

*Tomasz Żuradzki*

Institute of Philosophy, Jagiellonian University

# **Time-biases and Rationality: The Philosophical Perspectives on Empirical Research about Time Preferences**

## **Introduction**

The empirically documented fact is that people's preferences are time-biased. The main aim of this paper is to analyse in which sense (if any) do time-biases violate the requirements of rationality, as many authors assume. I will demonstrate that contrary to many influential views in psychology, economy and philosophy it is very difficult to find why the bias toward the near violates the requirements of rationality. I will also show why the bias toward the future violates the requirements of rationality in a very basic sense.

We prefer pains to be in the distant future rather than in the immediate future, and pleasures to be in the immediate future rather than in the distant future (the bias toward the near). We also prefer pains to be in the past rather than in the future and pleasures to be in the future rather than in the past (the bias toward the future). On the one hand, it can be argued that these phenomena of human psychology are similar to some other well-known preferences that are explicable in evolutionary terms, e.g. partiality towards immediate kin or partiality towards people living in the same community. If so they could be taken into account by normative moral theories and could affect the scope of our moral and prudential obligations, in particular the way in which we weigh our future interests and well-being against our pre-

sent interests and well-being. However, on the other hand, it seems that some forms of time-biases violate the requirements of rationality. Therefore it can be claimed that we should get rid of these biases as irrational when determining the scope of moral or prudential obligations in the intertemporal cases.

The objective of the article is to analyse some of the most important recent findings in empirical economy and psychology that document different kinds of time-biased preferences with the help of the analytical tools of philosophy, in particular the recent discussions about the nature of the requirements of rationality. I also would like to evaluate the possible significance of these empirical findings for normative theories. The premise of the article is that moral norms are grounded in some features of human nature and reflect the specific human sensitivity and therefore findings in empirical economy and psychology that document the different kinds of time-biases are important for normative theories and can influence the shape and the scope of our moral or prudential obligations.

In the first section I discuss a representative sample of philosophical views on time-biases; in the second I make some clarifications. The third section analyses different evidence from empirical psychology and economy about time-biases; the fourth tries to answer the question of why is it commonly thought that rationality does not permit us to be time-biased. In the fifth section I explain how present day normative orders (i.e. legal and economic regulations) have emerged in the context of different forms of the time-biased preferences of individuals.

## **1. Philosophers on time-biases: Hume, Sidgwick, Rawls, Parfit**

Many philosophers claim that there is something wrong with preferences that have no other reason than time. As we will see, some of them suggest that these kind of preferences are irrational and that tem-

poral neutrality is an important requirement of rationality. According to many philosophers, temporal neutrality demands that agents attach no significance to the temporal location of goods within someone's life. Despite the apparent novelty of experimental research on time biases, the conceptual analysis of this phenomenon is not new. The sources of this view can be found in David Hume's *A Treatise on Human Nature*:

In reflecting upon any action which I am to perform a twelvemonth hence, I always resolve to prefer the greater good, whether at that time it will be contiguous or remote. My distance from the final determination makes all those minute differences vanish, nor am I affected by anything, but the general and more discernible qualities of good and evil. But on my nearer approach, those circumstances, which I at first over-looked, begin to appear, and have an influence on my conduct and affections. A new inclination to the present good springs up, and makes it difficult for me to adhere inflexibly to my first purpose and resolution.<sup>1</sup>

In this chapter of *A treatise...* Hume discusses in which way people are governed by their own interests. Although in the first sentence of this chapter he claims: "Nothing is more certain, than that men are, in a great measure, governed by interest", he also tries to understand why people sometimes violate the rule of self-interest. He maintains that if they reflect from a temporal distance they govern themselves only by the real qualities of things and the gains they can expect from them. He assumes that value judgments made from a temporal distance are more reliable, whereas when "an inclination to the present good" appears, their judgments very often become disturbed. Hume calls this bias toward the near "natural weakness" or "natural infirmity".

---

<sup>1</sup> D. Hume, *A treatise of human nature*, Courier Dover Publications, New York 2003 (1739), III, II, 7.

Adam Smith had similar views on time-biases and he claimed that temporal neutrality is a normative requirement of prudence which demands an equal concern for all parts of life.

The impartial spectator does not feel himself worn out by the present labour of those whose conduct he surveys; nor does he feel himself solicited by the importunate calls of their present appetites. To him their present, and what is likely to be their future situation, are very nearly the same: he sees them nearly at the same distance, and is affected by them very nearly in the same manner.<sup>2</sup>

“The impartial spectator” is an ideal impartial person who fully empathizes with the emotions and actions of “normal” people. Smith assumed – as many philosophers before him – that the soul is composed of two different faculties or orders (“the faculty by which we judge of truth and falsehood” and “passions and appetites which are so apt to rebel against their master”). This lower faculty (“inferior and brutal appetites”) is responsible for time-biases, which result from cognitive limitations or the weakness of will. It is why the impartial spectator, who can be understood also as a metaphor for the higher faculty, does not manifest time-biases (“to him their present, and what is likely to be their future situation, are very nearly the same”). Both Hume and Smith maintain that it is impossible to correct these lower inclinations, because it is impossible “to change or correct anything material in our nature” (Hume). The only way to restraint them is to use some contrary impulses and the mild paternalism, that corrects human value judgments.<sup>3</sup>

Henry Sidgwick in *The Methods of Ethics* defended a similar view, but he took one step further claiming that the location of some experience in time within a single human life is not relevant:

---

<sup>2</sup> A. Smith, *The theory of moral sentiments*, Penguin, London 2010 (1759/2010), VI, I, 11

<sup>3</sup> I. Palacios-Huerta, “Time-inconsistent preferences in Adam Smith and David Hume”, *History of Political Economy* 35(2) (2003), pp. 241–268.

[P]roximity is a property that it's reasonable to disregard except in its effect on uncertainty. My feelings next year should be just as important to me as my feelings next minute, if only I could be equally sure of what they will be. This impartial concern for all the temporal parts of one's conscious life is a prominent element in the common notion of the rational as opposed to the impulsive pursuit of pleasure.<sup>4</sup>

Sidgwick assumes that the moment in time in which particular feelings exist is no reason to prefer one of them over another. There are no reasons to care more about one part of my life than to care about any other. Moreover, temporal neutrality is an important element of the rational pursuit of pleasure. He writes: “[T]he mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment than to that of another”.<sup>5</sup> He also holds that the good of one person consists in the aggregation of the different goods of each moment as they follow one another in time. Similarly, the universal good is constructed by the comparison and integration of the good of many different individuals.

It seems, however, clear that the time at which a man exists cannot affect the value of his happiness from a universal point of view; and that the interests of posterity must concern a Utilitarian as much as those of his contemporaries, except in so far as the effect of his actions on posterity – and even the existence of human beings to be affected – must necessarily be more uncertain.<sup>6</sup>

Nevertheless, he claims that there is an important factor that can influence preferences both in individual and collective cases: the uncertainty of the effects “on posterity”. Yet when there is no uncer-

---

<sup>4</sup> H. Sidgwick, *The methods of ethics*, Hackett Publishing, London 1907, II, 2, 1.

<sup>5</sup> *Ibidem*, III, 13, 3.

<sup>6</sup> *Ibidem*, IV, 1, 2.

tainty, we should not be affected by the time-biases which are not “reasonable” and which do not suit the rational “pursuit of pleasures”.

John Rawls in *A Theory of Justice* accepted Sidgwick’s view and explored the connection between time-neutral preferences and rationality:

In the case of an individual the avoidance of pure time preference is a feature of being rational. As Sidgwick maintains, rationality implies an impartial concern for all parts of our life. The mere difference of location in time, of something’s being earlier or later, is not in itself a rational ground for having more or less regard for it. Of course, a present or near future advantage may be counted more heavily on account of its greater certainty or probability, and we should take into consideration how our situation and capacity for particular enjoyments will change. But none of these things justifies our preferring a lesser present to a greater future good simply because of its nearer temporal position.<sup>7</sup>

And then he is even more radical:

Mere temporal position, or distance from the present, is not a reason for favoring one moment over another. Future aims may not be discounted solely in virtue of being future, although we may, of course, ascribe less weight to them if there are reasons for thinking that, given their relation to other things, their fulfillment is less probable. The intrinsic importance that we assign to different parts of our life should be the same at every moment of time. These values should depend upon the whole plan itself as far as we can determine it and should not be affected by the contingencies of our present perspective.<sup>8</sup>

In those two fragments Rawls presents very strong views on the irrationality of time-biases. First of all, he claims that “rationality im-

---

<sup>7</sup> J. Rawls, *A theory of justice*, Belknap Press, Cambridge, MA 1999 p. 259.

<sup>8</sup> *Ibidem*, p. 369.

plies an impartial concern for all parts of our life". He assumes that maximizing values during someone's lifespan is the requirement of rationality. Therefore the mere location in time of some value is not "a rational ground" for preferring some value. In the second quotation he repeats this view in a slightly modified form. He claims that temporal position of some value is not a reason for favouring this value in one moment over another. It also seems that he treats as one the requirements of rationality and "being a reason for favouring something".

The longest and the most sophisticated discussion about the rationality of time-biases can be found in Derek Parfit's *Reasons and Persons*. At the beginning of chapter 8 he asks the following question: "On the Desire-Fulfilment Theory, should I give equal weight to all of my desires, past, present, and future?" According to the Self-interest Theory and Desire-Fulfilment Theory, which are the subject of his criticism in the first chapters of *Reasons and Persons*, the strongest reasons speak in favour of these actions which lead to the fullest fulfilment of someone's desires, no matter how they are situated in time. But, Parfit argues, it cannot be true.

Firstly, regarding past desires, he proposes the following thought experiment: let us assume that for fifty years someone has two strongest desires: trying to ensure that Venice will be saved and being one of its saviours. Yet after 50 years this person ceases to have these desires. Does he still have a reason to contribute to the Venice Fund? If we assumed total time-neutrality, we should answer positively to this question, because acting in this way this person could help to fulfil two of his past desires. Although he no longer has these desires, they were his strongest desires for fifty years and he will not have any other so strong and so long-lasting. So it would be irrational to cease to contribute even though he do not now have, and shall never later have, any desire to contribute but this solution seems awkward.

Secondly, regarding the future desires that depend on value judgments, Parfit quotes the well-known principle: we cannot honestly claim that  $q$  is true, but not believing in  $q$ . Let us imagine that in a particular moment in time someone comes to the conclusion that in

the future his desires that depend on value judgments will change: he will stop believing that *q* and he will start believing that not-*q*. Again, if we assumed total time-neutrality, this man would have to give the same weight to his present and his predicted future values. This would be giving the same weight to what he now believes to be justified and to what he now believes to be worthless and this is clearly irrational – as Parfit writes. It is impossible to believe that *q* is true and in the same time be neutral towards this *q*. If someone believes that some of his later beliefs will be better justified, he should have this belief now. Parfit concludes that “a rational agent must give priority to the values or ideals that he now accepts” over those which accepted in the past or will be accepted in the future.

The next problem related to time-biases that Parfit discusses is a distinction between two phenomena. The first is the bias towards the near, the second is the bias towards the future. This first phenomenon was partially noticed by Hume in the quotation I have discussed: we have an inclination to the present goods or, in other words, we discount the value of future goods. This second phenomenon is described by Parfit in the following way: “The thought of such events affects us more when they are in the future rather than the past. Looking forward to a pleasure is, in general, more pleasant than looking back upon it. And in the case of pains the difference is even greater.” (Parfit, 1984, 160) Some preferences or behaviours that seem to be incoherent with the first phenomenon can be explained by the second. For example, we sometimes bring pains into the nearer future, and postpone pleasures. The second phenomenon can explain this easily: we may want to get the pains behind us and to keep the pleasures before us. Parfit suggests that there is something wrong with this these two biases. Although his views on the bias toward the future are not always very clear, his views on the bias toward the near are explicitly critical:

But on all plausible theories one point is agreed. When we are deciding what is in someone’s interests, we should discount for uncertainty, but not for mere remoteness. We should not give less weight to this



person's further future, or give greater weight to his present desires... We should give equal weight to all the parts of this person's life.<sup>9</sup>

It is not clear what is a nature of this requirement of "equal weight to all the parts of someone's life". It can be guessed that it is neither the requirement of morality, nor the requirement of rationality (understood as a mere coherence of attitudes), but the requirement of prudence: it is in everybody's interests not to have it.

It is quite difficult to deduce what is Parfit's view about the bias towards the future (ultimately he seems to be neutral regarding its irrationality). On the one hand, he claims that the bias toward the future is sometimes profitable for us and sometimes not:

Our bias towards the future is bad for us. It would be better for us if we were like Timeless. We would lose in certain ways. Thus we should not be relieved when bad things were in the past. But we should also gain. We should not be sad when good things were in the past.<sup>10</sup>

On the other hand, Parfit finally suggests that it would be better for us if we were like Timeless, that is someone who experiences past harms which are just as strong as those that will follow. Why? He presents two arguments that show that, taking everything into account, the bias toward the future is not beneficial for us. The first refers to the mechanisms of memory.

When we look backward, we could afford to be selective. We ought to remember some of the bad events in our lives, when this would help us to avoid repetitions. But we could allow ourselves to forget most of the bad things that have happened, while preserving by rehearsing all of our memories of the good things.<sup>11</sup>

---

<sup>9</sup> D. Parfit, *Reasons and persons*, Oxford University Press, Oxford 1984, p. 162.

<sup>10</sup> *Ibidem*, p. 174.

<sup>11</sup> *Ibidem*.

Parfit's second argument for preferring neutrality is related to the more calm attitude toward death we might acquire if we got rid of the bias toward the future. As we age, we have less and less to look forward to and this is why we fear death. Of course, as we age we also have more and more to look back upon but, because of our bias, that fact does not compensate for our having less of a future to look toward. If we didn't have our bias, we could enjoy looking back at our past as much as we enjoy looking toward the future now, and hence we would not fear death.

Now suppose that our lives have nearly passed. We shall die tomorrow. If we were not biased towards the future, our reaction should mirror the one that I have just described. We should not be greatly troubled by the thought that we shall soon cease to exist, for though we now have nothing to look forward to, we have our whole lives to look backward to.<sup>12</sup>

So, on the one hand, Parfit presents arguments in favour of the view that the bias towards the future is not beneficial for us, but on the other he presents in the widely discussed example of a hospital which is designed to demonstrate that this bias is very common and, moreover, it seems to be justified and those lacking it would seem irrational.

I am in some hospital, to have some kind of surgery. Since this is completely safe, and always successful, I have no fears about the effects. The surgery may be brief, or it may instead take a long time. Because I have to co-operate with the surgeon, I cannot have anaesthetics. I have had this surgery once before, and I can remember how painful it is. Under a new policy, because the operation is so painful, patients are now afterwards made to forget it. Some drug removes their memories of the last few hours.

---

<sup>12</sup> *Ibidem.* p. 176.

I have just woken up. I cannot remember going to sleep... I may be the patient who had his operation yesterday. In that case, my operation was the longest ever performed, lasting ten hours. I may instead be the patient who is to have a short operation later today. It is either true that I did suffer for ten hours, or true that I shall suffer for one hour... [I]t is clear to me which I prefer to be true. If I learn that the first is true, I shall be greatly relieved.<sup>13</sup>

The problem Parfit noticed here is a reversal of time preferences: it seems that before the operation every rational agent should prefer the later, much shorter operation. But if you have just woken up and you are not sure whether you had the operation or not, the desire to be after the operation seems to be completely justified. Moreover, this bias revealed in this moment at first sight seems to be a paradigm of rational attitude, we would not regard the person without this bias as “normal”.<sup>14</sup>

Nevertheless, the general view which emerges from these classical texts is that a rational person will be temporally neutral and will assign no independent significance to the temporal location of goods and harms within his life. In the next section I will clarify some issues related to the time-biases.

## 2. Clarifications

The most important clarification concerns the very notion of time-bias preferences. The kind of preferences I am interested in are called “the pure time preferences” and refer solely to the psychological effects with respect to remoteness in time. They could be defined (with regard to the bias toward the near) as follows “I prefer to have X to-

---

<sup>13</sup> *Ibidem*, p. 165.

<sup>14</sup> D. Moller, “Parfit on pains, pleasures, and the time of their occurrence”, *Canadian Journal of Philosophy* 32(1) (2002), pp. 67–82.

day rather than having *X* at some point in the future, for no other reason than the fact that I dislike waiting”.<sup>15</sup> According to another definition, a pure time preference is “a preference for something to come at one point in time rather than another, not because this will make the benefit greater or more certain, but merely because of when it occurs in time”.<sup>16</sup>

The mere possibility of separating pure time preference – both in empirical research or in conceptual analyses – from other factors underlying time preferences can be questioned. Many preferences that seem to be based solely on time do not involve time-biases in my sense. For example you can prefer to be rich now than in 30 years because you will have more time to enjoy your fortune. In such cases preferences are not time-biased because it is clear that the mere location in time of the same event (becoming rich) can result in different gains. Therefore it is not the mere difference of location in time that makes you prefer to be rich today, but the fact that you will gain more if you receive the wealth earlier.

Probably the most important is “the uncertainty effect” visible in the quotations from Sidgwick and Rawls: given the uncertainty of many future events, in many cases it is reasonable to attribute less value to future outcomes than to immediate ones. The uncertainty does not have to be limited to the results of our actions. We also cannot be certain in which way we will value these outcomes. For example I can prefer go skiing now than in 30 years not because of a pure time preference, but because I am not sure whether I will enjoy skiing when I am in my sixties.

The other important clarification concerns the distinction between inter-temporal and intergenerational issues. In this paper I am interested mostly in the rationality of the preferences of an individual on her own behalf (the exception is the last section where I dis-

---

<sup>15</sup> G. Tinghög, “Discounting, preferences, and paternalism in cost-effectiveness analysis”, *Health Care Analysis* 20(3) (2012), p. 302.

<sup>16</sup> R. Lowry, M. Peterson, “Pure time preference”, *Pacific Philosophical Quarterly* 92(4) (2011), p. 490.

cuss the problem of in which way some present normative orders, like legal regulations, have emerged in the context of time-biased preferences).

The next problem, closely related to the previous one, is a distinction between the criteria of individual and collective actions. I will only deal with a problem of the inter-temporal allocation of values within someone's life and not with the allocation of values between different individuals separated by time. It is worth noting that that many discussions by philosophers or economists concern not the individual decisions about the distribution of goods within someone's life, but policy-makers who must decide in which way to distribute goods among different generations of people. Much of the disagreement about discounting, in particular between economists and philosophers, seems to originate from the lack of recognition of this distinction.<sup>17</sup> It was visible for example in the quotation by Sidgwick, where he claims that "the interests of posterity must concern a Utilitarian as much as those of his contemporaries". Sidgwick (but also Rawls) seemed to use these two interchangeably: he held that if individual time biases are irrational, the same must be said about policy makers who must not prefer the interests of present people over the interests of future people only because of the time of their existence. Of course there is no easy way from inter-temporal to inter-generational issues, although the individual tendencies to care more about the near is often considered to be a reason to place less value on future generations.

In the final section I will discuss briefly whether behavioural effects related to time preferences tested empirically provide any reason for society to favour (or not) the present over the future when making intergenerational choices. The initial objections would say that individual discounting does not imply social discounting: someone may discount his own future heavily, but in the same time be deeply con-

---

<sup>17</sup> G. Brennan, "Discounting the future, yet again", *Politics, Philosophy & Economics* 6(3) (2007), pp. 259–284.

cerned about the welfare of distant generations, and favouring some social policies.<sup>18</sup>

The important factor in discounting future goods is growth and diminishing marginal utility: it is assumed that the utility derived from a unit of good will decrease as the amount of this good increases (if there is more of some good, an individual derives from a given unit of that good less utility). The value of 2000 PLN is much higher for a person who earns this amount during one month than for a person who earns it in one day. So, if an individual expects that he will have more goods in the future, he can reasonably place a higher value on present goods, because this is growth and not pure time preference.

It is easy to notice that different categories of things can be discounted: commodities, wellbeing, health. It was noticed that this could also be a reason why different authors have different views on discounting. For example, John Broome observed:

Economists typically discount the sorts of goods that are bought and sold in markets, which I shall call commodities. Philosophers are typically thinking of a more fundamental good, people's wellbeing. There are sound reasons to discount most commodities, and there may well be sound reasons not to discount wellbeing. It is perfectly consistent to discount commodities and not wellbeing.<sup>19</sup>

He claims that it makes sense to employ a discount rate (probably even within one life) for commodities because the price of most them falls over time either because we can produce them in a more efficient way or because natural resources grow naturally or because we will be able to buy more because of the positive return from investments. But according to Broome there is no reason to discount our own future wellbeing.

---

<sup>18</sup> S. Frederick, "Valuing future life and future lives: A framework for understanding discounting", *Journal of Economic Psychology* 27(5) (2006), pp. 667–680.

<sup>19</sup> J. Broome, "Discounting the future", *Philosophy & Public Affairs* 23(2) (1994), p. 144.

One can also argue that the concept of personal identity, especially the link that connects a person in one time period with the same person in a later period, can influence the rate of discounting. For example, I am now closely connected with “myself” on the next day, but the psychological link that is guaranteed by memory and other processes, between me now and me in 30 years time is much weaker. I partially treat this future me as a different person, and I do not treat his potential future experiences as mine. The implication of this argument would be that some degree of short-sighted pure time preference would be rational if individuals become less psychologically attached to their future selves.

Finally, some authors have argued that the time at which goods occur can affect the narrative structure of a person’s life. They claim that the contribution of an individual action, event, or experience to well-being depends not only on its intrinsic properties, but also on what has gone on before and what will go on to happen. For example, one can believe that a life that begins badly but improves is better than a life that begins brilliantly, but deteriorates. So he can prefer to receive some achievements later rather than earlier, not because of some strange reversed bias toward the near, but because of his views about the structure of a good life.<sup>20</sup>

In this section I have distinguished pure time preferences (either in a form of a bias toward the near or in the form of bias toward the future) from other factors that can influence our time preferences, like the probability of future events, diminishing marginal utility, the kind of goods that are vulnerable to time-biased preferences, connectedness with my future self, the narrative structure of life. I have also delineated the problem I am interested in (the inter-temporal decisions of an individual) from three other problems: the inter-temporal decisions of policy makers, the intergenerational decisions of individual, the intergenerational decisions of policy makers. In the next section I will review briefly the most important empirical findings about time-biases.

---

<sup>20</sup> J.D. Velleman, “Well-being and time”, *Pacific Philosophical Quarterly* 72(1) (1991), pp. 48–77.

### 3. The anomalies of time-biases

Evidences from empirical psychology and economy suggest that in making intertemporal choices people discount future values. These results were obtained both from field studies (preferences were inferred from the economic decisions that people make in their real lives) and from experimental studies (people were asked to evaluate intertemporal prospects with hypothetical outcomes).

The most important and widely confirmed feature of time preferences is the fact that people discount future values and their discount rates are usually not constant over time (that is discounting is not exponential), but decline in the form of a hyperbolic discounting function. When subjects are asked to compare smaller but sooner rewards to larger but later ones, the discount rate over longer time horizons that was inferred from their behaviour was lower than the implicit discount rate over shorter time horizons. For example, the behavioural surveys show that most of us would prefer 110 PLN in 31 days over 100 PLN in 30 days, but also prefer 100 PLN now over 110 PLN tomorrow (it is also assumed that such preference reversals would also hold if subjects who currently prefer 110 PLN in 31 days over 100 PLN in 30 days were asked again thirty days later, they would prefer 100 PLN at that time over 110 PLN one day later). This is the best documented phenomenon which implies that different discount rates are used to evaluate close and far away goods. In one well-known survey, subjects were asked how much money they would require in one month or one year or ten years to make them indifferent to receiving 15 USD now. The median responses (20 USD / 50 USD / 100 USD) demonstrated an average annual discount rate of “345 percent over a one-month horizon, 120 percent over a one-year horizon, and 19 percent over a ten-year horizon”.<sup>21</sup>

---

<sup>21</sup> R. Thaler, “Some empirical evidence on dynamic inconsistency”, *Economics Letters* 8(3) (1981), pp. 201–207.



This pattern is characteristic not only of humans: pigeons have also had to choose between a smaller-sooner reward and a larger-later reward (food). When the smaller reward was available after 28 seconds and the larger after 32 seconds, all the pigeons strongly preferred the larger reward. However, when the smaller reward was available after 2 seconds and the larger after 6 seconds, all the pigeons overwhelmingly preferred the smaller reward.<sup>22</sup>

Nevertheless there are some important reservations to these seemingly clear views on discount rates that are revealed in behaviour. Firstly, when we exclude studies with very short time horizons (less than one year) there is no behavioural evidence that the discount rate declines. Secondly, there is great variability in the estimates based on behavioural surveys (the implicit annual discount rates range from –6 percent to infinity). Thirdly, in contrast to estimates of physical phenomena, there is no evidence of methodological progress; the range of estimates does not shrink over time.

### 3.1. Some anomalies of the bias toward the near

The hyperbolic discounting is not the only regularity tested in an empirical way. There are also a few other characteristic patterns.<sup>23</sup> Firstly, gains are discounted more than losses (“the sign effect”). For example, subjects were asked to imagine they had received a traffic ticket that could be paid either now or later and to state how much they would be willing to pay if payment could be delayed (by three months, one year, or three years). The discount rates imputed from these answers were much lower than the discount rates imputed from comparable questions about monetary gains.<sup>24</sup>

---

<sup>22</sup> L. Green, E. Fisher, S. Perlow, L. Sherman, “Preference reversal and self control: Choice as a function of reward amount and delay”, *Behaviour Analysis Letters* 1 (1981), pp. 43–51.

<sup>23</sup> S. Frederick, G. Loewenstein, T. O'Donoghue, “Time discounting and time preference: A critical review”, *Journal of Economic Literature* 40(2) (2002), pp. 351–401.

<sup>24</sup> R. Thaler, “Some empirical evidence...”, *op. cit.*

Secondly, smaller amounts are discounted more than larger amounts (“the magnitude effect”). In one study, for example, respondents were, on average, indifferent between \$15 immediately and \$60 in a year, \$250 immediately and \$350 in a year, and \$3000 immediately and \$4000 in a year, implying discount rates of 139 percent, 34 percent, and 29 percent, respectively.<sup>25</sup>

Thirdly, “greater discounting is shown to avoid delay of a good than to expedite its receipt” (“the delay-speedup asymmetry”).<sup>26</sup> It was demonstrated that imputed discount rates can be affected by whether the change in the delivery schedule of an outcome is presented as an acceleration or a delay. For example, respondents who didn’t expect to receive a good for another year would pay much less to receive it immediately, in comparison to those who thought they would receive it immediately demanded much more money to accept a delay in its receipt by a year.<sup>27</sup>

Fourthly, in choices over sequences of outcomes, improving sequences are often preferred to declining sequences although positive time preference dictates the opposite (“preference for improving sequences”). For example in one study it was found that, for an otherwise identical job, most subjects prefer an increasing wage profile (i.e. improving sequence) to a declining or flat one.<sup>28</sup> This is surprising since the economic calculus suggests that it should be the other way round: it is better to earn more earlier, and less later, because you can invest money and exploit interest rates. Other studies found an even more surprising effect: increasing salary sequence was rated as highly as a decreasing sequence that earn subjects much more money. Daniel Kahneman, in a series of well-known studies, found that subjects strongly preferred decreasing discomfort (pain)

---

<sup>25</sup> *Ibidem*.

<sup>26</sup> S. Frederick, *et al.*, “Time discounting and time preference...”, *op. cit.*

<sup>27</sup> G.F. Loewenstein, “Frames of mind in intertemporal choice”, *Management Science* 34(2) (1988), pp. 200–214.

<sup>28</sup> G. Loewenstein, N. Sicherman, “Do workers prefer increasing wage profiles?”, *Journal of Labor Economics* 9(1) (1991), pp. 67–84.

to increasing discomfort, even when the decreasing discomfort last longer.<sup>29</sup>

Fifthly, in choices over sequences people have a preference for spreading consumption over time (“preference for spread”). In one experiment subjects were asked to imagine that they were given two coupons for restaurant dinners, and were asked to determine when they would use them. Subjects were either told that “you can use the coupons at any time between today and two years from today” or were told nothing about any constraints. Subjects who faced the two-year constraint scheduled both dinners at a later time than those who had no constraints. This second group delayed the first dinner for eight weeks (rather than three) and the second dinner for 31 weeks (rather than thirteen). This counterintuitive result can be explained in terms of a preference for spread if the explicit two-year interval was greater than the implicit time horizon of subjects in the unconstrained group.<sup>30</sup>

### 3.2. The bias toward the future

In contrast to the bias toward the near, the bias toward the future was hardly tested empirically. Probably there is only one experiment explicitly aimed at investigating this bias and testing the hypothesis that people place a greater value on events in the future than in the past.<sup>31</sup> Subjects in this experiment read pairs of stories describing two events, one of which occurred in the future and one of which occurred in the past (the distance from the present was the same in both cases). In the first of the five experiments, subjects were asked to imagine that they would spend (or had spent) five hours on work one month in the future (or one month in the past) and to indicate how much money

---

<sup>29</sup> D. Kahneman, B.L. Fredrickson, C.A. Schreiber, D.A. Redelmeier, “When more pain is preferred to less: Adding a better end”, *Psychological Science* 4(6) (1993), pp. 401–405.

<sup>30</sup> G.F. Loewenstein, D. Prelec, “Preferences for sequences of outcomes”, *Psychological Review* 100(1) (1993), pp. 91–108.

<sup>31</sup> E.M. Caruso, D.T. Gilbert, T.D. Wilson, “A wrinkle in time asymmetric valuation of past and future events”, *Psychological Science* 19(8) (2008), pp. 796–801.

they thought would be fair compensation for their work. Subjects responded that they wanted to receive twice as much money for performing the task in the future as in the past.

The researchers assumed that the subjects' valuation of the past and the future may be related to differences in their affective responses. To test this, subjects were asked to imagine that they had agreed to help their neighbour move out of his apartment (either they helped the neighbour move one week in the past and one in which they would help him one week in the future). The story stipulated that the neighbour would afterwards give him a voucher for a bottle of wine. Subjects asked to imagine helping the neighbour in the future chose a bottle of wine that was 38 percent more expensive than the bottle of wine they selected when given the past version of the story. Moreover, subjects who imagined helping their neighbour move in the future felt significantly more tired, stressed, and dreadful than subjects who imagined helping the neighbour in the past.

This study is important but the problem is that empirical research on this time-bias is limited, so firstly, these results have not been checked, secondly, there are probably many other important phenomena related to this kind of bias. For example, Caspar Hare<sup>32</sup> suggested (it was an armchair philosophy intuition) that future bias can be selective: we are not future-biased with respect to things that we take to be bad for us but badness is not related to having bad experiences. He assumes that in this case we would not care about whether the bad event was in the immediate past or the immediate future (I will return to this problem in the section 4.1).

### 3.3. The interpretation of these results

One of the most surprising results of experiments of time-biases is the fact that very often people do not treat them as errors. This is in con-

---

<sup>32</sup> C. Hare, "A puzzle about other-directed time-bias", *Australasian Journal of Philosophy* 86(2) (2008), pp. 269–277.

trast with the effects documented in other domains of judgments and choice. For example, in the famous conjunction fallacy described by Kahneman and Tversky, most subjects were able to recognize their initial mistakes when the nature of the experiment was explained to them (e.g., that it cannot be more likely for a woman to be a feminist bank teller than just a bank teller). But regarding the preferences for improving sequences, subjects were quite reluctant to abandon their preferences even when the nature of the experiment has been explained to them (that a decreasing wage profile would permit more consumption in every period than the corresponding increasing wage profile). In many experiments the respondents did not attempt to coordinate their responses, suggesting that they had not considered the different discount rates as something wrong. For example, the magnitude effect is more clearly visible when respondents evaluated both small and large amounts than when they evaluate one of those. The same results were obtained with the sign effect where discount rates were larger if subjects evaluated both delayed gains and delayed losses than when they evaluated either one.

This feature of time-biases shows that these preferences (or at least not all of them) belong to no standard category of preference failures. On the one hand, they do not belong to myopic failures which relate to a cognitive inability to compute information adequately to make choices that maximize the wellbeing of an individual. These failures refer to cases where there are difficult to predict outcomes, where the quantity of information is great to assess it properly, or where individuals are not able to interpret and calculate the implications of certain statistical or probabilistic situations.<sup>33</sup> In these cases respondents are able to recognize their own mistakes. On the other hand, time-biases do not belong to acrotic failures which relate to situations in which individuals are not able to control their own preferences effectively and act against their own better judgment. When making acrotic failures people's judg-

---

<sup>33</sup> A. Tversky, D. Kahneman, "The framing of decisions and the psychology of choice", *Science* 211(4481) (1981), pp. 453–458.

ments are different from the course of action they choose. So it is also not a case of at least many time-biased preferences, because in this case there is no inconsistency between judgments and choices.

#### **4. Time-biases and the requirements of rationality**

In this section I will analyse time-biases with the help of the analytical tools of recent philosophy. I try to do so in light of the recent discussions on the nature of the requirements of rationality and to evaluate the possible significance of the empirical findings for normative moral theories. The general question of this part is normative: should we think that rationality does not permit us to prefer receiving some value in one part of one's life over receiving the same number of units of value in another part of one's life?

The simple answer seems to be that rationality requires the maximization of our values, no matter where or when they exist or will exist. And this is the reason why time-biases are irrational: they interfere with this aim. Therefore it might seem that if we want to act rationally we should decide as if from a universal point of view, from which the time at which values exist does not matter. This kind of normative answer is characteristic mainly for the utilitarian perspective (implemented implicitly in some economic analyses that are conducted in terms of cost-benefit analysis or in many evaluative judgments by psychologists), which assumes that the goodness of outcomes is measured by the total utility resulting from the actions, irrespective of who gets the utility and when. According to this normative position we should act in such a way despite the fact that we have strong psychological inclinations to do otherwise.

##### **4.1. Time-biases and rationality as the coherence of attitudes**

The preference to receive some intrinsic value in one part of one's life over the same number of units of intrinsic value in another part

of one's life might be just a matter of taste, and as such it could be beyond being rational or irrational. In this understanding, if only an agent's set of preferences is consistent he is rational. This understanding of normativity is related to an attitude-dependent conceptions of reasons, where there are no facts about how an agent has the most normative reason to do that are valid independently of that agent's evaluative attitudes.<sup>34</sup> Therefore the only failures of rationality are failures to have a coherent system of beliefs and intentions. For example, we can enumerate the following requirements of rationality:<sup>35</sup>

- 1) Rationality requires of you that you do not both believe p and believe not-p.
- 2) Rationality requires of you that, if you believe p and you believe (if p then q), and if it matters to you whether q, then you believe q.
- 3) Rationality requires of you that, if you intend to G, and if you believe your Fing is a necessary means to your Ging, and if you believe you will not F unless you intend to F, then you intend to F.
- 4) Rationality requires of you that, if you believe you ought to F, and if you believe you will not F unless you intend to F, then you intend to F.

In this paradigm we are also irrational if we violate the basic axioms of decision theory (completeness, independence, continuity, transitivity): for example we assign a lower degree of probability to the occurrence of event A than A and some distinct event B (like in Kahneman's case of a bank-teller), or we are irrational if we prefer strawberry ice-cream to blueberry ice-cream, blueberry to blackberry and blackberry to strawberry. To the extent that a person's judgments and decisions accord with formal logic and the standard axioms of decision theory, they are rational.

---

<sup>34</sup> S. Street, "In defense of future Tuesday indifference: Ideally coherent eccentrics and the contingency of what matters", *Philosophical Issues* 19(1) (2009), pp. 273–298.

<sup>35</sup> The list is inspired by: J. Broome, "Does rationality give us reasons?", *Philosophical Issues* 15(1) (2005), pp. 321–337.

Of course, even within this paradigm some time-biases could be treated as irrational and they could be understood in a similar fashion to akrasia: we would be irrational when the time-biases revealed by our motivations fails to match the judgments we make about correct structure of time preferences. For example, if you believe that you should not discount the value of your future health, but you do not intend to change some habits that demonstrate that in fact you do discount your health, you would be irrational. The problem is that empirical evidence suggests that it is not the case that time-preferences do not match value judgments. Most people think that their time-biases are perfectly justified.

So should we accept that: “The mere fact that you prefer to consume something now rather than in the future, for no other reason than time itself, need not violate any of the internal properties of rationality researched in decision theory”?<sup>36</sup> I will demonstrate that this is indeed true in the case of future bias, but does not have to be true in the case of the bias towards the near (at least when an agent does not exhibit dynamic inconsistency).

Recently at least two arguments have been put forward to demonstrate that our future-bias can lead to such a violation. Caspar Hare noticed that we are selective in our future-biases and this can lead to incoherence in our attitudes.<sup>37</sup> The fact that we are selective means that sometimes we have future-bias and sometimes no. For example, we prefer to have a very painful surgery behind us then a less painful surgery ahead. But when we evaluate what would we prefer for other people we do not exhibit this bias: we probably prefer a public health care provider to do less painful surgeries than more painful, no matter when. To demonstrate that this kind of selective bias can lead to incoherence, Hare presented the following example (I present here a shorter version of this story from his entry to *A Companion to the Philosophy of Time*).

---

<sup>36</sup> R. Lowry, M. Peterson, “Pure time preference”, *op. cit.*

<sup>37</sup> C. Hare, “A puzzle about other-directed time-bias”, *op. cit.*



I learn, by letter, that my daughter, away at a distant, monastic retreat, far from phones or email, was scheduled either to have her wisdom teeth removed under a weaker local anaesthetic on Wednesday, or to have her wisdom teeth removed under a stronger local anaesthetic on Thursday. But this time I am unsure where the letter came from. She is either staying in a monastery in the far-north of Japan or staying in a monastery in the far-south of Japan. Undeterred, my wife and I jump on planes – hers heading to the south, mine heading to the north. On early Thursday morning I arrive at the northern monastery and am confronted with a sleeping figure. In the dim light I cannot quite tell if it is my daughter, and certainly cannot tell if my daughter has had her operation yet. One thing I do know is this: I am significantly better than my wife at allaying preoperative anxieties (with games, funny stories, etc.) and my wife is significantly better than me at providing post-operative comfort (with hugs, grapes, etc.).<sup>38</sup>

He noticed that our preference changes when the object (in this case our daughter) is close or far away. If she is close, we prefer she has had this operation behind her, if she is far away we evaluate her total well-being from a more impersonal perspective and we care about minimizing her suffering, no matter when it occurs (it is only his armchair intuition, it has not been verified behaviourally). Hare writes: “The majority view is that in the far-away case you should prefer that she have a better life, and in the nearby case you should prefer that she be in a better predicament”.<sup>39</sup> The problem is that he claims that in such a situation the structure of preferences would be incoherent and thus irrational. Why? Let us assume that there are four possible options here.

- A. She is near, and she has had the more painful operation.
- B. She is far, and she has had the more painful operation.

---

<sup>38</sup> C. Hare, “Time – the emotional asymmetry”, in H. Dyke, A. Bardon, *A Companion to the Philosophy of Time*, Wiley, Oxford 2013, pp. 507–520.

<sup>39</sup> C. Hare, “A puzzle about other-directed time-bias”, *op. cit.*, p. 271.

- C. She is near, and she will have the less painful operation.
- D. She is far, and she will have the less painful operation.

In this case Hare claims that the structure of preferences would be cyclical (I assume that  $B > A$  means that an agent prefers B to A):  $A > C > D > B > A$ . And having this kind of preferences is a clear example of irrationality according to many authors. Of course, this example does not show that we should get rid of future-bias because of the requirements of rationality. To avoid incoherency we can just become less selective in our future-biases. Indeed, Hare suggests that since you shouldn't be selectively future-biased because of the requirements of rationality, you should be constantly future biased because of the requirements of morality. It means that in the above example you should prefer that you daughter has have the worse operation no matter whether she is near or far, because you know that this worse surgery has already been finished.

It seems to solve the problem of irrationality of some future-biases, but there is a similar argument but with a much stronger conclusion.<sup>40</sup> It suggests that rationality requires not being future biased at all. Dougherty claims that two preferences that seem to be permissible independently, that is future bias and risk aversion, produce an irrational preference set when held simultaneously. He demonstrated that anyone who has the future bias and is risk averse can be turned into "pain pump": he would accept a series of deals that guarantees he will suffer more pain overall and be better off in no respect. To prove this he use the following example, a modified version of Parfit's hospital case:

On Monday, you are admitted into a hospital. You are told you will have one of two courses of operations, but you are not told which. If you have the early course, then you will have a painful, four-hour operation on Tuesday and a painful, one-hour operation on Thursday. If you have the late course, then you will have a painful, three-hour ope-

---

<sup>40</sup> T. Dougherty, "On whether to prefer pain to pass", *Ethics* 121(3) (2011), pp. 521–537.

ration on Thursday. After any operation, you will have amnesia for several days, and so you will not be able to remember if you have just had an operation. There is a calendar next to your bed, and so you always know what day it is.<sup>41</sup>

Dougherty proposes that on Monday you will be offered a deal that will take 29 minutes of pain off the Thursday operation if you are on the early course, and add 31 minutes to the Thursday operation if you are on the late course. And, either way, on Wednesday you will be offered a deal that will add 30 minutes to the Thursday operation if you are on the early course, and take 30 minutes off the Thursday operation if you are on the late course. Your decision on Monday will have no causal influence over your decision on Wednesday.

What happens if you are future-biased and risk averse? On Monday you accept the offer, because you want to “insure” yourself against the worst expected outcome: you are ready to add one minute of pain to the better course to reduce pain by one minute in the worse outcome. But on Wednesday you are also ready to accept the offer. Accepting it does not affect your expected amount of pain but it again reduces the risk you face: it reduces the gap between the bad outcome and the good outcome. Moreover, in this case this “insurance” is also for free from the perspective of expected utility. But if you take both pills you can be sure to have one minute more pain on Thursday and gain nothing: you become “a pain pump”, you can be easily exploited (in terms of pain) without any compensation (if you had the early course you took off 29 minutes of pain on Monday, but added 30 minutes on Wednesday; if you had the late course you added 31 minutes on Monday, but added 30 minutes on Wednesday). In effect, Dougherty claims that we have four possible options in cases like this:

(1) We could deny that you are rationally forbidden from having preferences that allow you to be turned into a pain pump. Alternatively, we

---

<sup>41</sup> *Ibidem*, p. 526.

could deny that you are permitted to be time-biased and risk averse. Notice that there are several ways in which we could deny this: (2) we could claim that separately time-bias and risk aversion are rationally permissible but the combination of the two is impermissible, (3) we could deny that risk aversion is rationally permissible, or (4) we could deny that time bias is rationally permissible.<sup>42</sup>

He suggests that the fourth option is probably the best, although at first glance it seems counterintuitive: it requires that you are indifferent as to whether you have already experienced strong pain or you are about to experience it.<sup>43</sup>

#### 4.2. Time-biases and rationality as reason responsiveness

There is also a second dimension in which our time-biases can be irrational. This type of criticism is applied to the bias toward the near. Some philosophers claim that there exist objective and external facts about the correctness of time-biases and responding to these facts would be a requirement of rationality.<sup>44</sup> On this view some beliefs and intentions are intrinsically irrational. There are some beliefs and intentions such that even if an agent is internally consistent, he or she:

is nevertheless making *a normative mistake* in the sense that he or she is valuing something that is not in fact valuable, or desiring something that is not in fact desirable. To put the view in reasons terminology, there are facts about how an agent has the most normative reason to live that hold independently of that agent's evaluative attitudes and what follows from within the standpoint constituted by them; instead, an agent's normative reasons are always ultimately a function of that agent's own evaluative attitudes and what is logically or instrumen-

<sup>42</sup> *Ibidem*, p. 533.

<sup>43</sup> See also: P. Greene, M. Sullivan, "Against Time Bias", *Ethics* 125(4) (2015), pp. 947–970.

<sup>44</sup> D. Parfit, *On what matters*, vol. 1, Oxford University Press, Oxford 2011.

tally entailed by those attitudes in combination with the non-normative facts.<sup>45</sup>

The important point here is that rationality must require much more than mere coherence between attitudes. It must also require that an agent responds properly to some reasons or at least to apparent reasons.<sup>46</sup>

There are at least two main problems with such an understanding. First of all the understanding rationality as responsiveness to reasons is controversial. I have argued elsewhere that Parfit's views in this matter lead to problematic conclusion: he does not have the tools to distinguish truth from rationality in the case of normative beliefs, so he must assume that every normative belief that is rational must also be true. This is contrary to the case of non-normative beliefs when it is quite usual that a rational belief does not have to be a true belief, for example because we do not know – without our fault – some important piece of information. In the case of normative beliefs it is not possible, leading to awkward results.<sup>47</sup>

Secondly, the claim that time-biases are irrational in this sense is based on a conception that overlooks a distinction between preferences that are rationally required and preferences that are rationally permitted. The standard argument that time-biases are irrational can be summarized as follows:

- (i) Rationality implies an impartial concern for all parts of one's life.
- (ii) A preference for something which is based merely on a difference of location in time is a preference which is not based on an impartial concern for all parts of one's life.

---

<sup>45</sup> S. Street, "In defense of future Tuesday...", *op. cit.*

<sup>46</sup> D. Parfit, "Rationality and reasons", in D. Egonsson, J. Josefsson, B. Petersson, T. Rønnow-Rasmussen, *Exploring practical philosophy: From action to values*, Ashgate, London 2001, pp. 17–39.

<sup>47</sup> T. Żuradzki, *Internalizm i eksternalizm w metaetyce*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2012; T. Żuradzki, "Trzeci rodzaj normatywności – wymóg reagowania na racje", *Filozofia Nauki* 88(4) (2014), pp. 35–51.

(iii) A preference which is based merely on a difference of location in time is irrational.<sup>48</sup>

Lowry and Peterson target the first point: they argue that sometimes rationality allows time-biases and therefore it does not require that everyone must have an impartial concern for his or her life. They consider ways in which a preference can be irrational. Apart from the mere incoherence that I have discussed above there are two other ways. Firstly, because a preference conflicts with the other preferences that agent can have, and secondly, because it is arbitrary. Now I will discuss shortly these two ways and I will demonstrate that there can be time-biases that are rationally permissible.

The first reason why time-biases are called irrational can be the fact that a time-biased preference can conflict with the other preferences that an agent can have. For example Rawls discusses the case of a blade-of-grass counter, whose only preference is “to count blades of grass in various geometrically shaped areas such as park squares and well-trimmed lawns”.<sup>49</sup> Rawls does not call his preference irrational. He assumes that there are many other activities (from cycling to studying philosophy) that make our life better than grass counting, but there is nothing intrinsically irrational in having a preference of counting blades of grass.

Nevertheless, according to Parfit’s view, there are reasons to make other things that compete with the preference to count blades of grass and these other preferences should prevail if an agent is rational (that is reacts correctly to his or her reasons). In other words, a rational agent should have a proper “mechanism” that helps him to recognize and respond properly to the strength of reasons. .

The case of time-biases is similar: Parfit does not claim that time-biases are irrational as such, but only that they conflict with other reasons. It is worth noting that for Parfit there are no direct reasons not to

---

<sup>48</sup> R. Lowry, M. Peterson, “Pure time preference”, *op. cit.*, p. 492.

<sup>49</sup> J. Rawls, *A theory of justice*, *op. cit.*

be time-biased. All of Parfit's examples are designed to demonstrate only that it is wrong to be time-biased because of some other prudential reasons. For example: Parfit would claim that if someone had a belief that some dish would damage his health, this belief would make the desire to eat this dish irrational, because this desire would respond to objectively existing reasons in an inappropriate way. He claims that it is always irrational for me to act in a way that I know will be worse for me. The case of time-biases is different: they are irrational because they conflict with stronger prudential reasons.

The second reason why time-biases seem to be irrational is their arbitrariness.<sup>50</sup> A preference is arbitrary, when there are no reasons to have it (but also there are no reasons not to have it). For example, Parfit writes:

It is irrational to care less about future pains because they will be felt either on a Tuesday, or more than a year in the future... In these cases the concern is not less because of some intrinsic difference in the object of concern. The concern is less because of a property which is purely positional, and which draws an arbitrary line. These are the patterns of concern that are, in the clearest way, irrational.<sup>51</sup>

He claims that a preference is arbitrary when it does not reflect some intrinsic differences in objects of concern but is based on "purely positional property". What is wrong in these purely positional preferences? I will assume that he mistakenly claims that when a preference is based on a neutral basis then it must be irrational. A neutral basis means that there no reasons in favour and against this preference. Similar reasoning could be found in the above citations from Sidgwick<sup>52</sup> and Rawls<sup>53</sup> who write about "a reasonable ground" or "a rational ground".

---

<sup>50</sup> P. Greene, M. Sullivan, "Against Time Bias", *op. cit.*, p. 950,

<sup>51</sup> D. Parfit, *Reasons and persons*, *op. cit.*, pp 125–126.

<sup>52</sup> H. Sidgwick, *The methods of ethics*, *op. cit.*, III, 13. 3.

<sup>53</sup> J. Rawls, *A theory of justice*, *op. cit.*, p. 259.

Let me distinguish three possible ways in which preferences can be grounded.<sup>54</sup> Firstly it can be grounded on some good reasons: for example I prefer to eat a dish because I believe that it is healthy. Secondly, it can be grounded on some bad reasons: I prefer to eat it because I believe that it is poisonous. Thirdly a preference can be grounded on a neutral basis. Here is an example: I live in a city where there are two concert halls and I would like to listen to music this evening. In one of the concert halls, Beethoven will be played, in the other, Schumann. I believe that both concerts will be equally pleasurable, and I value Schumann as much as Beethoven. I do not have any reasonable grounds to prefer one concert over the other, but I must decide. So I finally decide to listen to Schumann. Is my decision, which is based on neutral grounds, necessarily irrational? I believe that this case is similar to the case of some time-biases. For example, let us imagine that the decision concerns not to which should concert I go, but when. Should I listen to Schumann today or tomorrow? Again I do not have any relevant preferences in this matter, and experiencing the concert gives me a fixed number of units of intrinsic value whenever I go, so you have no rational grounds for preferring to go on a particular occasion. But I must decide and finally I decide to go today. I decide only because of the pure time preference: I just prefer to have a pleasurable experience earlier than later. I do not see why (as Parfit claims) this kind of preference which is purely positional must be irrational. It is not a preference similar to the preference of this kind: "I prefer to eat because I believe that it is poisonous". And I also assume that there are no other stronger reasons to do something else (like in the example of blade-of-grass counter).

It seems that the problem with Parfit's solution is that he assumes that to have a rational preference we must have a good reason to have it. Therefore he does not distinguish between the preferences that are required from the preferences that are merely permitted. And a preference that is based on time-bias can sometimes be a rationally per-

---

<sup>54</sup> R. Lowry, M. Peterson, "Pure time preference", *op. cit.*



mitted preference. For example: the fact that I prefer to listen to music something rather earlier than later only because I have a pure time preference is not irrational itself. It can become irrational if there are some reasons against my earlier listening.

Parfit's tough position on the irrationality of time-biases can be partially explained by his reason externalism.<sup>55</sup> This position (he calls it "value-based theory of practical reasons" in contrast to "desire-based theories") claims that normative reasons for actions are all "provided by facts about what is relevantly good, or worth achieving."<sup>56</sup> If we add that Parfit assumes that time when an action is realized is not "relevantly good" and that he claims that rationality requires that we respond correctly to normative reasons for actions, we obtain the result that we are irrational when we are time-biased. In contrast, according to internalist view an agent has a reason for action if and only if that action is related to agent's "subjective motivational set" in a particular way. It must be the case that "A could reach the conclusion that he should  $\phi$ ... by a sound deliberative route from the motivations that he has in his actual motivational set – that is, the set of his desires, evaluations, attitudes, projects, and so on."<sup>57</sup> If normative reasons for action are related to an agent's subjective motivational set it is much easier to justify time-biased preferences. An agent who has this kind of preference can be perfectly rational, because the rationality of his desires depends (at least partially) on his own motivational set.

There are two main conclusions to this section. Firstly, there are good arguments that the bias toward the future is irrational in a very strong sense: it violates the requirement of rationality as coherence. It is a surprising conclusion since at first sight it might have seemed that this bias is less problematic than the bias toward the near. Secondly, arguments against the rationality of time preferences are more

---

<sup>55</sup> S. Street, "In defense of future Tuesday...", *op. cit.*

<sup>56</sup> D. Parfit, "Rationality and reasons", *op. cit.*

<sup>57</sup> B.A.O. Williams, "Internal and external reasons", in *Moral Luck*, Cambridge University Press, Cambridge 1981, pp. 101–113.

problematic, because they must assume that a preference that is not grounded on some objective reasons is irrational, and that there are no rational preferences based on neutral grounds.

## 5. Time-biases and the emergence of normative orders

The article is intended to help understand how the present day normative orders (i.e. popular moral judgments, legal and economic regulations) have emerged in the context of different forms of the time-biased preferences of individuals. Policy makers very often treat time-biases on the individual level as a one of main reasons to justify discounting future benefits and losses (that is to place a lower value on a future gain or loss than on the same gain or loss occurring now). The other justification for discounting is the productivity of capital (evaluating some project we should take into consideration that instead we can invest our money and have more money in future). The standard argument for the justification of discounting based on the preferences of individuals is the following. Firstly, policies in democratic societies should reflect the well-established preferences of individuals. For example, the first sentence of a review article about discounting says: “Prescriptive economics requires that, unless there are very good reasons to the contrary, economic policy should be based on the principle that individuals’ preferences should count”.<sup>58</sup> Secondly, empirical research unambiguously shows that people tend to care more about the present than about the future and this phenomenon concerns many different values (money, health, lives etc.). So if this behaviour reflects what people prefer, then discounting is required. These time-biases are visible in many normative orders: major infrastructure programs, health care programmes, environmental protection and actions against climate change, energy

---

<sup>58</sup> D. Pearce, B. Groom, C. Hepburn, P. Koundouri, “Valuing the future”, *World Economics* 4(2) (2003), pp. 121–141.

policies. For example, an European Union guide for a long term investment program claims:

Costs and benefits occurring at different times must be discounted. The discount rate in the economic analysis of investment projects the social discount rate (SDR) reflects the social view on how future benefits and costs should be valued against present ones. It may differ from the financial discount rate when the capital market is inefficient (for example when there is credit rationing, asymmetric information and myopia of savers and investors, etc.).

The EU commission assumes that costs and benefits must be discounted and that the discount rate reflects the social view what can be understood as the aggregation of an individual's preferences. The EU commission also assumes that the social discount rate can differ from the financial discount rate only when market is inefficient. Unfortunately, it does not propose any method in which it should be counted.

Publicly funded healthcare systems provide one of the most visible and ethically intriguing aspect of discounting. For example, a recently published study by WHO *Assessing health needs: the Global Burden of Disease Study* claims:

Individuals commonly discount future benefits against current benefits similarly to the way that they may discount future dollars against current dollars. Whether a year of healthy life, like a dollar, is also deemed to be preferable now rather than later, is a matter of debate among economists, medical ethicists, and public health planners, since discounting future health affects both measurements of disease burden and estimates of the cost-effectiveness of an intervention.<sup>59</sup>

---

<sup>59</sup> C.J.L. Murray, A.D. Lopez, C.D. Mathers, *Assessing health needs: The global burden of disease approach*, Oxford University Press, New York 2006.

The study mentions that the most visible effect of discounting is the value of health interventions that provide benefits largely in the future, for example “vaccinating against hepatitis B, which may prevent thousands of cases of liver cancer, but some decades later.” The question is how should we compare the cost of these interventions with others that can prevent or treat present day illness. For example, the cited WHO study notices that if future life years were discounted by 3 per cent per year, this would mean that a year of healthy life bought for 10 years hence is worth around 24 per cent less than one bought for now. It would also mean that we state that 1 life saved today is equivalent to saving 19.2 lives in 100 years. So the discount rate is a crucial parameter in healthcare analyses, it is often employed arbitrarily.

At first glance the discount rate, especially when concerns the value of health of future people, seems not to have any moral justification. Why should policy makers value the present health of the population more than the health of the population in five years time? On the one hand, it seems that the value of health in these two cases should be the same. On the other, if we get rid of discount rates, we shortly reach very counterintuitive results.

For example, it was seriously argued that publicly funded healthcare systems should not pay for patented treatments and wait until their protection expires, even if it would mean the death of many present people who could be rescued by these drugs.<sup>60</sup> The author of this proposition assumes, firstly, that it is better to “save the greater number in cases where all are equally deserving and we cannot save both groups”<sup>61</sup>; and secondly, that it would be wrong to apply a discount rate to future health benefits. He claims that “there is no justification for thinking that the time at which a medical need occurs should make

---

<sup>60</sup> J. Wilson, “Paying for patented drugs is hard to justify: an argument about time discounting and medical need”, *Journal of Applied Philosophy* 29(3) (2012), pp. 186–199.

<sup>61</sup> *Ibidem*, p. 186.

a difference to its moral weight”.<sup>62</sup> Therefore, since we can reasonably expect that future lifesaving will be cheaper than present lifesaving (for example because patents will expire), we should concentrate solely on the number of saved people, and not on the time when they are saved. A similar line of argumentation was presented (but as a *reductio ad absurdum* of the views that opposes using a discount rate to future health benefits) with regard to delaying aid intended to prevent relatively easily preventable death in the developing world.<sup>63</sup> The author claims that if we assumed strict neutrality between the present and future interests of people, there would never be sufficient reasons to rescue present people (for example, by humanitarian aid), because it would be always more efficient to invest money and save more people in the future.

One of the most controversial and discussed examples of the use discount rate is energy policy and mitigating human induced climate change. For example, the Stern Review estimated that “if we don’t act, the overall costs and risks of climate change will be equivalent to losing at least 5% of global GDP each year, now and forever. If a wider range of risks and impacts is taken into account, the estimates of damage could rise to 20% of GDP or more”.<sup>64</sup> Many economists were surprised by the scale of this prediction since it was dramatically different from earlier economic models. The main factor that underlies these very high estimates was a discount rate which is at the heart of the Review’s radical view of the damages from climate change. The review assumed – because of moral reasons – an extremely low discount rate and that future losses should be valued almost as much as present. The result was that we should now spend about 1 % of global GDP (500 billion USD) annually to fight global warming. Although the Stern Review was a serious report prepared for the British

---

<sup>62</sup> *Ibidem*, p. 187.

<sup>63</sup> D. Moller, “Should we let people starve—for now?”, *Analysis* 66(291) (2006), pp. 240–247.

<sup>64</sup> N.N.H. Stern, *The economics of climate change: the Stern review*, Cambridge University Press, 2007.

government, some economists treated this scale of expenditures as a *reductio ad absurdum* of the views that oppose using a discount rate to evaluate future environmental harms.<sup>65</sup>

The conclusion of this section is that it is relatively easy to understand the way in which present day normative orders have emerged in the context of individual time-preferences. The social discounting can be understood as an expansion of individual preferences. The standard model says that individuals have clear time-biased preferences, and societal value is maximised when aggregating these personal preferences within a community. The much harder question is how to evaluate these propositions which suggest that policies should not discount future values (e.g. the health or well-being of future population). To evaluate these views we have to take into account not only whether these individual time-biased preferences are rational but also many other problems. For example, there are probably differences in individual time-biased preferences between our own life and society as well as between intraand intergenerational distributions. Perhaps we have many different sets of time-biased preferences: one that reflects individuals' preferences for their own wellbeing extended in time, a second where individuals take social and intergenerational considerations into account. Some authors assume that if these time-biased preferences turn out to be irrational (or incorrect, or immoral), there would be "no more reason that public policy should reflect people's inability to weight time neutrally than that it should reflect people's incapacity to think rationally about large numbers or perform fancy arithmetic."<sup>66</sup> But this answer is too quick: I have demonstrated that time-biased preferences are not similar to myopic failures (like people's incapacity to think rationally about large numbers). Even if there were, there would still be a problem of paternalism: if adult citizens care less about the further future, are policy makers allowed to force people to care more?

---

<sup>65</sup> P. Dasgupta, "Discounting climate change", *Journal of Risk and Uncertainty* 37(2-3) (2008), pp. 141–169.

<sup>66</sup> R.E. Goodin, "Discounting discounting", *Journal of Public Policy* 2(1) (1982), pp. 53–71.

## 6. Conclusions

I have demonstrated that time-biases (or at least many of them) are very specific preferences and they do not belong to any standard category of obvious preference failures discovered recently by experimental psychology or economy. This is why it is not easy to settle the matter of their rationality. On the one hand, I have reached a surprising conclusion that the bias toward the future (together with a non-controversial premise that we are risk averse) violates the requirement of the transitivity of our value preferences, and therefore this type of time-bias seems to be irrational. On the other, I have argued that the standard arguments against rationality of the bias toward the near are surprisingly weak. I have also noticed that some arguments against time-biases are based on armchair intuitions about people's reactions that have not been checked as yet empirically. This is an important task since, as research on time-biases shows, empirical investigations can often inform normative analyses.

## References

- Brennan, G. (2007). Discounting the future, yet again. *Politics, Philosophy & Economics*, 6(3), 259-284.
- Broome, J. (1994). Discounting the future. *Philosophy & Public Affairs*, 23(2), 128-156.
- Broome, J. (2005). Does Rationality Give Us Reasons? *Philosophical Issues*, 15(1), 321-337.
- Caruso, E. M., Gilbert, D. T., & Wilson, T. D. (2008). A wrinkle in time asymmetric valuation of past and future events. *Psychological Science*, 19(8), 796-801.
- Dasgupta, P. (2008). Discounting climate change. *Journal of Risk and Uncertainty*, 37(2-3), 141-169.
- Dougherty, T. (2011). On Whether to Prefer Pain to Pass. *Ethics*, 121(3), 521-537.
- Frederick, S. (2006). Valuing future life and future lives: A framework for understanding discounting. *Journal of Economic Psychology*, 27(5), 667-680.
- Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2), 351-401.
- Goodin, R. E. (1982). Discounting discounting. *Journal of Public Policy*, 2(01), 53-71.
- Green, L., Fisher, E., Perlow, S., & Sherman, L. (1981). Preference reversal and self control: Choice as a function of reward amount and delay. *Behaviour Analysis Letters*.
- Greene, P., Sullivan, M. (2015). Against Time Bias, *Ethics*, 125(4), 947–970.
- Goodin, R.E. (1982). Discounting discounting, *Journal of Public Policy*, 2(1) (1982), 53-71.
- Hare, C. (2008). A Puzzle about Other-directed Time-bias. *Australasian Journal of Philosophy*, 86(2), 269-277.
- Hare, C. (2013). Time—The Emotional Asymmetry. *A Companion to the Philosophy of Time*. In: *A Companion to the Philosophy of Time*, H. Dyke, A. Bardon (eds). Oxford: Wiley, 507–520.
- Hume, D. (2003). *A treatise of human nature*: Courier Dover Publications: New York.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4(6), 401-405.
- Loewenstein, G. F. (1988). Frames of mind in intertemporal choice. *Management Science*, 34(2), 200-214.
- Loewenstein, F. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological review*, 100(1), 91.
- Loewenstein, G., & Sicherman, N. (1991). Do workers prefer increasing wage profiles? *Journal of Labor Economics*, 67-84.
- Lowry, R., & Peterson, M. (2011). Pure Time Preference. *Pacific Philosophical Quarterly*, 92(4), 490-508.
- Moller, D. (2002). Parfit on pains, pleasures, and the time of their occurrence. *Canadian journal of philosophy*, 32(1), 67-82.



- Moller, D. (2006). Should we let people starve—for now? *Analysis*, 66(291), 240-247.
- Murray, C. J. L., Lopez, A. D., & Mathers, C. D. *Assessing health needs: The global burden of disease approach*. New York, NY: Oxford University Press.
- Palacios-Huerta, I. (2003). Time-Inconsistent Preferences in Adam Smith and David Hume. *History of Political Economy*, 35(2), 241-268.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
- Parfit, D. (2001). Rationality and reasons. In: *Exploring practical philosophy: From action to values*, D. Egonsson, J. Josefsson, B. Petersson, T. Rønnow-Rasmussen (eds), London: Ashgate, 17–39.
- Parfit, D. (2011). *On what matters* (v. 1). Oxford: Oxford University Press.
- Pearce, D., Groom, B., Hepburn, C., & Koundouri, P. (2003). Valuing the future. *World Economics*, 4(2), 121-141.
- Rawls, J. (1999). *A theory of justice*. Cambridge: Belknap Press.
- Sidgwick, H. (1907). *The methods of ethics*. London: Hackett Publishing.
- Smith, A. (2010). *The theory of moral sentiments*. London: Penguin.
- Stern, N. N. H. (2007). *The economics of climate change: the Stern review*. Cambridge: Cambridge University Press.
- Street, S. (2009). In defense of future tuesday indifference: Ideally coherent eccentrics and the contingency of what matters. *Philosophical Issues*, 19(1), 273-298.
- Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, 8(3), 201-207.
- Tinghög, G. (2012). Discounting, Preferences, and Paternalism in Cost-Effectiveness Analysis. *Health Care Analysis*, 20(3), 297-318.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
- Velleman, J. D. (1991). Well-being and time. *Pacific Philosophical Quarterly*, 72(1), 48-77.
- Williams, B. A. O. (1981). *Internal and external reasons*. In: *Moral Luck. Philosophical Papers 1973-1980*, Cambridge: Cambridge University, 101–113.
- Wilson, J. (2012). Paying for Patented Drugs is Hard to Justify: An Argument about Time Discounting and Medical Need. *Journal of Applied Philosophy*, 29(3), 186-199.
- Żuradzki, T. (2012). *Internalizm i eksternalizm w metaetyce*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Żuradzki, T. (2015). Trzeci rodzaj normatywności - wymóg reagowania na racje, *Filozofia Nauki*, 88(4), 35-51.