over-readiness to abandon one's resolutions” (2003: 41). Therefore, we can know that people are weak-willed when they try to fight with some unwanted desires by sticking to their resolutions. For example, Janet needs to study for the forthcoming exam, so she dismissed her friends' proposal to go to the pub in the evening because she wants to study (resolution). However, after 30 min of studying, she is tired and watches a movie instead. In the end, she failed to study, but at least she tried to do so and was able to dismiss her friends' invitation. According to Holton, such behaviour indicates weakness of will.

Moreover, he makes a distinction between weakness of will (overriding one's resolution) and akratic action (overriding one's judgment), which was criticized by some philosophers (e.g., Mele 2010) but is very useful for this paper. Holton writes:

“Imagine someone who is convinced that all the arguments point the same way: he should give up meat. Yet suppose he does not. He is, therefore, akratic. But is he weak-willed? That, I contend,

⁶More discussion is provided in the works of Kennett (2001), Mele (2010) and Bermúdez (2018).



depends on other factors. Has he repeatedly vowed to give up, only to find himself succumbing time and again in the face of rare steaks and slow-cooked offal? Or does he unblushingly affirm while conceding practical inconsistency that he has never had any intention of giving up, and never will? In the first case we surely would accuse him of weakness of will; in the second I think we would not.” (Holton 2003: 52)

This distinction is the same as Schelling’s when he argued that we should help only those who act against desires they do not want to have (by precommitment). Holton and Schelling consider that it is not enough if one says that one does not want to smoke; one needs to try to quit smoking. Otherwise, it can only be a cheap talk. The distinction between *akrasia* and weakness of will is useful because in the case of *akrasia* we do not know whether somebody really wants to quit smoking or only says so. When one acts against unwanted desires and forms a resolution, it indicates that this person really wants something else. It also provides a better framework than a vague one used by many behavioural economists who rely on New Year’s resolution test where people’s declarations are treated as their real preferences (criticism: Rizzo 2016; Sugden 2018).

Holton uses the idea of resolution to explain why some philosophers could not accept that weakness of will exists. This is because they use the Humean account (belief – desire account) where all intentional actions are explained in terms of an agent’s beliefs and desires. The agent simply acts on the strongest desire and resolutions are reducible to beliefs and desires (Holton 2003). Therefore, when someone tries to quit smoking but fails to do so, it means that deep down one still wants to smoke. Intentions have the same value as desires because in the end “reason is, and ought only to be the slave of the passions” (Hume 1896: 213). This reasoning is very close to neoclassical economics due to instrumental rationality where preferences are given and people satisfy them. Holton is a supporter of the “Augmented Humean account” (belief – desire – intention account) where beliefs and desires are not deemed enough. He writes, “Intentions, of which resolutions are a species, should be seen as a third mental kind, irreducible to the other two” (Holton 2003: 51). This approach works similarly to the Humean account in terms of the end result. Therefore, “If a resolution is stronger than any contrary desires, the agent will stick to it; if the contrary desires are stronger, then the agent will act on them instead” (Holton 2003: 51). However, the qualitative difference between desires and intentions makes weakness of will possible, and also, explains why people regret after following unwanted desires. In the context of economics, in some cases, intentions can be perceived as people’s real preferences.

It is necessary to make a qualitative distinction between desires and intentions in relation to the intertemporal choices when we need to choose which self should decide normatively. This cannot be done by utilitarian measurement which is used by the neoclassical economists where desires and values are compared within the same framework as in the Humean account. The limitations of this approach are visible in the example of Mill who tried to save utilitarianism from criticism. He differentiated higher pleasures from the lower ones with regard to quality and argued that “some kinds of pleasure are more desirable and more valuable than others” (Mill 1863: 11). However, people choose “lower” pleasures. Sometimes, they prefer lying on the couch and watching comedy to reading Shakespeare. At first sight, it contradicts Mill’s belief that people want higher pleasures. He answers this criticism by explaining the difference between human choosing in the short-term and the long-term perspective. People prefer lying on the couch and watching a comedy because it gives them more pleasure; however, they value



Shakespeare more. Mill considers that the true choice is in the long-term perspective. If somebody asks people what they want to do in their whole life, the majority will choose Shakespeare. Mill explains that we choose the “lower” pleasures even if we want the higher ones as “occasionally, under the influence of temptation, postpone them to the lower” (Mill 1863: 14). In other words, we know the difference between comedy and Shakespeare, but sometimes we succumb to our weaknesses. We find *Hamlet* a great play, not because it gives us the most pleasure, but because it challenges us and makes us better.

In this context, Mill departed from utilitarianism where everything can be compared by one all-encompassing notion of utility because he assumed that more worthy deeds exist. His credit for the higher needs of humans leads to the problems of neoclassical economics which relies on RPT. Economists do not like declarations because often they do not coincide with the observable choices (Beshears et al. 2008). For example, when people are asked which programs they want to watch on TV, they point out performances or theater, but they watch soap operas when it comes to choose. However, it does not mean that people are hypocrites. As was stated earlier, weakness of will results in situations when people do not achieve what they really want.

In order to choose what people should do, it is necessary to know the preferences that are worth following. Many years ago, neoclassical economists abandoned this endeavour as they did not want to be entangled in this normative discussion. However, due to the present bias, they do not have the privilege not to choose between selves and people’s goals.

7. REASON VERSUS EMOTIONS

In the end, we can ask why people should follow commitments (reason) and not short-term pleasures (emotions).⁷ This assertion is controversial in neoclassical economics which aspires to be value-free. Therefore, economists do not decide what is best for people and perceive rationality instrumentally. However, as was indicated earlier, this position is only seemingly value-free. When economists do not decide what is best for people and assume that people choose what is best for them, they support “short-term human.” Therefore, economists cannot escape from choosing between selves (not choosing is also a choice, as Jean-Paul Sartre said). The discussion in this section allows deciding which self should be in control: the self that follows short-term pleasures or the one that has long-term goals? Of course, the topic of reason versus emotions is too large to be analyzed thoroughly here. Therefore, we focus on autonomy and free will in the context of neoclassical economics where these qualities are taken for granted because economists presuppose the existence of individuals (Davis 1995, 2003, 2010).

One clarification is needed to explain why the discussion in this section is based on reason versus emotion. In this article, desires are identified to some degree with preferences. In general, economists perceive preferences as desires constrained by time and budget (Hausman 2012). However, here we focus on short-term preferences which are very similar to desires because people irrationally focus on the present moment. At some level, we can identify preferences with desires as both are given and people cannot fully choose them.

The discussion between emotions and reason started with philosophy itself. We deal only with Kant and Hume because they are essential for this paper. The first one is identified with

⁷The author uses Plato’s famous dichotomy between reason and desires.



reason and can be perceived as an unconscious supporter of goal-oriented economics in which human goals and values are discussed. On the other hand, Hume makes a case for emotions and can be perceived as the father of neoclassical economics which is value-free because it is not entangled in the discussion on goals. Hume is a champion of instrumental rationality which is also cherished by neoclassical economists. He writes, “reason is, and ought only to be the slave of the passions” (1896: 283). According to him, passions in themselves cannot be called unreasonable because:

“A passion must be accompany’d with some false judgment, in order to its being unreasonable; and even then ‘tis not the passion, properly speaking, which is unreasonable, but the judgment [...] a passion can never, in any sense, be call’d unreasonable, but when founded on a false supposition, or when it chuses means insufficient for the design’d end.” (Hume 1896: 284)

Hume means that for an act to be irrational, the desire which was a cause of the action must be founded on a false belief or be ineffective (not to use proper means to achieve the end). Therefore, we can say that a drunk is irrational not because he is drunk but because he drinks apple juice instead of vodka. Effectiveness is the only way for neoclassical economics to judge if some action is irrational or not.

In the instrumental view of rationality, the action is finally determined by psychological states. Thus, an instrumental theory of choice presents a casual explanation of human action which has the same structure as the explanation found in the natural sciences.⁸ Kant understood that we need to accept this approach in order to explain the choices made by people because they are part of the physical world, and we need to perceive that their actions have physical explanations (Sugden 1991). However, Kant thinks that when we reason we have to perceive ourselves as autonomous. We must conceive ourselves as capable of forming beliefs and reach conclusions that are not determined by external causes. Therefore, when people reason what actions to take, they must perceive themselves as capable of determining their actions. This account differs from neoclassical economics where self is a bundle of preferences that are given and people just act upon them (Davis 1995). Kant argues that a reason has to regard itself as the creator of its principles which are independent of the alien influences. From this perspective, Humean passions and desires can be perceived as alien influences because it is impossible to reason coherently about what to do when we think that our choices are fully determined by our psychological states (economics as “closed system”; Lawson 1997). However, Kant’s goal is not to prove that people are really autonomous. He argues that we cannot engage in any form of reasoning without presupposing our autonomy.

The second feature of Kantian autonomy is that an autonomous individual is the one whose actions are governed by the laws that one imposed on oneself. Kant agrees with Humean instrumental rationality and perceives it as the source of hypothetical imperative (if you want to achieve X, then do Z). However, there is another form of rationality which refers to categorical imperatives (do X, regardless of your wants). From this perspective, a choice is rational if it is prescribed by some principle which can be a universal law for all rational agents (Sugden 1991). There is a qualitative difference between reason and instrumental rationality. In the end,

⁸However, causal account in RPT is widely criticized (Ross 2005; Bimmore 2009; Gul – Pesendorfer 2008).



Kantian categorical imperative cannot be embraced by the traditional economic framework. The limitations of RPT in the case of ethics have been presented by [White \(2011\)](#).

According to Kant, reason is strictly connected with freedom and autonomy. Why can emotions not be perceived in the same way? Can we even assess whether emotions are rational or irrational?⁹ A negative answer comes to these questions from two premises: first, emotions being predominately involuntary cannot be perceived as actions; second, only freely chosen actions can be assessed as rational or irrational ([Elster 1996](#)). Each of these premises was questioned, but the explanations given are far from being convincing. The first premise (emotions are involuntary) was questioned by the authors arguing that emotions are chosen by people ([Schafer 1976](#); [Solomon 1980](#); [Sartre 2015](#)). They believe that emotions are actions in the same way as mental calculations are actions. However, Elster commonsensically objects to these theories. He writes, “If we can choose our emotions, why do we not choose to be happy all the time? And why would anyone ever choose to be sad?” ([Elster 1996](#): 8). The second premise seems to be at odds with the idea of rational belief. People form beliefs but do not choose them. Actually, it is widely accepted that the idea of deciding to believe is self-contradictory ([Williams 1970](#)).

We have presented the argument that acting upon desires and given preferences is not enough to establish freedom. Now, we move toward two contemporary philosophers (Frankfurt and Dworkin) who argue that autonomy comes from the reflection about desires and who try to establish a hierarchical self where reason and commitments should be protected. Both arguments are essential to support the conception of “long-term human” presented in this paper.

[Frankfurt \(1971\)](#) argued that emotions (desires) do not make us free. He called the condition when people are tossed by their desires as “wanton.” Its predominant feature is that it does not care about people’s will. The desires move people to do a particular thing without asking them either they want to be moved by those desires or they prefer to be moved by others. Wanton is the feature of young children and animals. We do not perceive wanton as having freedom of the will, even though we recognize that animals can be free to run in whatever direction they want. Therefore, Frankfurt stated that “having the freedom to do what one wants to do is not a sufficient condition of having a free will” ([1971](#): 9). Here, Frankfurt distinguished between being free to do what one wants and being free to decide what one wants. These concepts are completely different:

“A person who is free to do what he wants to do may yet not be in a position to have the will he wants. Suppose, however, that he enjoys both freedom of action and freedom of the will. Then he is not only free to do what he wants to do; he is also free to want what he wants to want.” ([Frankfurt 1971](#): 14)

To explain this difference, Frankfurt presented the conception of the second-order desires. This can be described using the previously analyzed case of cigarettes. Having the choice to take or not to take cigarettes, the addicted people will choose cigarettes, which is their first-order preference. However, if they had the choice, they would rather rank their preferences as no cigarettes > cigarettes than cigarettes > no cigarettes. According to Frankfurt, the capacity to have second-order desires (“metapreferences”; [Sen 1977](#)) distinguishes humans from other

⁹The discussion on rationality of emotions is extensive and could not be presented here (see, [de Sousa 1990](#)).



forms of life, which differentiates his account from Becker's rational addict who incorporates different desires in the moment of choice and performs a cost-benefit calculation. Moreover, [Frankfurt \(1971\)](#) considers that free will should be perceived as the ability to choose according to one's second-order desires (metapreferences). It means the potential to construct the preferences that one would prefer to have.

Philosophers frequently use metapreferences and the idea of hierarchical self in debates on personal autonomy. Frankfurt's conception of second-order desires is much closer to Kant (goal-oriented economics) than Hume (value-free economics). Although individuals have preferences (first-order desire) as in Humean approach and neoclassical economics, they can reason and reflect on their preferences, which is where the "true" self lies. However, the metapreference approaches are not as strict as Kant's because they allow the influence of society in shaping values ([Sen 1973](#); more on differences between Kant and Sen: [Paul-Struder 2006](#)).

In the 1970s, Frankfurt and Dworkin along with other philosophers started the contemporary discussion on personal autonomy. Their goal was to decide what kind of individual freedom should be protected and how it can be presented from the perspective of contemporary conceptions of free will. They concluded that people's actions are worth protecting if they are ingrained in values, commitments and objectives that people have. On the other hand, their actions are not worth protecting if they go against those values, objectives and commitments. Both Frankfurt and Dworkin use the hierarchical models when the agent is autonomous only if her first-order desire is accepted by a second-order desire which is identified with volition ([Frankfurt 1988](#): 12–25). This approach is neutral when considering the origins of higher-order desires. It does not exclude the desires and values that are socially constructed. The cause of such desires does not matter, but only the agent's identification with them ([Frankfurt 1988](#): 53). Autonomy concerns our ability to examine and ask whether we do identify with our desires or whether we may wish to override them. "We," in this case, is established by our higher-order preferences. Dworkin, like Frankfurt, speaks of them as the agent's "true self" ([Dworkin 1988](#): 59).

Identifying freedom with reason and reflection on desires is a common theme in philosophy. In general, freedom refers to the self-determination of a subject possessing both reason and will ([Ballet et al. 2013](#)). More fundamentally, it indicates the choice of a particular way of life that reflects one's values and commitments. It is a different account than negative freedom used in neoclassical economics, where it is enough if people act upon their preferences. Therefore, freedom presented in this article means choosing values and commitments reflecting the life one is living. Self-determination means not only the choice as in neoclassical economics (revealed preference) but also the capacity to choose. In other words, it refers to the capacity to evaluate and reevaluate one's choices which can be identified with precommitment strategies and intentions. The conclusion from the discussion is that the self that should decide is connected with reason and commitments, and not with desires as unintentionally supported in RPT.

8. CONCLUSIONS

This article demonstrated that neoclassical economics cannot admit that weakness of will exists due to the axiomatic assumptions of RPT. However, these assumptions lead to many problems. Their main concern is choice over time because present bias shows that the consistency



assumption is unrealistic. This argument is well established in the literature. However, it is not sufficient to dismiss the axiomatic assumptions of RPT. The example of Becker's rational addict has been presented to show how the descriptive criticism concerning intertemporal choices is absorbed and economists can argue that people always choose what is best for them. The reason for this is a tautological understanding of rationality and utility. This paper presents the problems with these assumptions on the example of weakness of will. Denying the existence of this phenomenon leads to substantial problems in real life (economics' conception of well-being).

The philosophical discussion on the weakness of will was necessary to establish a distinction between desires (first-order desires/short-term preferences) and intentions (second-order desires/long-term preferences). Precommitment strategies allow finding what preferences individuals really want to have. This knowledge is essential in the context of intertemporal choices because comparing them by the utility measurements seems very problematic. The framework distinguishing between intentions and desires should help economists decide which self should be in control. The conclusion is that people should follow reason and commitments because they support freedom and autonomy. However, it is more a methodological framework than a justification for some behavioural economists who want to nudge people toward "purified preferences" which can be discovered thanks to New Year's resolution test (Thaler – Sunstein 2008: 6,73). Despite having a methodological framework, it is hard to distinguish between real commitments and cheap talk. Hence, we should take some declarations with a pinch of salt (Bryan et al. 2010; Laibson 2015; Sugden 2018).

This paper is intended to be a new impulse for multiple self-conception which had not gained much popularity, precisely because it did not have a framework that enables economists to decide between selves. Moreover, the demonstration of the qualitative difference between desires and intentions can convince economists not to focus solely on the individuals' choices. However, the take-home message of this paper is not that we should not rely on RPT because people are not rational and choices are context-dependent as the behavioural economists argue. The message is that economists should not shy away from discussing human goals and cannot treat RPT as a value-free framework. It is impossible not to decide which self should be in control.

ACKNOWLEDGEMENT

The project is financed by the Ministry of Science and Higher Education in Poland under the programme "Regional Initiative of Excellence" 2019–2022, project number 015/RID/2018/19, total funding amount 10 721 040,00 PLN.

REFERENCES

- Ainslie, G. (1975): Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control. *Psychological Bulletin*, 82(4): 463–496.
- Ainslie, G. (2001): *Breakdown of Will*. New York: Cambridge University Press.
- Andreoni, J. – Sprenger, C. (2012): Estimating Time Preferences from Convex Budgets. *American Economic Review*, 102(7): 333–356.



- Angner, E. (2018): What Preferences Really Are. *Philosophy of Science*, 85(4): 660–681.
- Ballet, J. – Bazin, D. – Dubois, J. – Mahieu, F. (2013): *Freedom, Responsibility and Economics of the Person*. London: Routledge.
- Barro, R. J. (1974): Are Government Bonds Net Wealth? *Journal of Political Economy*, 82(6): 1095–1117.
- Becker, G. – Murphy, K. (1988): A Theory of Rational Addiction. *Journal of Political Economy*, 96(4): 675–700.
- Berg, N. – Gigerenzer, G. (2010): As-If Behavioral Economics: Neoclassical Economics in Disguise? *History of Economic Ideas*, 18(1): 133–165.
- Bermúdez, J. L. (ed.) (2018): *Self-Control, Decision Theory, and Rationality: New Essays*. Cambridge University Press.
- Beshears, J. – Choi, J. J. – Laibson, D. – Madrian, B. C. (2008): How Are Preferences Revealed? *Journal of Public Economics*, 92(8–9): 1787–1794.
- Binmore, K. (2009): *Rational Decisions*. Princeton University Press.
- Boland, L. A. (1981): On the Futility of Criticizing the Neoclassical Maximization Hypothesis. *American Economic Review*, 71(5): 1031–1036.
- Boldyrev, I. – Svetlova, E. (eds) (2016): *Enacting Dismal Science: New Perspectives on the Performativity of Economics*. New York: Palgrave Macmillan.
- Brickman, P. – Coates, D. – Janoff-Bulman, R. (1978): Lottery Winners and Accident Victims: Is Happiness Relative? *Journal of Personality and Social Psychology*, 36(8): 917–927.
- Bruni, L. – Porta, P. (eds) (2007): *Handbook on the Economics of Happiness*. Cheltenham: Edward Elgar.
- Bruni, L. – Sugden, R. (2007): The Road Not Taken: How Psychology Was Removed from Economics, and How It Might Be Brought Back. *The Economic Journal*, 117(516): 146–173.
- Bryan, G. – Karlan, D. – Nelson, S. (2010): Commitment Devices. *Annual Review of Economics*, 2(1): 671–698.
- Callon, M. (2006): *What Does It Mean to Say that Economics Is Performative?* CSI Working Papers Series, No. 005. Centre De Sociologie De L'innovation (CSI), Mines Paristech.
- Cohen, J. D. – Ericson, K. – Laibson, D. – White, J. M. (2020): Measuring Time Preferences. *Journal of Economic Literature*, 58(2): 299–347.
- Davidson, D. (1980): How Is Weakness of the Will Possible? In: Davidson, D. (ed): *Essays on Actions and Events*. Oxford: Clarendon Press, pp. 21–42.
- Davis, J. B. (1995): Personal Identity and Standard Economic Theory. *Journal of Economic Methodology*, 2(1): 35–52.
- Davis, J. B. (2003): *The Theory of the Individual in Economics: Identity and Value*. London: Routledge.
- Davis, J. B. (2010): *Individuals and Identity in Economics*. Cambridge: Cambridge University Press.
- De Sousa, R. (1990): *The Rationality of Emotion*. Cambridge, MA: The MIT Press.
- Dworkin, G. (1988): *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Elster, J. (ed.) (1987): *The Multiple Self*. Cambridge: Cambridge University Press.
- Elster, J. (1996): Rationality and the Emotions. *The Economic Journal*, 106(438): 1386–1397.
- Ericson, K. M. – Laibson, D. (2019): Intertemporal Choice. In: Bernheim, D. – DellaVigna, D. – Laibson, D. (eds): *Handbook of Behavioral Economics: Applications and Foundations 2*. Amsterdam: North-Holland, pp. 1–67.
- Frankfurt, H. G. (1971): Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68(1): 5–20.
- Frankfurt, H. G. (1988): *The Importance of What We Care About*. Cambridge: Cambridge University Press.



- Frederick, S. – Loewenstein, G. – O'Donoghue, T. (2002): Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature*, 40(2): 351–401.
- Frey, B. S. (2008): *Happiness: A Revolution in Economics*. Cambridge, MA: The MIT Press.
- Frey, B. S. – Stutzer, A. (2002): What Can Economists Learn from Happiness Research? *Journal of Economic Literature*, 40(2): 402–435.
- Friedman, M. (1957): The Permanent Income Hypothesis. In: *A Theory of the Consumption Function*. Princeton University Press, pp. 20–37.
- Friedman, M. – Friedman, R. (1980): *Free to Choose: A Personal Statement*. Boston: Houghton Mifflin Harcourt.
- Fudenberg, D. – Levine, D. K. (2006): A Dual-Self Model of Impulse Control. *The American Economic Review*, 96(5): 1449–1476.
- Fudenberg, D. – Levine, D. K. (2012): Timing and Self-Control. *Econometrica*, 80(1): 1–42.
- Gilbert, D. (2007): *Stumbling on Happiness*. Vintage.
- Gul, F. – Pesendorfer, W. (2001): Temptation and Self-Control. *Econometrica*, 69(6): 1403–1435.
- Gul, F. – Pesendorfer, W. (2004): Self-Control and the Theory of Consumption. *Econometrica*, 72(1): 119–158.
- Gul, F. – Pesendorfer, W. (2008): The Case for Mindless Economics. In: Caplin, A. – Schotter, A. (eds): *The Foundations of Positive and Normative Economics: A Handbook*. Oxford University Press, pp. 3–39.
- Hands, D. W. (2013): Foundations of Contemporary Revealed Preference Theory. *Erkenntnis*, 78(5): 1081–1108.
- Hands, D. W. (2014): Paul Samuelson and Revealed Preference Theory. *History of Political Economy*, 46(1): 85–116.
- Hands, D. W. (2015). *Normative Rational Choice Theory: Past, Present, and Future*. <https://ssrn.com/abstract=1738671>.
- Hare, R. M. (1952): *The Language of Morals*. Oxford: Clarendon Press.
- Hare, R. M. (1963): *Freedom and Reason*. Oxford: Clarendon Press.
- Hausman, D. M. (1992): *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hausman, D. M. (2008): Mindless or Mindful Economics: A Methodological Evaluation. In: Caplin, A. – Schotter, A. (eds): *The Foundations of Positive and Normative Economics: A Handbook*. Oxford University Press, pp.125–151.
- Hausman, D. M. (2012): *Preference, Value, Choice, and Welfare*. Cambridge, New York: Cambridge University Press.
- Hausman, D. M. – McPherson, M. S. (2006): *Economic Analysis, Moral Philosophy and Public Policy*. Cambridge: Cambridge University Press.
- Hédoine, C. (2016): Sen's Criticism of Revealed Preference Theory and Its 'Neo-Samuelsonian Critique': A Methodological and Theoretical Assessment. *Journal of Economic Methodology*, 23(4): 349–373.
- Herfeld, C. (2020): The Diversity of Rational Choice Theory: A Review Note. *Topoi*, 39(2): 329–347.
- Hodgson, G. M. (2012): On the Limits of Rational Choice Theory. *Economic Thought*, 1: 94–108.
- Holton, R. (1999): Intention and Weakness of Will. *The Journal of Philosophy*, 96(5): 241–262.
- Holton, R. (2003): How Is Strength of Will Possible. In: Stroud, S. – Tappolet, Ch. (eds): *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press, pp. 39–67.
- Hume, D. (1896): *Treatise of Human Nature*. Oxford: Clarendon Press.
- Jevons, W. S. (1911): *The Theory of Political Economy*. (Fourth edition) London: Macmillan.



- Kahneman, D. – Tversky, A. (1979): Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2): 263–291.
- Kennet, J. (2001): *Agency and Responsibility: A Common-Sense Moral Psychology*. Oxford: Clarendon Press.
- Laibson, D. (1997): Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics*, 112(2): 443–477.
- Laibson, D. (2015): Why Don't Present-Biased Agents Make Commitments? *American Economic Review*, 105(5): 267–272.
- Lawson, T. (1997): *Economics & Reality*. Abingdon: Routledge.
- Loewenstein, G. – Elster, J. (1992): *Choice Over Time*. New York: Russell Sage Foundation.
- Loewenstein, G. – O'Donoghue, T. (2004): *Animal Spirits: Affective and Deliberative Processes in Economic Behavior*. CAE Working Paper, No. 04–14, Carnegie Mellon University. <https://Cae.Economics.Cornell.Edu/04-14pdf>.
- Mele, A. R. (2010): Weakness of Will and Akrasia. *Philosophical Studies*, 150(3): 391–404.
- Mill, J. S. (1863): *Utilitarianism*. London: Parker, Son and Bourn.
- Modigliani, F. (1966): The Life Cycle Hypothesis of Saving, the Demand for Wealth and the Supply of Capital. *Social Research*, 33(2): 160–217.
- Morgan, M. S. (2012): *The World in the Model: How Economists Work and Think*. Cambridge, New York: Cambridge University Press.
- Ostapiuk, A. (2019): Human Now versus Human over Time. When Instrumental Rationality and Utility Are Not Enough. *Panoeconomicus*, 66(5): 633–657.
- Ostapiuk, A. (2020): *The Eclipse of Value-Free Economics. The Concept of Multiple Self versus Homo Economicus*. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- O'Donoghue, T. – Rabin, M. (1999): Doing It Now or Later. *American Economic Review*, 89(1): 103–124.
- O'Donoghue, T. – Rabin, M. (2001): Choice and Procrastination. *The Quarterly Journal of Economics*, 116(1): 121–160.
- Pauer-Studer, H. (2006): Identity, Commitment and Morality. *Journal of Economic Methodology*, 13(3): 349–369.
- Plato (1961): Protagoras. In: Hamilton, E. – Huntington, C. (eds): *The Collected Dialogues of Plato*. Princeton: Princeton University Press, pp. 308–352.
- Reiss, J. (2013): *Philosophy of Economics: A Contemporary Introduction*. Routledge.
- Rizzo, M. J. (2016): Behavioral Economics and Deficient Willpower: Searching for Akrasia. <https://Ssrn.Com/Abstract=2731818>.
- Robbins, L. (1932): *An Essay on the Nature & Significance of Economic Science*. London: Macmillan.
- Rodrik, D. (2015): *Economics Rules: The Rights and Wrongs of the Dismal Science*. W. W. Norton & Company.
- Ross, D. (2005): *Economic Theory and Cognitive Science: Microexplanation*. Cambridge, MA: MIT Press.
- Samuelson, P. A. (1937): A Note on Measurement of Utility. *The Review of Economic Studies*, 4(2): 155–161.
- Samuelson, P. A. (1938): A Note on the Pure Theory of Consumer's Behaviour. *Economica*, 5(17): 61–71.
- Sartre, J. (2015): *Sketch for a Theory of the Emotions*. London: Routledge.
- Schafer, R. (1976): *A New Language for Psychoanalysis*. New Haven: Yale University Press.
- Schelling, T. C. (1980): The Intimate Contest for Self-Command. *Public Interest*, 60: 94–116.
- Schelling, T. C. (1984): Self-Command in Practice, in Policy, and in a Theory of Rational Choice. *The American Economic Review*, 74(2): 1–11.
- Sen, A. K. (1973): Behaviour and the Concept of Preference. *Economica*, 40(159): 241–259.



- Sen, A. K. (1974): Choice, Orderings, and Morality. In: Korner, S. (ed): *Practical Reason*. Oxford: Blackwell, pp. 54–67.
- Sen, A. K. (1977): Rational Fools: A Critique of the Behavioural Foundations of Economic Theory. *Philosophy and Public Affairs*, 6(4): 317–344.
- Sen, A. K. (1993): Internal Consistency of Choice. *Econometrica*, 61(3): 495–521.
- Solomon, R. C. (1980): Emotions and Choice. In: Rorty, A. (ed): *Explaining the Emotions*. Berkeley, Los Angeles: University of California Press, pp. 251–282.
- Strotz, R. H. (1955): Myopic and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies*, 23(3): 165–180.
- Stroud, S. – Tappolet, Ch. (eds) (2003): *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press.
- Sugden, R. (1991): Rational Choice: A Survey of Contributions from Economics and Philosophy. *The Economic Journal*, 101(407): 751–785.
- Sugden, R. (2018): *The Community of Advantage: A Behavioural Economist's Defence of the Market*. New York, Oxford: Oxford University Press.
- Thaler, R. H. (1980): Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization*, 1(1): 39–60.
- Thaler, R. H. (2015): *Misbehaving: The Making of Behavioral Economics*. New York: W. W. Norton & Company.
- Thaler, R. H. – Shefrin, H. M. (1981): An Economic Theory of Self-Control. *Journal of Political Economy*, 89(2): 392–406.
- Thaler, R. – Sunstein, C. (2008): *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- Watson, G. (1977): Skepticism about Weakness of Will. *Philosophical Review*, 86(3): 316–339.
- Wedgwood, R. (2003): Choosing Rationally and Choosing Correctly. In: Stroud, S. – Tappolet, Ch. (eds): *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press, pp. 201–229.
- White, M. (2011): *Kantian Ethics and Economics: Autonomy, Dignity, and Character*. Stanford: Stanford University Press.
- Williams, B. A. O. (1970): *Problems of the Self*. Cambridge: Cambridge University Press.
- Wong, S. (2006): *Foundations of Paul Samuelson's Revealed Preference Theory: A Study by the Method of Rational Reconstruction*. London: Routledge.

