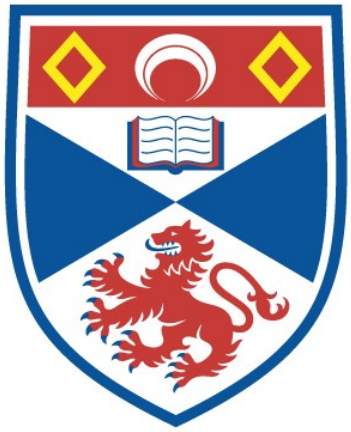
No tears for spilt milk:
temporal neutrality and the rationality of future-bias

Anh-Quân Nguyen

A thesis submitted for the degree of PhD
at the
University of St Andrews

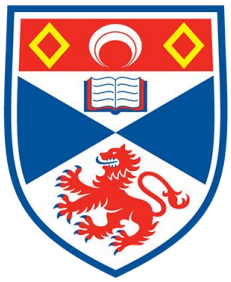


2021

No Tears for Spilt Milk

Temporal Neutrality and the Rationality of Future-bias

Anh-Quân Nguyen



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

September 2020

Candidate's declaration

I, Anh Quan Nguyen, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 66,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2016.

I, Anh Quan Nguyen, received assistance in the writing of this thesis in respect of language and spelling, which was provided by Svenja Niederfranke.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

4.01.2021

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

4.01.2021

Date Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the

University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Anh Quan Nguyen, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

4.01.2021

Date Signature of candidate

4.01.2021

Date Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Anh Quan Nguyen, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

4.1.2021

Date Signature of candidate

Acknowledgements

I would like to thank Meghan Sullivan for providing me access to an early unpublished draft of her book *Time-Biases: A Theory of Rational Planning and Personal Persistence,* as well as David Brink, Dale Dorsey, Antti Kauppinen, Thomas Hurka, and Preston Greene for taking the time to discuss my arguments over email, at conferences, and talks. Even though all disagreed with most of what I say in this thesis, they were thoughtful, generous, and encouraging in their discussions.

I would also like to thank Rebecca Rühle, Joe Slater, Madeline Hyde, Nick Allen, Lisa Bastian, Deryn Thomas, and Christine Bratu for their comments and suggestions on a final draft of this thesis, as well as Svenja Niederfranke for her invaluable help with the final editing of this work.

I am greatly indebted to my second supervisor Simon Prosser, who has provided comments, support and suggestions for my writing and thinking that have broadened my philosophical horizon beyond what I had expected to explore in this thesis.

My deepest thanks go to my first supervisor Theron Pummer for his patience, his academic and non-academic support, and his challenging, but insightful and encouraging thoughts, comments and arguments. Writing this thesis would not have been possible without his help, and I am a better writer and philosopher thanks to him.

I would also like to thank the PhD community of the St Andrews and Stirling Philosophy Graduate Programme, especially the members of the Minorities and Philosophy (MAP) Group, for their support and solidarity. I would also like to acknowledge the Mlitt Philosophy Class 2015-2016, who have remained a friend and support group long after graduation. I am also indebted to the Heinrich-Böll Foundation and its PhD research community and would like to thank Sevilay Karaduman as my scholarship advisor for her encouragement and support over these years.

Special thanks go to my office mates Lisa Bastian and Sara Vikesdal, who accepted my whining with few complaints only, to my flatmates in Flat 9 for enduring me for years, and to Claire Fogarty for publishing my woes about this thesis online.

Funding

This work was generously supported by the Heinrich-Böll Foundation with a research scholarship and conference support. Scholarship Number: P124686

Thesis Abstract

Temporal neutrality has become widely accepted as a rational requirement for agents, due to recent arguments from Sullivan (2018), Greene and Sullivan (2015), Dougherty (2011, 2015) and Brink (2010), which build on older remarks from Rawls (1971), Parfit (1984), and Sidgwick (1874). According to Brink, an agent is temporally neutral if she does not prefer an event over another solely based on temporal location and gives equal significance to all parts of life. Temporally neutral agents are required not to be time biased.

There are two forms of time-biases in the debate, near-bias and future-bias. An agent is near-biased if she prefers positive events to be closer to her and negative events to be further away in the future. An agent is future-biased if she prefers positive events to be future rather than past, and negative events to be past rather than future. Both forms of time-biases are seen as irrational by proponents of temporal neutrality due to (1) concerns of arbitrariness, (2) concerns around pragmatic loss, and (3) irrelevant influences.

While I accept near-bias as rationally impermissible, I develop a systematic defence of the rationality of future-bias in this thesis. My thesis provides defences against all types of arguments from temporal neutralists and aims to set out rational grounds for being future-biased based on rational agency, control-asymmetry, and emotional prudence.

Additionally, I explore implications of future-bias for some moral theories. I argue that we should understand future-bias as a comprehensive preference pattern that includes all evaluative aspects of life. If we accept a comprehensive reading of future-bias, this undermines some moral theories by making it permissible to focus moral evaluation on the present and future.

In short, my thesis argues that we are permitted to, and sometimes should be future-biased, and should reassess our moral theories accordingly.

Table of Contents

1 Introduction: Temporal Neutrality and Time-Biases

As rational and moral agents, we often face two competing temporal perspectives in our rational and moral commitments.

On the one hand, we are temporally embedded agents that evaluate our rational and moral commitments from a point of view in time which we call the present moment. In this moment, we face events that are ahead of us and can look back to events that have already passed. Additionally, it makes sense for us as temporally embedded agents to look back at our past selves and what we thought, wanted, believed and valued, and draw a distinction between what our reasons, desires and preferences were in the past and what they are now. What I was like as a rational and moral agent five years ago is so markedly, radically different to who I am now that it seems justified to draw a line.

On the other hand, we understand ourselves as temporally extended agents that persist over time, and with our agency, our rational and moral commitments persists with us. We may think that we are, to some extent at least, the same person today as we were yesterday, the week before, or five or ten years ago. Our past, present and future selves seem to be just snapshots of us being agents over time and should not influence our rational and moral commitments too much if at all, as our past, present and future are all parts of the same agent.

Part of these competing temporal elements are expressed directly in some of our rational and moral commitments as requirements constraining our preferences and choices. Philosophers like John Rawls and Henry Sidgwick, and more recently Meghan Sullivan, Preston Greene, Tom Dougherty and David Brink have argued that, as temporally extended agents, we ought to be temporally neutral. Temporal neutrality is a moral and rational requirement that asks us to not place any normative significance solely on the temporal location of goods and harms. Furthermore, temporal neutrality requires us to treat all parts of a life as equally significant, whether it is past, present or future.[1]

Another way of describing temporal neutrality is as a requirement for us not to be time-biased. A time-bias is a preference pattern we exhibit when we prefer goods or harms to be scheduled in a way that is sensitive to their temporal location, e.g. when you would like to have all the bad things in your life to be scheduled on Tuesdays, you are Tuesday-biased and not temporally neutral. Less absurdly, and much more common, are the *bias towards the near* and the *bias towards the future*.

If you are near-biased, you prefer good things to be scheduled closer to you in time, and bad things to be further away in the future. For example, you might prefer your holiday to be scheduled next week rather than next month, even if it would be just as enjoyable either way – you would even pay extra to some extent to have your holiday next week, as a good is worth more to you if it is temporally nearer to the present moment. On the

[1] See Brink (2010), p.1.

other hand, you would rather have your next appointment at the dentist next month rather than next week, or next year if possible, as you prefer bad things to be farther away in the future rather than closer to the present.

If you are future-biased, you prefer bad things to be behind you in the past, and good things to be in the future, expressing an asymmetry in how you evaluate past and future events. You would rather have your dentist appointment in the past rather than in the future, even if your past appointment would be more painful, as it would be over and done with, and not ahead of you. And you would much more like to have your holiday ahead of you in the future rather than behind you in the past.

Both near-bias and future-bias (and all other time-biases) are irrational according to proponents of temporal neutrality. These two time-biases are expressions of our temporal embeddedness, as we evaluate events, good or bad, from our personal point of view in time. Temporal neutrality, on the other hand, stems from our sense of temporally extended agency that tells us that past, present and future belong to us equally. If we fully appreciate that we are temporally extended agents, we will come to embrace temporal neutrality, and recognize that both near- and future-bias are irrational attitude patterns.

1.1 The Rationality of Future-Bias

This dissertation aims to defend the rationality of future-bias against arguments from temporal neutralists. I will provide both a rebuttal to existing arguments that show that future-bias is irrational as well as a justification for future-bias being a

rationally fitting, coherent, and beneficial attitude pattern for us to hold. Additionally, the thesis will explore future-bias as an attitude pattern, how to best conceptualise it, and what implications this holds for moral theories. I will argue that we should understand future-bias in a strong, comprehensive way that includes not only pleasures and pains but all normative aspects of life: if you are future-biased, your past reasons will be trumped by present and future reasons. Given that future-bias is rationally permissible, this will undermine moral theories relying on diachronic links between different parts of life, such as lifetime egalitarianism and narrative ethical theories.

The thesis will unfold as follows. In the rest of this introductory chapter, I provide an overview of the different arguments employed by proponents of temporal neutrality by explaining how they apply to near-bias. As most arguments against the rationality of future-bias are attempts to extend the arguments against near-bias to future-bias, this will be helpful to outline the structure of the concerns surrounding time-biases generally as well as the overall appeal of temporal neutrality.

In chapter 2, I discuss the intuitive foundations of future-bias's rationality and argue that attempts to undermine those fail. Proponents of temporal neutrality such as David Brink and others tried to show that while future-bias seems intuitively permissible in self-regarding cases concerning hedonic goods and harms, this intuitive appeal disappears when applied to other-regarding cases or non-hedonic goods and harms. As these cases are relevantly similar according to Brink and others, we ought to revise our intuitions supporting future-bias's rationality, or at

least not treat them as evidence for its permissibility. I show that the argument fails on several grounds, by showing that future-bias is more widespread than friends of temporal neutrality acknowledge and can appear intuitively permissible even in non-hedonic cases, as well as pointing out a systematic flaw in the temporal neutralists argument that causes it to either fail or backfire against temporal neutrality.

In chapter 3, I explore the so-called arbitrariness-argument proposed by Meghan Sullivan, who builds on remarks from John Rawls and Henry Sidgwick to show that future-bias lacks rational grounds and is not a result of rational processes. As spatial location between two goods and harms is an arbitrary difference, it should not factor into our evaluative assessments of goods and harms, and as temporal location of goods and harms are relevantly similar to spatial location, future-bias is irrational on the grounds of arbitrariness, according to Sullivan. I will not only show that the argument can be resisted, but aim to suggest a new way of how future-bias is a result of rational processes due to the attitude pattern being rationally grounded in our sense of agency.

Chapter 4 examines attempts by Meghan Sullivan and Preston Greene to show that evolutionary influence on our tensed emotions give rise to future-bias, and as a result, we should reduce our confidence in its permissibility. As soon as we realise the root cause of future-bias, we should use our capabilities of rational reflection to weed out irrelevant influences like evolution and emotions and move towards a temporally neutral perspective. After shortly rebutting these arguments, I then proceed to show that evolutionary influence and tensed emotions can instead

provide rational grounds for future-bias's rationality, as future-bias evolved to track a control-asymmetry between past and future events. As we cannot influence the past, but sometimes can influence the future, future-bias is both fitting and beneficial to us as rational agents.

In Chapter 5, I focus on attempts by Meghan Sullivan, Preston Greene, and Tom Dougherty to show that being future-biased will lead to pragmatic loss. They argue that having this attitude pattern will cause us to at least sometimes choose worse options, leaving us worse off overall. The authors argue that being future-biased in combination with being risk-averse (Dougherty) or regret-averse (Sullivan and Greene) leads us to making irrational choices leading to a worse life overall. I will discuss ways of resisting these arguments, but also explore why I remain unconvinced that they lead to the overall rational impermissibility of future-bias even if I admit that sometimes, future-bias may lead to pragmatic losses. I additionally provide a reversed version of the argument showing that, being temporally neutral in combination with changing preferences will commit agents to rationally problematic sunk-cost cases. As sunk cost fallacies are much more common than the creative but surreal cases provided by Dougherty, Sullivan, and Greene, being future-bias will be better for us than temporal neutrality.

Chapters 6 and 7 do not deal with arguments for and against the rationality of future-bias directly, but rather explore what it means to be future-biased, and what this means for moral theories. In Chapter 6, I set out why we ought to understand future-bias as a comprehensive attitude pattern that not only

includes past events, but our past parts of life. I then show that if we understand future-bias like this, it makes more sense for us to treat the discounting involved with future-bias as lexical priority of future over past events, and not as an absolute discount function. Finally, I show that this undermines whole-life egalitarianism. Chapter 7 continues this exploration by examining how being future-biased will undermine ethical theories that rely on narrative connections between different parts of life to determine what we ought to do.

Each chapter is written in a way that it is accessible on its own, while contributing to my overall aims, which are to provide a comprehensive defence of the rationality of future-bias, to show that it holds implications for the way we conceptualise moral theories, and that temporal neutrality should not be as universally accepted as some authors think it is.

In the rest of this chapter, I will explore the appeal and the arguments in favour of temporal neutrality by explaining how they apply to near-bias. I will ultimately remain neutral on whether these arguments against near-bias succeed, even though I am sympathetic to the view that near-bias is irrational. But I hope to show with my thesis that it is justified to treat near-bias and future-bias asymmetrically, and that it is rational for us to focus on the future over the past.

1.2 What about Near-Bias?

Imagine that you're a young person considering going to university in the US, maybe studying something like philosophy, which seems somewhat interesting to you. After looking into it,

you become disillusioned by the amount of tuition fees you would have to pay to gain a degree that mostly won't pay off later in economic benefits. You don't have more than $100 000, aren't able to access any scholarships and are resigned to not going to university.

But you're in luck! You are offered a loan by some banks, who will provide you with money now to pay your tuition and cover your living costs. The only thing you have to do is to pay it back afterwards with interest, so on average 6-9% more of what they gave you in the first place. Should you take up that offer?

There is increasing evidence that this is a bad deal.[2] What makes it more likely for you to accept this, however, is if you don't particularly care about the far future, and prioritise the near future. While you might accept the offer based on other considerations, e.g. you could think that a philosophy degree is worth more than can $100 000, or that you estimate that you will be in a better position to pay after graduation, but caring more about the near than the far future will make you more likely to take on student loans.

Let's restate what near-bias is more formally:

> Near-Bias: An agent is near-biased iff for two exclusive future events E_1 and E_2,

[2] See Best & Keppo (2011) for an overview of the costs of university. See Cooke et al. (2004) for an overview on the effects of student loans on mental health.

where E_2 is at least as positive as E_1, the agent prefers E_1 because it would occur earlier to now than E_2, or

where E_2 is at least as negative as E_1, the agent prefers E_2 because it would occur later to now than E_1.

Near-bias is the most common time-bias, sees wide applications in psychology, economics, neurobiology, and is known outside of philosophy in the wider scientific literature simply as a time preference. Note that it is important to distinguish between temporal discounting, which broadly describes any behaviour that leads agents to care less about future consequences, and time preference, which is a genuine time-bias that describes a preference for nearer utility over delayed utility.[3] Temporal discounting may include discounting the far future based on uncertainty, or genuine value differences between immediate and future utility such as incommensurability of immediate survival and long term benefits. You wouldn't really be near-biased if you're in poverty and need to focus on the immediate future rather than distant future goods, as nearer and future utility would be genuinely different and not comparable to nearer utility for you. Time preference, on the other hand, describes near-bias as a genuine temporal preference between nearer goods and harms and far-future goods and harms that can be compared properly.

[3] For an overview, see Frederik, Loewenstein, and O'donoghue . (2002).

Near-bias is usually represented in the literature as a hyperbolic discount function in which the values of events decrease the more they are scheduled into the future. A hyperbolic discount means that the value of a future event will decrease at a slow and steady rate for some time before spiking and decreasing very fast. In contrast to that, an exponential discount function would describe a steady and constant decrease in value for future events. There is quite a lot of evidence in the literature that we are hyperbolic and not exponential discounters.[4]

In what follows, I will summarise the three arguments advanced against the rationality of near-bias, based mostly on work by David Brink and Meghan Sullivan. After that, I will discuss whether the arguments can be resisted, and the rationality of near-bias be defended. While I am somewhat sympathetic to neutrality about the future, I will examine several ways forward for opponents of temporal neutrality.

1.3 The Arbitrariness Argument

The first argument against the rationality of near-bias is the so-called arbitrariness argument, first defended in detail by Henry Sidgwick. He writes in *The Method of Ethics* that

> "Hereafter as such is to be regarded neither less nor more than Now. [...] the mere difference of priority and posteriority in time is not a reasonable ground for having

[4] See for example, Lowenstein and Elster (1992) as well as Green and Myerson (2004). See also Sullivan (2018), pp. 12-17 for a good overview.

more regard for the consciousness of one moment than to that of another."[5]

This was later taken on by John Rawls in *A Theory of Justice*:

> "The different temporal position of persons and generations does not in itself justify treating it differently. [...] There is no reason for the parties to give any weight to mere position in time."[6]

This is spelled out by Sullivan to a fully developed argument:

Arbitrariness-Argument against near-bias[7]

(1) It's not rationally permissible to have preferences sensitive to arbitrary differences. (Non-Arbitrariness Principle)
(2) Relative temporal distance from the present is an arbitrary difference between two future events.
(3) If an agent is near-biased, her preferences are sensitive to mere temporal distance between events relative to the present.
(4) Therefore, near-bias is not rationally permissible.

(3) follows from our definition of near-bias above. (1) and (2) will be more controversial. Why should we believe (1)? Sullivan appeals to cases involving other seemingly arbitrary differences, such as a mere difference in physical location to argue for (1).

---

[5] Sidgwick (2019), p. 380.

[6] Rawls (1971), p. 259.

[7] Sullivan (2018), p. 36, Brink (2010), p. 4.

> Detergents: My local grocery store stocks three kinds of detergent: Wisk, Surf, and Tide. Each is composed of the same cleaning agents, and I'm aware of this. Wisk and Surf are stocked on the same shelf at about waist height. Tide is one shelf above them at eye level. Tide is more expensive than either of the other brands. I'm indifferent between Wisk and Surf but prefer Tide to either of the other detergents. In fact, I'm willing to pay the difference to get Tide and regularly choose that brand.[8]

If my shopping were to be like this, Sullivan would criticise me for my preference pattern, as the mere difference in physical location is no reason for preferring one thing over another. If Sullivan were to ask me about my shopping preferences, and I'd reply, "It's because it's on the shelf that meets my eyes and the others aren't.", she would dismiss it as a response, as it's not being a reason at all.

Another case Sullivan uses in support for the arbitrariness-principle is the Future-Tuesday case by Derek Parfit.

> Future Tuesdays: A certain hedonist cares greatly about the quality of his future experiences. With one exception: he has Future-Tuesday-Indifference. Throughout every Tuesday he cares in the normal way about what is happening to him. But he never cares about possible pains or pleasures on a future Tuesday. Thus, he would choose a painful operation on the following Tuesday rather than

[8] Sullivan (2018), p. 38.

> a much less painful operation on the following Wednesday. This choice would not be the result of any false beliefs. The man knows that the operation will be much more painful if it is on Tuesday and agrees that it will be just as much him who will be suffering on Tuesday, and knows that Tuesday is merely part of a conventional calendar with an arbitrary name. Nor has he any other beliefs that might help to justify his indifference to pain on future Tuesdays. This indifference is a bare fact. When he is planning his future, it is simply true that he always prefers the prospect of great suffering on a Tuesday to the mildest pain on any other day.[9]

Future-Tuesday indifference is a time-bias, albeit a very obscure and eccentric one that seems a lot less natural than near-bias. But it seems irrational for a hedonist to prefer pain to be scheduled on a Tuesday just because it is a Tuesday, as it would be arbitrary to do so and "being Tuesday" as such is no rational ground to base a preference on.

Sullivan, together with Finco, also uses the cheerleader effect to demonstrate the arbitrariness-principle.

> Cheerleader Effect: Ted is at a singles bar with his friends. At one end of the bar he sees a woman, Amy, who is enjoying an amaretto sour by herself. At the other end of the bar he sees Trudy, who has just been given her amaretto sour. Being perhaps too shallow, he finds Amy

[9] Shortened from Parfit (1984), pp. 123-4.

> more attractive than Trudy and so prefers to go talk to her. He decides to introduce himself to Amy after buying himself another drink.
>
> In the interim, Trudy has re-joined her friends. Ted glances in her direction on the way to Amy. Ted's judgment changes when he sees her with a group. He now finds her more attractive and prefers to go talk to her. With no shame Ted pivots and heads toward Trudy's table.[10]

While this is a fairly common phenomenon, it is at best questionable dating behaviour, and Ted should probably reconsider his approach to finding a partner significantly. One ground on which we could help him improve his game is to advise him that it should be irrelevant to a person's attractiveness whether they are alone or not. Basing a preference on whether a person is in a group seems like arbitrary grounds, for dating and otherwise.

Note, however, that, while these cases are fairly convincing in supporting the arbitrariness-principle, they don't really spell out what exactly amounts to an arbitrary difference generally. Without a proper recipe for what counts as arbitrary or non-arbitrary difference in cases, it is quite possible to resist them.

For example, Sharon Street rejects (1) by providing a defence of the most obscure of these cases, and argues that ideally coherent eccentrics like our Future-Tuesday hedonist are not being

[10] Sullivan/Finocciaro (2016), pp. 142-3.

irrational as long as their belief system stays consistent.[11] Why is it relevant to him to schedule pains on a Tuesday? It might be, so Street, that his ancestors evolved in a way that made it more likely to survive if pains were scheduled regularly on a certain date – imagine a background origin story that shaped his beliefs and preferences in a way that makes it understandable and coherent for him to be indifferent about future Tuesdays. Then, it wouldn't be arbitrary for him to prefer as he does. And while it might be bad in terms of pragmatic considerations – the Tuesday-man could be scheduling more pains in his life than without his Tuesday-indifference - the arbitrariness of his preference wouldn't be a bad thing in itself. In a similar manner, we could provide equally compelling evolutionary stories for Detergents and Cheerleader effect (which quite likely is actually a result of evolutionary pressure) – as long as we are being coherent and no harm is done, why should arbitrariness be a concern?

Sullivan, for this reason and to avoid infinite regresses about reasons, endorses externalism about prudential reasons: some facts just are normatively significant to agents, regardless of what they happen to desire or prefer, and some aren't, without the need for further justification.[12] With this, we can put aside Street-style arguments for a while, and while this still leaves open what exactly counts as arbitrary, we can get the arbitrariness-argument against near-bias off the ground.

---

[11] Street (2009).

[12] Sullivan (2018), p. 45.

1.4 The Compensation Argument

The second kind of argument against the rationality of near-bias is called the compensation-argument, and the idea is simple: you will get compensated in terms of utility if you're not near-biased. To put it differently, agents when choosing between two different future events will make good trades that will benefit them more if they are temporally neutral. Say, for example, that I managed to save 10 000€ from my grad student stipend. I could choose to spend it on a few months holiday in the near future to visit my relatives in Vietnam. I could also put it into a savings account that will (assuming money and banks will still exist in the far future) provide me with a financial safety net with interests when I am older. If I am near-biased, I will be inclined to choose the former option, even if the latter option will be better for me in lots of ways. Sullivan outlines the argument like this:

Compensation-Argument against Near-bias[13]

(1) A rational agent prefers her life to go forward as well as possible.

(2) If you are near-biased, you will choose earlier lesser goods over later greater goods just because of them being earlier.

(3) Your life would go better if you chose the later, greater good over the lesser, nearer good.

---

[13] Sullivan (2018), pp. 22-23. "Life going forward" is borrowed from Sullivan herself, as the so-called success-principle. If you feel that this stacks the deck against temporal neutrality, think of it as

(1) A rational agent prefers her life to be as good as possible.

(4) Therefore, if you're near-biased, your life will not go forward as well as possible.

(5) Hence, a rational agent would not be near-biased.

Note that the argument is neutral on what constitutes goodness or well-being. Whether you are a hedonist, hold a preference-satisfaction view or adhere to some kind of objective list of well-being, you should accept (1), as "going forward as well as possible" can be achieved under all of those theories. Note also that this argument also doesn't rule out what Sullivan calls "Bookstore Buddhism", which is a life-in-the-moment rule asking you to focus and enjoy the present instead of worrying too much about the future.[14] This view is consistent with (1), as your life wouldn't go forward as well as possible if you were to lose well-being by worrying about the future all the time. (1) is also consistent with structuralist views on a good life: some theories, e.g. narrative ethical theories, think that a life is better overall if its parts are ordered in a certain way that provides narrative unity. This is consistent with both (1) and temporal neutrality in general, as we wouldn't prefer things to be scheduled at a certain time just based on the temporal location, but because of its contribution to narrative or structural well-being.[15]

While (1) can be questioned on Humean grounds – not every rational agent might want her life to go best – I find the argument fairly convincing. It is simple, straightforward, and is applicable to a wide range of cases, from individual saving problems to

[14] Sullivan (2018), pp. 26-28.

[15] See Brink (2010).

collective decision making. There is an issue here that arises when we try to apply the compensation argument to intergenerational justice – we should not prioritise the closer future generation over the far-future generation – as it isn't clear who is being compensated by whom. To solve this issue, we would need additional tools that appeal to personal identity, theories of time, or moral principles supporting the compensation argument. The topic, however, goes beyond the scope of my thesis.[16]

1.5 Evolutionary Influence

Let's assume that the arguments from arbitrariness and compensation are both successful. It is very likely that your behaviour will not change a lot, as it is generally very hard to get rid of cognitive biases, and because time-biases in particular are very deep seated and as a result it is hard to fully believe them to be irrational.[17] However, Greene and Sullivan offer an explanation why this is the case, and with that explanation, we should be more able to move towards temporal neutrality.

Why are we near-biased? Because we have evolved to be sensitive to probabilities in the future, according to many philosophers and psychologists. We developed near-bias, according to Greene and

---

[16] For interesting arguments on this, see Beckstead (2013), and Caney (2020), section 3.

[17] For reasons why it is hard to get rid of cognitive biases, see Kahnemann (2013), who explains that, our capacities for rational reflection (System 2) are fairly limited and cost-intensive, and are as a result usually not used as much as they could be. Very often, they merely affirm what our cognitive intuitions and hunches tell us very quickly (System 1).

Sullivan, because it helps us track uncertainty about the future, and how likely it is that we will benefit from one event over another.[18] If an event is far away in the future, we cannot as easily assume that it will take place, while a near future event will create less uncertainty in our mind. While this may have been very useful generally for our ancestors, the problem is, according to Greene and Sullivan, that the heuristic "backfires in situations in which there is a significant divergence between probabilities and our rate of discounting." For example, you might have a discount rate of 50% for events that are in a year: a necessary but painful operation will be evaluated to be half as bad for you compared to an equally painful operation next week. However, (assuming your healthcare system is somewhat reliable) there is no uncertainty about the pain, and maybe just a 10% chance of the painful operation to be rescheduled, and as such, your 50% discount rate doesn't track the actual uncertainty and probabilities.

Greene and Sullivan give us a hint as to why evolutionary influences and not rational processes are responsible for us being near-biased: a lot of tensed emotions are associated with near-bias. We use anxiety about future event as a heuristic to track uncertainty about the future, and these emotions are not the best guides to what a rational agent would do. Greene and Sullivan provide us with an example:

> "Imagine that you are given a choice between undergoing a moderately painful dental surgery tomorrow or delaying

[18] See Greene and Sullivan (2015), p. 966.

> the surgery for a year, risking the problem becoming worse. Prudently, you choose to make an appointment for tomorrow. However, you may find yourself thinking about tomorrow's surgery, and even feeling anxiety, and this anxiety may build as the time of the appointment draws nearer. At the same time, you know that if you had scheduled the distant surgery, you would not be anxious. Does the fact that you would feel anxiety only about the near surgery show that the distant surgery is preferable to the near one?"[19]

Greene and Sullivan think that this is obviously not the case – following your anxiety may be counterproductive to your well-being and as a result irrational. There is a reason, Greene and Sullivan argue, why psychologists are researching ways to regulate our emotions like anxiety, so that we are not slaves to them and can follow our rational reflection instead.

I believe this argument to be mistaken, as I think Greene and Sullivan mischaracterise the rationality of emotions, as well as evolutionary influence, but I won't engage with it here. I will argue in chapter 4 that there is a good argument to be made as to why both evolution and emotions may provide rational grounds for being time-biased, and future-biased in particular.

[19] Greene and Sullivan (2015), p. 967.

1.6 Can there be a Defence of Near-Bias's Rationality?

The irrationality of near-bias is usually accepted quite widely, but a few philosophers have attempted to provide a defence of its permissibility. In this last section, I will explain and discuss arguments provided by Caspar Hare and Dale Dorsey and argue that they are on their own not sufficient to overcome the arguments against the rationality of near-bias.

Hare uses the so-called A-theory of time to justify time-biases via metaphysical asymmetry. The A-theory states that there is something metaphysically different about the present, that "now" cannot be described in terms of earlier or later moments only. This contrasts with the so-called B-theory, or four-dimensionalism, which states that there is no fundamental metaphysical difference between past, present, and future. He argues that "If we accept four- dimensionalism, then [the problems of justifying time-biases] really are insoluble." [20] According to Hare, each A-theory provides an account of "all that there is"[21], so what the maximal state of affairs is like. And if we accept the A-theory, one feature of the maximal state of affairs is that present things are different from past and future things. So other things being equal, near future negative events are just

[20] Hare (2009), p. 16.

[21] Hare (2009), p. 17.

worse than far future negative events because near future negative events exist more.[22]

Thereby, we arrive at a justification for near-bias:

> Harmony: When a near-biased person favours a near future scenario over a far future scenario, she thereby simply favours a better maximal state of affairs.[23]

This is supposed to explain both why near-bias is not arbitrary, as the far future is simply less real, as well as why the compensation argument does not apply, as far future utility is simply less real than near-future utility.

I'm not convinced that this is a strong defence. Firstly, this defence of near-bias doesn't come cheap and needs to face all objections against the A-theory, including concerns about compatibility with special relativity, and at what speed the passage of time is supposed to take place. Also, Hare endorses not only the A-theory, but one of the most radical versions of it, a (mild) form of presentism, according to which the presence exists "more" than the past and future. This is not universally accepted amongst all a-theorists. For example, Sullivan herself endorses an A-theory that treats past present and future as equally existent.[24] That is not to say that Hare's version of presentism won't turn out to be the correct theory of time, but until that is established, this severely limits Hare's call for harmony.

---

[22] Hare (2009), p.18.

[23] Amended for just near-bias from Hare (2009), p.10.

[24] See Sullivan (2012).

Secondly, to really establish near-bias as we defined it, as a hyperbolic discount function, Hare's presentism would not only need to establish that the present exists "more" than the future, it would need to establish that the future gradually comes into existence, resembling a hyperbolic function. So, while near future events are gradually less existent, at some point in time, they would suddenly have to become less existent in a much faster way. I have no idea how a plausible version of presentism could develop a theory of time resembling that. Standard versions of presentism that claim that past and future simply do not exist would also not be able to justify near-bias in the way Hare wants it to, as both the near-future and the far-future would be equally non-existent. Presentism would be able to explain a bias towards the present, which Parfit discusses briefly, but is for our purposes irrelevant.[25]

Dale Dorsey suggests a different way of providing a rationale for near-bias, even though he ends up rejecting it himself. According to a lot of classic moral theories, we have special reasons to favour family and friends over strangers. While some philosophers like Singer and Godwin may deny this, most accept that those who are close to us stand in a particular relation to us that is morally significant and generates reasons for us to prefer them over those we do not stand in a relation with.[26] Caring and valuing these relations are at least sometimes non-instrumental, and are value-generating in a way that cannot be captured by impartial moral

[25] See Parfit (1984).

[26] See, for example, Scheffler (1994), Kolodny (2003).

reasoning, and can therefore provide special normative significance to those we care about in this way.

Dorsey argues that if "what we owe to each other can be shaped by the structure of our interpersonal bonds of concern, it's not clear why what we owe to ourselves cannot similarly be structured by our intrapersonal bonds of concern."[27] So according to Dorsey, our current self sometimes has similar bonds to our future selves just as we have bonds with family and friends, and that can sometimes justify near-bias. To illustrate the idea, the relationship between my present self and my one month in the future self might be a relationship of care, while the relationship between my present self and my one year in the future self is not – even if I know equal amounts about both future selves. Or in Dorsey's words:

> "My moral perspective will be different than yours, insofar as we bear such bonds to different people. But if we allow such considerations to cross the intrapersonal barrier, prudential evaluation becomes similarly perspectival. My prudential perspective, now, may very well be different than my yesterday's prudential perspective or my tomorrow's prudential perspective, insofar as I, today, may have very different intrapersonal bonds than I will tomorrow. And given plausible, and common,

[27] Dorsey (2019), p. 464.

> psychological assumptions, for many these prudential perspectives will be biased toward the near."[28]

So, if I have a strong bond to my near future selves and a weaker bond to my far future selves, this provides a rationale for being near-biased.

While this defence has an advantage over Hare insofar as it does not rely on sketchy assumptions of metaphysics of time, the defence does not hold. Dorsey himself rejects his own suggestion because it would end up justifying problematic patterns of reactive attitudes.[29]

I don't believe, however, that we need to go this far to detect problems with Dorsey's suggestion. There are quite obvious asymmetries between partiality concerns to family and friends, and bonds to future selves. Firstly, bonds of partiality are usually built through special kinds of interactions.[30] We develop special concerns with loved ones by interacting with them in a certain way. For example, two people will become friends by building trust, interests, shared values, and so on. This is not the case for our present and future selves whose abilities to interact are severely limited. While our present selves can probably make reasonable predictions about our future selves (I can predict that my future self will need to find a new flat in a month), this does not provide the same kind of shared bond as an interpersonal

---

[28] Dorsey (2019), p. 464.

[29] See Dorsey (2019), pp. 473-475.

[30] See for example, Brink (2001).

relation due to the lack of any kind of interaction. Secondly, Dorsey's suggestion also doesn't really provide an explanation *why* it is permissible or appropriate for us to care more about our near-future selves than our far-future selves. Even if we grant that there can be special bonds between our present and near-future selves, it isn't an explanation why we ought to care more about our near-future just because we care about it more. Admittedly, this is an issue in the debate with interpersonal moral concerns too – it is difficult to explain why special concerns for our families and friends should be special moral considerations. But in contrast to near-bias, partiality to friends and family has such an intuitive weight to it that going against this renders moral theories implausible for a lot of people.[31] A moral theory that tells us to let our loved ones die in favour of strangers will fail to accommodate our moral intuitions in a way that will make us abandon the theory in practice. This is not the case for near-bias: Our concern for our near-future selves does not carry the same intuitive weight to be justifying special concern in itself without further argument. Hence, Dorsey's suggestion should be rejected.

So where does this leave the rationality of near-bias? I am inclined to believe that proponents of temporal neutrality are generally right to say that concern over our near-future over our far-future is not justified, and that we should be future-neutral. It should be noted that in some cases, near-bias will still be an advantageous disposition to have. For example, a person in

[31] For the classic arguments, see Williams (1981).

poverty will instinctively focus on the immediate future instead of caring about the long-term impact of their actions, as this will help them to survive. This is not ruled out as rationally impermissible by temporal neutrality, as caring about the far future just as much as about the immediate future would not lead to the agent's life to go best. But it shows that in a limited range of cases where there is an urgent need of focusing on the immediate future, near-bias will remain a useful heuristic for our actions.

1.7 Conclusion

In this chapter, I have outlined the positions of temporal neutrality, future-bias and near-bias, and explained how they are tied to two different temporal perspectives of agency. I have outlined the three main arguments against the rationality of near-bias: the arbitrariness-argument, the compensation-argument, and concerns about evolutionary influence. I then discussed and dismissed arguments by Hare and Dorsey, who try to provide a rationale for near-bias by appealing to the A-theory of time and partiality to near-future selves respectively. In summary, near-bias, while sometimes still a useful heuristic, should be treated as rationally impermissible.

2 Intuitions about Past and Future Value

I'm a socially awkward person who doesn't enjoy social events. Hanging out with others costs me energy and causes anxiety. Unexpectedly, someone invites me to a party! I'm too shocked to say no, so the person suggests that we either go out tomorrow, or next week. Let's say all other things are equal, like the risk of me getting ill, etc. Is it rational to prefer the party to be next week rather than tomorrow, just because it's further away in time?

If I have a preference like that, I'm time-biased. More specifically, I'm biased towards the near. In general, an agent is near-biased if, all else equal, she prefers positive future events to be nearer rather than further away, and negative future events to be further away rather than nearer.

There is also another form of time-bias. Let's say I wake up, hung-over, disoriented, and can't remember what happened yesterday. I recall that I was invited to a party, and that I reluctantly agreed to go. I run to my calendar to check when the party was scheduled – it was yesterday! (Explains the hangover.) Should I now feel relief? The social event, which caused me anxiety, is already over. It's past rather than future – I already lived through it.

If I have this attitude, I'm also time-biased, in a different way. More specifically, I'm biased towards the future. An agent is future-biased if, other things being equal, she prefers positive

events to be future rather than past, and negative events to be past rather than future.

Are these attitudes irrational? Proponents of temporal neutrality say yes. Temporal neutrality requires agents to not prefer events, goods or harms based on their temporal location per se, and to give equal significance to all parts of their lives.[32] So, whether the event is tomorrow or next week, past or future, shouldn't matter in my evaluation of the event. In short, I shouldn't be time-biased.

While it's generally agreed that near-bias is rationally impermissible, it's more controversial to make the same claim about future-bias.[33] It just seems so natural to prefer bad things to be past and good things to be future that we also tend to support future-bias's rational permissibility. However, several authors such as Brink, Dougherty and Hurka have challenged this: The intuitive appeal of future-bias's permissibility is limited to a set of isolated cases that involve only hedonic and self-regarding goods and harms. If we look at non-hedonic goods or concerns about other people, the intuitive appeal behind future-bias disappears. On the contrary, because future-bias seems rational to us only concerning self-regarding hedonic goods and harms, but not others. The authors suggest that as a result, we

---

[32] Brink (2010), p. 1.

[33] See Greene and Sullivan (2015), pp. 952-953, Sullivan (2018), p. 46, for an overview of the difference treatment near- and future-bias have received in the literature, and Heathwood (2008), pp. 56-57 for an example of a view defending future- but not near-bias.

should revise our intuitions accordingly and treat future-bias as irrational in self-regarding hedonic cases too, or at the very least not treat the intuitive appeal in these cases as evidence for future-bias's permissibility, since hedonic and non-hedonic cases are relevantly similar.

In this chapter, I will defend the rationality of future-bias against this concern. I will first outline which non-hedonic cases are most promising to advance the argument on behalf of proponents of temporal neutrality. I will then argue for two points.

Firstly, even if we concede to the friends of temporal neutrality that future-bias is intuitively permissible only in self-regarding individual cases, it does not follow that we ought to revise intuitions about our past and future pleasure and pain towards temporal neutrality: hedonic goods and harms mostly concern well-being at a time, while the proposed non-hedonic goods concern lifetime goodness. Even if momentary well-being and lifetime goodness are linked, this is a relevant structural difference between hedonic and non-hedonic goods and harms that warrant different attitude patterns, including the applicability of future-bias.

And lastly, even if we assume that hedonic and non-hedonic goods and harms were relevantly similar, the degree of idealisation required to make our intuitions favour temporal neutrality obscures the weight of our intuitions to a level that makes it unreliable as evidence against the permissibility of future-bias. As temporal neutralists need this degree of idealisation of their cases,

because otherwise we can still make a reasonable case for our intuitions to favour future-bias's rationality, the argument fails.

In summary, we can keep using intuitive support as evidence for future-bias's rationality, at least when it comes to hedonic goods, but to some extent for its general permissibility, as suspicions about the rationality of future-bias based on shifts in intuitive support are unfounded. So, next time you wish for an awful social event to be over, don't despair – that's absolutely okay.

2.1 The Best Case against Non-Hedonic Future-Bias

Let's first restate the positions more precisely.

> Future-Bias: An agent is future-biased iff for two exclusive events E_1 and E_2, with E_1 being in the past, and E_2 in the future,
>
> - where E_1 is at least as positive as E_2, the agent prefers E_2 to E_1 because E_1 is in the past and E_2 is not, or
> - where E_1 is at least as negative as E_2, the agent prefers E_1 to E_2 because E_1 is in the past and E_2 is not.
>
> Past Neutrality: An agent is temporally neutral iff for two exclusive events E_1 and E_2, with E_1 being in the past and equally good as E_2, the agent is indifferent between E_1 and E_2.

Future-bias is incompatible with temporal neutrality. Now, what seems more natural to accept as rational? Here's the classic case against temporal neutrality:

> My Past and Future Operations: I am in a hospital to have a safe, but painful surgery. Because the operation is so painful, patients are afterwards made to forget it.
>
> I have just woken up. I cannot remember going to sleep. I ask my nurse if it has been decided when my operation is to be, and how long it must take. She says that she knows the facts about both me and another patient, but that she cannot remember which facts apply to whom. I may be the patient who had his operation yesterday, lasting ten hours. I may also be the patient who will have a short operation later today. I either suffered for ten hours yesterday, or will suffer for one hour later today.
>
> It is clear to me which I prefer to be true. If I learn that the first is true, I shall be greatly relieved.[34]

This case shows how most of us would prefer bad things like pain to be past rather than future; even if the past pains are worse than the future pains. So, considering Parfit's operations-case, temporal neutrality looks implausible – it just seems permissible that we care more about events ahead of us, that we want bad things to be in the past.

[34] Shortened from Parfit (2018), pp. 165-166.

Some proponents of temporal neutrality have argued that the claim that future-bias is permissible has limited appeal. In particular, they have suggested that its appeal is limited to self-regarding cases involving hedonic goods and harms. Brink has two suggestions he focuses on: past and future disgraces and other-regarding concerns to show that, future-bias isn't intuitively permissible beyond self-regarding hedonic cases like Parfit's operations. Both don't work particularly well. Let's start with other-regarding cases.

Brink's suggestion to demonstrate instability in intuitive support for future-bias based on concerns for other people is based on one of Parfit's cases[35].

> Past and Future Pains of Others. You receive a message about your daughter who lives in another country. The message says that your daughter had an accident that injured her greatly, and that she will suffer great pain in an operation. This depresses you. But then you receive another message, telling you that the earlier message was delayed, and your daughter already suffered through the operation. Do you feel relief that it is already over?

Brink claims that you don't seem to have future-bias when it comes to concerns about others, past pains are just as bad as future pains. [36] This, Brink argues, shows that future-bias is

[35] Parfit (1984), pp. 181-182.

[36] See also Parfit (1984), p.181, for a similar case.

unstable, since as soon we move from concerns about ourselves to concerns about others, the intuitive appeal disappears. Which, according to Brink, should lead us to question whether our intuitions about rationality are correct – and lead us to revise our self-regarding attitudes in a way that fits our other-regarding attitudes. The same argument is also endorsed to a certain point by Brueckner and Fischer[37] and seems to run like this:

(1) Future-Bias is intuitively permissible in self-regarding cases.

(2) Future-bias is not intuitively permissible in other-regarding cases.

(3) If two cases are relevantly similar, our intuitive response should be the same in both.

(4) Self-regarding cases are relevantly similar to other-regarding cases.

(5) Our intuitive response for future-bias's permissibility in self-regarding and other-regarding cases should be the same.

[37] Brueckner and Fischer (1986), p. 217. Note that Brueckner and Fischer don't actually argue for Future-bias's overall irrationality, but mainly for future-bias not being applicable to others and to goods and harms we don't actively experience. On the contrary, Brueckner and Fischer defend the view that future-bias in combination of a deprivation account best explains the badness of death.

From this, we can go in both directions, and either revise our intuitions in favour of temporal neutrality or future-bias.[38] This needs to be avoided by the friends of temporal neutrality:

(6) Our intuitions in other-regarding cases are more reliable than our intuitions in self-regarding cases.

(7) We ought to revise our intuitive response to future-bias in self-regarding cases towards impermissibility.

Dorsey simply denies (2) and claims that intuition still favours future-bias.[39] Additionally, we should add that (6) is quite controversial, as the intuition behind temporal neutrality is not stable here either. As Hare argued, our intuition about our daughter varies according to spatial distance: If I'm nearby, I seem to be future-biased on behalf of my daughter – if I'm far away, I seem to be temporally neutral.[40]

This suggests a problem with Brink's case: Our intuition might shift because of the distance, not because of other-regarding concerns. The intuitive support behind future-bias's permissibility might not change based on who we care about, but how far away they are.

Would it help Brink's cause to investigate the asymmetry based on proximity further? After all, proximity as such shouldn't be a relevant factor in evaluations, as was famously argued by Singer

---

[38] See Hare (2008) for a systematic argument for (5), as the inconsistency leads to a puzzle, and we should therefore be consistently future-biased or temporally neutral.

[39] Dorsey (2016), pp. 7-8.

[40] Hare (2008), pp. 269-271.

and his allies[41] – so other things being equal, why should my intuitions about future-bias shift depending on how far away my daughter is?

In Singer's famous case, intuition tells us that not helping a person is wrong if the person is near us, and not wrong if the person is far away. Singer then argues that we should revise our intuition about the far away case, as it would be absurd for him to say: It is not wrong to not help someone far away, therefore it is also not wrong to not help someone nearby. He illustrates by outlining that we can easily use communication devices and trusted testimony from experts to find out about the suffering of those far away, and as soon as we see the suffering, our intuitions align – we should help those far away too.

We can apply the same in Brink's case: if we would introduce a communication device so that we would be exposed to the suffering of the person far away, our intuitions would not align in favour of temporal neutrality. If you were able to see your daughter's pain over video or audio, or if you get a convincing testimony from her doctor about her pain, you will become more empathetic to her, and will tend to revise your intuition accordingly. Even when far away, it would at least be permissible to prefer her pain to be past, just as it would be if she were nearby. So,

---

[41] See Singer (1972), p. 232 or Unger (1996), pp. 33-35 for why proximity doesn't matter. See Woollard (2015), pp. 133-136 for why proximity does matter.

(2) Future-bias is not intuitively permissible in other-regarding cases.

seems false, as if we account for proximity, future-bias seems intuitively permissible.

There's an additional problem with the case: Even if we might find a case where it is clear that we're not sensitive to proximity but clearly have an asymmetry between self- and other regarding concerns, you can still hold that we should either be future-biased or temporally neutral consistently in both cases.

Hare concludes that we should be future-biased concerning others, regardless of near or far, because our future-biased concern should follow from our imaginative empathy for our daughter's situation in this case.[42] His argument for this is interesting: Hare thinks that because the daughter herself would have a future-biased preference, we would, by being temporally neutral, contradict her preference on her behalf – which is not justified according to Hare.

We don't need to agree with him that paternalistically contradicting other person's preferences is wrong. What's interesting here is Hare's suggestion that our concerns about other people are based on the other person's rational concerns. Brink suggested that our temporally neutral intuition about other people narrows down our future-biased intuition about ourselves – in other words, we might take our other-regarding attitudes to revise our self-regarding attitudes. But we could also say that,

[42] Hare (2008), pp. 276-277.

because rationality says differently, we should re-examine our intuition about other people, since we care about other people on their behalf. So, (6) is also false, as it could just as well be the other way around, and rationality should inform our intuitions about other people.

What we can learn from this is that to demonstrate intuitive instability for future-bias's rationality, we need a case that is not easily debunked in terms of its intuitive support and is not easily reversible. I will suggest two cases that might do the trick. Let's look at Brink's other suggestion, and from there, let's build the best case against future-bias.

Moral Failure

Brink's other suggestion is that I might prefer pain to be past rather than future, but when it comes to disgraces, I am temporally neutral - e.g. it doesn't matter to me whether it's past or future that I disgraced myself with bad jokes and too many drinks at a party – suffers from not being a clean case.[43] As Dorsey[44] suggests, we think that disgraces are instrumentally bad, and if we isolated the case well enough from instrumental effects such as social standing and loss of confidence, we would prefer a past disgrace to a future one. Additionally, it's not clear how a non-instrumentally bad disgrace would not be a hedonic harm, as its badness would be a particular kind of pain we'd feel.

---

[43] This example can also be found with Rosenbaum (1989), pp. 364-365.

[44] Dorsey (2016), pp. 6-7.

But maybe we can improve on Brink's case: Brink tries to capture some sense of non-hedonic harms that involve not only me and my feelings, but someone else: non-hedonic goods and harms that are relational. Relational goods and harms like friendship, love, or their counterparts are difficult to reduce to their instrumental and hedonic benefits due to their value being based on the relation between several agents, and its goodness also contributing to the flourishing of an agent's life as a whole.[45] Brueckner and Fischer offer a case that involves relations to others, which can serve as a starting point:

> "Suppose, for instance, that you know that either some friends of yours have betrayed you behind your back nine times in the past or some friend will betray you behind your back once in the future. Here, it seems that you should prefer the one betrayal in the future (given that the betrayals are comparable, etc.). It also appears that, given a choice between being mocked once behind your back in the past and being similarly treated once in the future, you should be indifferent. (Of course, we assume here that you know that you can have no effect on the future events)."[46]

While the second example about mockery seems similar to Brink's case about disgraces, and is similarly unconvincing for the same reasons, the first case is more interesting for us: Here, Brueckner and Fischer suggest that in the case of trust, we should not be future-biased. Trust might be a good candidate for temporal

[45] See Brink (1999).

[46] Brückner and Fischer (1986), p. 216.

neutrality, as it is clearly relational and non-hedonic, and we maybe strengthen the case by making it more realistic:

a) Your partner has had a major, long-lasting affair with someone else in the past.
b) Your partner will sleep with someone else once in the future.

Assume that this is only about trust in your partner and that you'll never experience these betrayals, e.g. by catching your partner cheating with b) – which one should you choose? Here, it at least seems a lot less natural to say that you ought to choose a) just because it's in the past. Betrayal is betrayal, and if someone were to give you relationship advice, they might say that it's crazy to choose nine betrayals over one, just because the nine are past.

But the case might need to be made more precise, as I am not sure about where intuitions lie. Firstly, I'm personally inclined to deny that our intuitions say that we are not *allowed* to prefer the past betrayal. Note that we are not asking whether intuitions favour b) over a), but whether we'd say that it's *irrational* to choose a). Surely, it would be perfectly understandable that someone would want to have all betrayals behind them – we wouldn't accuse them of being irrational if they displayed that preference. Why would it be crazy wanting to have the cheating behind you, and not in front of you?

This points towards a problem with how the case is set up: what seems to mess with our (or my) intuition here is that the case is seen through how much it would hurt me – if it were ahead of

me, I wouldn't have to live *through* the betrayals, which seems preferable. But this seems to have a hedonic flavour about it, and needs to be cleared up: imagine that the past betrayals will not have any impact on your future, e.g. regarding your self-esteem, your future relationships, your ability to trust others. Ruling out instrumental harms, in both a) and b), your relationship to your partner overall would not change, as you will never notice the betrayals and their impact – your partner would behave just as they'd do without the betrayal. This is difficult to imagine and as a result difficult to have a clear intuitive direction towards either temporal neutrality or future-bias in my view.

I will pick up on these methodological difficulties later, but for now, for a clearer case that is more obviously relational and concerned with your lifetime wellbeing, let's turn the case around: it's *you* who did the betrayal!

a) You partner cheated on your partner nine times in the past.
b) You will cheat on your partner once in the future.

Assuming that you are in a loving, non-abusive, exclusive relationship with your partner without any mitigating circumstances that would justify cheating, this would clearly indicate a moral failure on your part, regardless of its instrumental and hedonic implications. Even if we don't factor out instrumental and hedonic benefits and harms of cheating such as having a bad conscience, there is still a clear sense on how this would be a stain on your life as a whole, assuming that being

moral plays a role on what makes your life a good one.[47] As with achievement, the goodness of being a virtuous agent, and the corresponding vices of moral failures illustrate concern for an agent's life as a whole, and thereby favours a temporally neutral picture. As a result, in this case, it's not difficult to imagine that you would like fewer moral failures in your life regardless of when they take place. I hope this case is in the spirit of Brueckner and Fischer and improves on their suggestion by making it less of a target for objections like Dorsey's.

Achievement

Achievement is a good acquired by fulfilling goals and projects true to the agent's values, through the agent's own efforts. When a scientist verifies her hypothesis after years of experiments, or a writer finishes her lifetime novel after a decade of writing, what gives rise to the good of achievement is long-term goal fulfilment and not pleasure. Note that while they are related, achievements are not necessarily tied to desire-satisfaction: you can achieve something even if at the moment of reaching your goal, you lack a present desire to fulfil the goal. On the other hand, fulfilment of desires might be good for you under desire-satisfaction theories

[47] Most people concerned with lifetime goodness will most likely agree that moral agency plays a role in explaining what a good life is, whether in terms of virtue and human flourishing or some other theory of goodness. See Griffin (2000), pp. 69-70, Finnis (2011), pp. 124-127, and Raz (2000).

even if you did not earn it with your own efforts, but it wouldn't be an achievement.[48]

Now, imagine that you're the scientist. You wake up, being dizzy from your hard work as a scientist and for a moment you cannot recall: did you complete your works a few weeks ago? Or are you still in the process of finishing the last bits and will reach your goals in a few weeks? You vaguely recall that someone has published a major ground-breaking contribution four weeks ago, and you also remember that someone was about to finish a smaller publication in four weeks. Would you rather be the first or the second person?[49]

This is a case where you'd be more obviously drawn towards temporal neutrality, and indeed Hurka uses this example to demonstrate how perfectionist theories of well-being just *have* to be temporally neutral. It wouldn't even occur to you to be future-biased here, of course you'd have the past major achievement over the future, smaller one. You surely wouldn't discount your past achievements in favour of future ones, would you?

The reason why this case is more effective for Brink's argument is not only that achievement isn't easily broken down to instrumental or hedonic components – you would clearly care about achievements even if you'd factor out positive effects like social standing, recognition and effects of your scientific

---

[48] See Bradley (2009), pp. 13-14 for a discussion of this. Also, see Hurka (1993) and Scanlon (1998).

[49] Vaguely based on Hurka (1996), p. 61. Also endorsed by Sullivan/Finocciaro (2016), p. 148.

breakthrough. Achievement is also a kind of good that is clearly linked to what makes a life good as a whole, and connects to a picture of a temporally extended agent due to its fulfilment not being fixed upon a momentary point in time: if you care about achievement, it seems that you care about what you as a person can do to fulfil your goals throughout your life, not only about your present. Achievements are based on temporally extended goals that often only get fulfilled in different parts of an agent's life, so the intuitive way of viewing achievement as a good is to treat all parts of a life as equal. Or to cite Hurka directly, achievement seems to be "good from a person's point of view", which suggests that it is "good from a person's point of view at all the times in their life".[50]

2.2 What Exactly Is the Argument?

Now armed with two promising cases for their argument, let's take a moment to look more closely at the argument offered by the temporal neutralist. The dialectic is this: future-bias just seems rationally permissible. Hence, the burden of proof falls to defenders of temporal neutrality, who need to provide an argument to show either its irrationality or that the intuitive support is misleading or very isolated and hence cannot serve as justification. This is where the achievement and moral failure cases come in to show that future-bias's appeal does not generalise well.

[50] Hurka (1996), p. 60. Also see Griffin (2000), pp. 64-65.

Take Dougherty, analysing Brink's remarks about intuition shifts:

> "Brink notes that future-bias is limited only to hedonic experiences like pleasure and pain, and notes that he might prefer a smaller future disgrace to a larger past disgrace. Building on this remark, we might put the point the following way: It is arbitrary to have future-bias about some gains or losses but not others. This arbitrariness suggests that the preferences are not formed by rational processes."
>
> "Second, Brink notes that we lack this preference about pains and pleasures when these are the pains and pleasures of other people who are not immediately present. Again, we could view this worry as a concern with arbitrariness: there seems to be no good reason for being future-biased about ourselves but not about others."[51]

As I've explained above, Brink's appeal to other-regarding cases fails due to future-bias still being intuitively permissible after we account for proximity, as well as due to the argument being reversible in a way that we should revise our intuitions in other-, not self-regarding cases. But with the non-hedonic cases involving achievement and moral failure, we can get a better argument going. The basic idea here seems as follows:

(1) Future-bias is intuitively permissible in hedonic cases.

[51] Dougherty (2015), p. 3.

(2) Future-bias is not intuitively permissible in non-hedonic cases.

(3) If two cases are relevantly similar, our intuitive response should be the same in both.

(4) Hedonic cases are relevantly similar to non-hedonic cases.

(5) Our intuitive response for future-bias's permissibility in hedonic and non-hedonic cases should be the same.

From this, we can revise intuitions in either direction, and friends of temporal neutrality need to argue specifically that it's our intuitions about the hedonic and self-regarding cases that should be revised. So, they need to add something like:

(6) Our intuitions in non-hedonic cases are more reliable than our intuitions in hedonic cases.

Then they can get the conclusion they want:

(7) We ought to revise our intuitive response to future-bias in hedonic cases towards impermissibility.

Premise (1) is established by Parfit's operations case, while Brink unsuccessfully tried to establish premise (2) with his cases on disgraces and other-regarding operations[52] – which we have supplemented with cases concerning achievement and moral failure. Premise (3) can be illustrated further: when we view two scenarios under rational reflection, there should be no reason to treat relevantly similar cases differently. This idea is mentioned by Dougherty[53], and developed to a full systematic argument by

[52] See Dorsey (2016), p. 5.

Sullivan by appealing to relevant similarity between time- and location-biases.[54] (3) should be universally supported in moral theory – even if you're endorsing particularism about reasons, you would insist that the different treatment of similar cases is down to a change in context, which index the reasons concerned – making two cases dissimilar from each other. If we were to look at two cases with identical contexts and reasons, we would treat them similarly, as they would basically be the same case. So even particularists should accept (3), they just think that the number of non-identical cases that are relevantly similar to each other is quite low.

Premise (6) is also a key premise that, from my impression, is not explicitly defended by Brink and others. As this is a necessary step for the temporal neutralist to take, to avoid the argument going against the intuitive support of temporal neutrality, it is important to outline shortly what speaks in favour of (6). One thing the temporal neutralist can appeal to is numerical advantage insofar as there are more non-hedonic cases pointing in the same direction than there are hedonic ones. As there is only one kind of hedonic good (pleasure), but many other non-hedonic ones (achievement, moral failures, and more), and all intuitions except the one in hedonic cases point towards temporal neutrality, hedonic cases seem less reliable.

Let's dismantle the argument. I will proceed to show that the argument above fails: premise (3) and (4) are going to be the first battlegrounds. I will explain how proponents of temporal

[54] Sullivan (2018), p. 24.

neutrality do not establish (3) and (4), as they do not specify what they mean by "relevantly similar", and that we have good reason to think that non-hedonic cases concerning achievement and moral failure are relevantly different to cases involving pleasure and pain, as the former concern lifetime goodness and the latter momentary well-being. After that, I will argue that (2) can still be plausibly questioned even with strengthened cases about achievement and moral failure, as future-bias can still show up even if not to the extent of absolute discounting known from hedonic cases, and if proponents of temporal neutrality try to idealise the cases further to factor out interfering hedonic thoughts we might have when looking at the cases, they will make the scenarios so abstract that our intuitions on the cases become unreliable – thereby undermining premise (6) of the argument.

So, in summary, the argument fails both to establish concerns of arbitrariness that future-bias isn't result of a rational reflection process due to its instability in intuitive support, and fails to show that we ought to revise our intuitions about future-bias's rationality.[55]

---

[55] It is worth mentioning that empirical evidence on how future-biased people are exists with Caruso (2008, 2010) amongst others. The cases examine demonstrate a wide range of goods and harms participants are future-biased about, including the value of work and labour, fairness and justice, solidarity, as well as the virtue of generosity. While the cases only examine whether participants actually are future-biased, and not whether their intuitions say that it's permissible to be future-biased, and are not sufficiently isolated to exclude hedonic factors, there were some participants indicating that they feel

2.3 Moments and Lifetimes

When I try to find out how someone is doing, I can ask about it in different ways. I can issue a simple inquiry into how a person feels now, by asking something like "How are you?" or "How is it going?". If I would like a more detailed report, I could ask a person something like "How have you been since I last saw you?" or "How was your week?", to receive a well-being report over a specific period of time. What these questions do not cover is how good their life is overall, as they are simple snapshots of how a person is doing.

Regardless of what you think well-being is exactly – the question is so loaded that I won't make an attempt to summarise the debate – there is an obvious difference between momentary well-being, well-being over periods of time, and lifetime well-being. I could be doing quite well right now, even though this year has been overall pretty bad for me, but overall, I could still be on track to living a good life (hopefully).[56]

Whether and how these three things interlink and contribute to each other are big and open questions – how momentary well-being contributes to lifetime well-being, what the smallest and biggest atoms within lifetime well-being are, and whether we know how good a life is overall if we know how good its value atoms are – these are questions that I won't be answering in this

---

drawn towards future-biased even if they know about the arbitrariness of preferring one good over the other just because it is past.

[56] For a good overview of the debate, see Bradley (2009), chapter 1.

chapter.[57] What is important for my purposes here is mainly that there is a significant difference between momentary well-being, temporally extended well-being, and lifetime well-being. Consequently, goods and harms that concern different types of well-being are relevantly different to each other.

So, a hedonic good like pleasure concerns momentary well-being, while a non-hedonic good like achievement is concerned with temporally extended well-being. Regardless of what theory of well-being you end up defending, any of them will hold that achievement and pleasure are significantly different goods that contribute to different kinds of well-being.

There is a systematic reason behind this: The difference runs so deep that it might even be questioned whether the notion of achievement as a good makes sense under a theory of well-being that reduces all well-being to momentary well-being, as achievement sometimes fulfils goals and projects we had in the past but don't necessarily have in the moment when they are achieved.[58] For example, if I'm a scientist that achieves a publication in a major journal, leading to a breakthrough in my field, it could well be that, when I submitted my manuscript, I cared very much about it being published – but at the time when my manuscript was accepted, my desire has long past. If I only

---

[57] For some answers, see again Bradley (2009) for a hedonist viewpoint, see King (2018 and 2019) for a discussion how lifetime goodness and well-being can come apart, see Bruckner (2013) for a defence of desire satisfaction.

[58] See Bradley (2009), p. 22 for an extended discussion on this that leads to a problem for non-hedonic theories of well-being.

look at momentary well-being, there is nothing that makes the publication good for me. Only if we look at temporally extended well-being, my scientific achievement is intelligible as a contribution to my well-being.

Now recall premises (3) and (4) of the temporal neutralists' argument:

(3) If two cases are relevantly similar, our intuitive response should be the same in both.

(4) Self-regarding cases are relevantly similar to other-regarding cases, and hedonic cases are relevantly similar to non-hedonic cases.

While we should accept (3), it is never specified by proponents of temporal neutrality what makes a case relevantly similar to another beyond that all cases involve past and future goods and harms – and in this particular comparison of hedonic and non-hedonic cases, (4) seems plainly false. Hedonic and non-hedonic cases are quite significantly dissimilar, as I have demonstrated above: the former concerns momentary well-being while the latter does not.

Why is it a relevant difference that one case concerns momentary and the other one temporally extended well-being? The reason is that it is not possible to hold up the other things equal condition between the two cases when comparing. When we compare intuitions between a case with hedonic and one with non-hedonic goods , we try to keep everything similar, especially the being-past and being-future components, with the only exception the kinds of goods involved – pleasure and achievement – to get a sense of whether and why our intuitions differ. However, as soon

as we look at the kind of good in question, we will automatically examine momentary well-being when it comes to pleasure, and temporally extended well-being when it comes to achievement. As I explained, it isn't possible to evaluate achievement under momentary well-being alone, and we therefore must view it under a temporally extended kind of well-being that, under most achievementist views, will include all of life. Hence, the good in question, achievement, isn't really only past or future anymore, as it concerns a temporally extended period that goes beyond being past. Therefore, the other things equal condition between the cases is violated.

That achievement concerns lifetime well-being may well be the reason why our intuition favours temporal neutrality, as a good that concerns lifetime can't really properly be over and done with. But this also is the reason why the cases are not relevantly similar, and we should therefore reject premise (4) of the temporal neutralist's argument. This is, however, not to say that future-bias can never appear when we are concerned about lifetime well-being: Even when lifetime well-being is concerned, we can prefer some goods to be later in life, and harms to occur earlier.[59] This roughly corresponds with a non-absolute variant of future-bias that only discounts past value in a hyperbolic, and not absolute way.[60] In any case, the difference in kinds of well-being between hedonic and non-hedonic goods should lead us to reject (4).

---

[59] For a discussion of this, see Dorsey (2015), Velleman (1991).

[60] See Sullivan (2018), p. 4-5 for a discussion of hyperbolic vs exponential discounting.

Let's shortly demonstrate the same thing about our second case concerning moral failures. When I cheat on my girlfriend, it might not be bad for me in the moment when the betrayal takes place. While the cheating as such might also be bad in some sense if we only view it from a perspective of momentary well-being, the true extent of the moral failing may not be fixed to that moment – what is bad may be the repercussions after, such as a bad conscience, the pains of a failing relationship and so on. But if we exclude these instrumental future factors, it will be difficult for us to explain fully what exactly is bad about the act of cheating if we only consider momentary well-being.[61] Only if we look at temporally extended well-being, we can make sense of my failing as a vice and say something about how my moral failure makes my life worse. Therefore, both cases I've discussed fail to establish premise (4) of the argument against future-bias.

What could be a reply on behalf of the friends of temporal neutrality? The first thing they could say is that I haven't really specified why different kinds of well-being are a relevant difference for intuitive responses to cases concerning future-bias and temporal neutrality. Without specifying what counts as a relevant difference, we could point out differences between hedonic and non-hedonic goods until the end of our days.

My first response to this would probably be: that's not really my job. As the argument against future-bias, which was set up by authors like Brink, Brueckner and Fischer, Sullivan and

[61] This is an argument that has been used against hedonism discussing the value of friendship and love.

Dougherty, relies on hedonic and non-hedonic cases being relevantly similar, the honour of specifying what that means falls to the defenders of temporal neutrality. So, it's their problem really.

Secondly, when we try to compare intuitions responding to thought experiments that primarily concern the temporal location of different kinds of goods and harms, what should be relevant to those cases should only concern the difference in what kinds of goods and harms we examine, while other factors in the scenarios stay fixed – all other things are equal, especially the temporal location bit, since we (as the temporal neutralists) are trying to show that, whether we look at hedonic or non-hedonic goods and harms should be irrelevant when it comes to temporal location of these goods and harms. So, when looking for a relevant difference between the two scenarios described, we should point out a factor within the comparison of the scenarios that nullifies the other things being equal condition without directly changing the scenarios. In other words, you need to show that both cases are not similar without invoking anything other than the issues at stake.

For example, in the classic child in a pond case, Peter Singer compares a case of close proximity to a drowning child to a case far away where a child is drowning, with everything else being the same, thereby establishing that proximity is an irrelevant factor, and helping someone in close proximity to you is relevantly similar to helping someone far away. If you don't want to help children in ponds far away, your strategy will usually be to show how there are factors in one of the scenarios that differ

other than proximity, which are linked to what is at stake in the debate, e.g. responsibility frames for agents, structural injustice issues underlying the cases, and so on. These are all factors that keep the scenarios described by Singer, while also only invoking factors that are at stake in the debate and aim to show how the other things being equal condition cannot hold between the two ponds.

The same is true for cases comparing hedonic and non-hedonic goods in terms of future-bias and temporal neutrality. When raising the point about difference of temporal extendedness of well-being, we're not looking at anything beyond hedonist or non-hedonist theories of well-being. But when we are considering the difference between momentary or temporally extended well-being, this makes it impossible for the ceteris paribus condition to hold between the hedonic and non-hedonic cases.

In short, even if we grant the intuitive plausibility of future-bias being irrational in cases of achievement and moral failure, the attempt to undermine the intuitive permissibility of future-bias when it comes to hedonic goods and harms fails as there are relevant differences between hedonic and non-hedonic goods and harms that may well warrant different intuitive responses – a general impermissibility of future-bias can't establish via this route.

2.4 Idealisation and Evidence

In the rest of the chapter, I will discuss shortly what happens if we don't immediately concede the intuitive pull towards temporal neutrality in cases of achievement and moral failure. I will argue

that, even in cases concerning achievement and moral failure, it's possible to make the case for future-bias being permissible, attacking

(2) Future-bias is not intuitively permissible in other-regarding and non-hedonic cases.

In a similar vein as Dorsey and Hare have for disgraces and other-regarding cases. I will also show that, if defenders of temporal neutrality try to rescue the cases by further isolating factors in the thought experiment, the cases become so abstract and idealised that intuitions in these cases cease to be reliable indicators of rational preference, thereby undercutting

(6) Our intuitions in other-regarding and non-hedonic cases are more reliable than our intuitions in self-regarding hedonic cases.

as our intuitions in the hedonic cases will turn out to be much clearer, cleaner, and more reliable.

Recall the case about achievement: You're one of two scientists, either one who has published a major breakthrough four weeks ago, or one who is about to publish a smaller contribution in a month. Which scientist would you prefer to be?

While your initial reaction, as should be typical for achievement, may point towards the bigger past achievement, let's continue and exclude all instrumental effects of the past achievement – recognition as a scientist, earnings, reputation etc. are the same in both cases. The same goes for the effect of your contribution – who made it doesn't make a difference, this should only be about you fulfilling your project, not about impact in science. Second,

exclude stress and labour involved: you might be influenced by the effort it takes to get to the future achievement to prefer the past achievement – precisely because of future-bias you'd prefer the effort to be past rather than future.

If these are excluded, I'd suggest that it's perfectly reasonable to prefer your achievement to be future rather than past. The meaningfulness and satisfaction from reaching a goalpost will be gone with your past achievement and matter a lot less to you, especially as a committed creator of scientific research. Note that I'm not pushing achievement towards hedonism with feelings of satisfaction – the value of achievement arises through a process towards a goal. But as a person who cares about fulfilling goals, as a "doer" as Thomas Hurka[62] would say, you'd like this process to be ahead of you, not behind you. This activist impulse fits with the value of achievements and makes it perfectly reasonable to prefer achievements to be future rather than past – even if the past achievement is bigger. At the very least, it shouldn't be seen as irrational to prefer achievements to be ahead of us.

The temporal neutralist would obviously not agree: What is happening here is that feelings of momentary well-being are interfering with our judgement in this scenario – when we imagine the case, we ask from the perspective of the scientist at the moment of waking up, and assume that we at that time have a desire for scientific achievement. However, this is a hedonist influence on our judgement that pulls us towards concerns about current feelings of satisfaction or pleasure – even without having

[62] Hurka (1987), p. 729.

this desire at the present time, an achievement would still be good for us. So, what we need to do is further isolate the case to exclude this lingering feeling of feeling pleased or satisfied about the achievement.

So, when waking up, you do not know which scientist you are, but you also do not have a desire for a publication at the current time and won't derive any pleasure from getting an achievement. For this to be the case, we need to assume that, at that time, you won't be conscious of the achievement and will never know which scientist you turn out to be.

This might sound puzzling first – how can something be good for me if I'm not even conscious of it? However, having an achievement without being conscious about it seems possible, and it might still seem to be good for me. Take for example, the case of Jamal Khashoggi, the journalist murdered by the Saudi-Arabian regime – a lifelong advocate against oppression. Because of his murder, he won't be conscious of any change happening in Saudi-Arabia. However, let's say in response to his murder, the international community sanctions Saudi-Arabia's regime and forces it to democratise the country and reduce the oppressive grip on its people – it wouldn't be absurd to claim that this is an achievement that is good for Khashoggi, even if he will never know.

So, should I be future-biased when choosing between a smaller future and a bigger past achievement that I both will never become conscious of, or even notice? A proponent of temporal neutrality might say: look, now that we have removed

consciousness, you would be temporally neutral – what counts is what the better achievement is, not whether it is past or future. If I can choose between two achievements, e.g. a smaller book publication in the future that I will never notice, or a bigger book publication in the past that I don't know about, I would just choose the better achievement regardless of whether it is past or future. And indeed, even proponents of future-bias like Brueckner and Fischer say that future-bias can't be extended to goods and harm we don't experience.[63]

However, I find it difficult to have clear intuitions on a case like this where I won't even know about the achievements but can still choose. A case like this appears frustrating to me because it is difficult to even imagine having clear intuitive responses about. That's not to say that it's an impossible or unrealistic case. After all, Parfit's Past and Future Operations are also a wee bit unrealistic. But it is not easy to construct the case of unconscious achievements clearly, so that I don't have consciousness of the achievements but can still choose between them and then have a clear intuitive response to what is better for me.

To construct a case like this, we require a high degree of idealisation. We first need to imagine the case of past and future achievements, second we need to remove instrumental benefits, third we must tweak the scenario such that I won't know about any of these achievements and not experience them, and fourth I need to be able to choose between them, and forget about the choice afterwards – and after all this, we still need the assumption

[63] Brueckner and Fischer (1986) , p. 216.

that these achievements are still good for me in some sense. This overtakes the degree of idealisation that is necessary in Parfit's original case, where we only have to do step one and two, and still have a clear intuitive response to it.

I don't know where exactly the point is where idealisation of thought experiments causes intuitions to become unreliable. But the case of unconscious past and future achievements is such a highly idealised case that makes the imagination of the scenario difficult, thereby making our unreliable. As the case is so far removed from our ordinary sense of how we think about achievement, we should not be taken this as evidence against future-bias's rationality more generally. In other words, the temporal neutralist is undermining

(6) Our intuitions in other-regarding and non-hedonic cases are more reliable than our intuitions in self-regarding hedonic cases.

of their argument in order to make sure that future-bias does not apply to cases of achievement. With the increase in idealisation, it becomes more difficult to form a clear intuitive response towards either temporal neutrality or future-bias, thereby making our intuitions on non-hedonic cases less reliable than our intuitions in hedonic cases.

To illustrate this, compare this case with cases concerning very large numbers:

> Hangnails for Torture: For any excruciatingly painful torture session lasting for at least two years to be experienced by one person, there is some large number of

> minute-long very mildly annoying hangnail pains, each to be experienced by a separate person, that is, other things equal, worse.[64]

Here, one might say that the number of minute long hangnail pains that would outweigh two years of torture is so absurdly large that we cannot sufficiently imagine it to form a clear intuition that reliably supports the truth or falsity of *Hangnails for Torture*. As Pummer argues, even if we cannot imagine the billions of minutes of hangnail pains, it is possible in large number cases to still have intuition-based reasons to deny the truth of *Hangnails for Torture.* As long as we can relevantly imagine any number of mild hangnail pains, we then can extrapolate our imaginative response about a medium-sized case towards an intuition that goes against the truth of Hangnails for torture.[65]

This avenue of response is not open for the temporal neutralist in this case for obvious reasons: if we'd try to extrapolate from a less idealised case to an intuition-based reason, we would, as I have argued above, potentially extrapolate a future-biased intuition.

What could be attempted on behalf of the temporal neutralist is the reverse of what Pummer suggests for large number cases: We can increase the stakes to a very large difference in benefits in the scenario, to a point where we would have a clear intuition, and then extrapolate from there to a case with lower stakes: Imagine that you could choose between a past scientific achievement that

---

[64] From Pummer (2013), p. 37.

[65] See Pummer (2013), p. 39.

has revolutionised your discipline and a future achievement where you successfully presented your paper at a small conference. Or, for the moral failure case: imagine that you could choose between having cheated on your partner hundreds of times and flirting with someone else in the future. In these cases, it seems pretty clear that intuition favours temporal neutrality, as the number of goods and harms at stake is sufficiently high to drown out distracting features and to explicate our intuitive response. From here, we can extrapolate that, if we are temporally neutral in a high stakes case involving achievement or moral failure, we should also be temporally neutral in cases with lower stakes.

The problem with the suggestion is two-fold: firstly, the same could be said about hedonic goods – if we'd be asked whether we'd prefer past or future pain, some very large number of past pain could outweigh a very small number of future pain, even if we're future-biased. However, that alone does not show that preferring ten hours of past pain over one hour of future pain is irrational. It merely shows that the discount function of future-bias is not totally absolute. In the same way, we could uphold that when it comes to non-hedonic goods like achievement and moral failure, we are still future-biased, the only difference is that the discount function is less steep than in hedonic cases – but future-bias is rational in both cases. A potential explanation for this difference in discounting is the difference in kinds of goods we're concerned about: As I outlined above, hedonic goods concern momentary well-being, while non-hedonic goods tend to concern temporally extended well-being or lifetime well-being. With lifetime well-being, a past good is not entirely past, as it

still matters for the good of one's life overall, so absolutely discounting it seems counterintuitive – but it may still be permissible to have certain preferences around the location of that good, e.g. to have good things later on in life while having bad things occur earlier. Viewed from an agent's perspective, this would resemble future-bias, with a non-absolute discount function.

Secondly, we could amend the scenarios in the opposite direction: if we'd have to decide between equally valuable achievements in past and future, would we really, as a temporal neutralist would suggest, flip a coin? Or would we want the achievement to be ahead, rather than behind us? I think it's reasonable to assume that we would be reluctant to flip a coin. It could be objected that, in the case of two equally valuable achievements, the small hedonic factors that always remain regardless of how well we isolate a case tip the balance, and that is what makes us reluctant to flip a coin. This may be true, but the phenomenon remains that the lower the stakes, the less strongly we are pulled towards temporal neutrality, and the best explanation for this is that this merely reveals something about the nature of discounting with future-bias: when it comes to non-hedonic goods and harms, we are still future-biased, just less absolutely.

Overall, however, this kind of response does not solve the issue of idealisation – we still can't really imagine what it means to have an achievement without being conscious about it, even if we amend the quantities of the goods and harms in question. In summary, the temporal neutralist can choose: Either they

concede that future-bias is intuitively permissible in non-hedonic cases like achievement, and give up

(2) Future-bias is not intuitively permissible in other-regarding and non-hedonic cases.

or they can defend (2) by further idealising the cases, but thereby undermining

(7) Our intuitions in other-regarding and non-hedonic cases are more reliable than our intuitions in self-regarding hedonic cases.

as the high degree of idealisation renders intuitions in these cases less reliable than intuitions in hedonic cases.

2.5 Summary

In this chapter, I argued against the following argument which has been proposed by authors like Brink, Brueckner, Fischer, Hurka and Dougherty:

(1) Future-Bias is intuitively permissible in self-regarding hedonic cases.

(2) Future-bias is not intuitively permissible in other-regarding and non-hedonic cases.

(3) If two cases are relevantly similar, our intuitive response should be the same in both.

(4) Self-regarding cases are relevantly similar to other-regarding cases, and hedonic cases are relevantly similar to non-hedonic cases.

(5) Our intuitive response for future-bias's permissibility in self-regarding and other-regarding cases / hedonic and non-hedonic cases should be the same.

(6) Our intuitions in other-regarding and non-hedonic cases are more reliable than our intuitions in self-regarding hedonic cases.

(7) We ought to revise our intuitive response to future-bias in self-regarding hedonic cases towards impermissibility.

I have argued that (2) can be contested, and a defence of (2) will end up undermining (6) due to the high degree of idealisation rendering intuitions in non-hedonic cases less reliable than those in hedonic cases. I have also shown that defenders of temporal neutrality do not explain sufficiently what qualifies as relevant similarities between cases, and that (4) is false, as the discussed hedonic and non-hedonic cases differ fundamentally in terms of what kinds of well-being is affected. In summary, we can keep being future-bias for now, and can continue using intuitive appeal behind future-bias as evidence for its rationality.

3 The Past Isn't Arbitrary

3.1 Introduction

Louise Banks is a remarkable person. Not only is she an accomplished field linguist, she also learns to perceive the world as being timeless. She doesn't view the world as a causal chain of events, where the future is followed by the present, then by the past: She sees everything as temporally equal. She achieves this by learning a language called "Heptapod B" from aliens visiting earth, who perceive time differently than humans do. "Heptapod B" is a timeless language, and because language influences thought and perception, Banks gains a timeless perspective through learning a timeless language.

Louise Banks is fictional. She is the main character of the short story *Story of your life* by Ted Chiang, and the movie *Arrival* by Denis Villeneuve. Let's set aside obvious problems with changing perspective via language, free will and determinism, and aliens teaching humans their language, and focus on another question: when Banks' perception of reality changes, do her attitudes towards events change too? How does she *feel* about past, present, and future? How does her change in temporal perspective influence her attitude towards bad events like death, failure, or a painful divorce?

In both the short story and movie, the change in her perception of time indeed changes her attitudes, most notably towards the death of her daughter. Banks' simultaneously perceiving all times leads her to "remember the future", and as a result, she can see

that if she marries and has a daughter, her child inevitably dies an unnatural death. She accepts this, because *for her*, her child never ceases to exist and never goes beyond her perception. In the movie, she tells her husband that she was fully aware of their child's fate, and proceeded to have her anyway, leading to a divorce from her upset husband, which she also willingly accepts, because in her perspective, neither her daughter, nor her husband are truly gone. For Banks, whether something is past, present, or future is an arbitrary difference, because of her perception of time. So, with her changing temporal perception from a tensed to a timeless perspective, her attitudes towards past, present and future events also change to a timeless outlook that is temporally neutral about events in her life.

What does it mean to be temporally neutral regarding attitudes towards events? Temporal neutrality is understood by Brink as a requirement for agents to not place any normative significance per se on the temporal location of goods and harms, and to give equal significance to all parts of one's life.[66] So, whether an event is tomorrow or next week, or whether it's in the past or future, shouldn't matter as such in my evaluation of the event. In short, I shouldn't be time-biased.

The two most common time-biases are the bias towards the near and the bias towards the future. If I'm near-biased, I'd prefer positive future events to be closer to the present and negative future events to be further away from the present. So, if Louise Banks were near-biased, she'd prefer to split with her husband

[66] Brink (2010), p. 1.

later rather than sooner, other things, like loving time spent together, being equal. But Banks is not near-biased, as she is temporally neutral, so she is indifferent to the temporal location of the divorce as such.

If I'm future-biased, I place little to no normative significance on events that are past[67]. For example, if I compare a positive event in the future and an at least as positive event in the past, I would always prefer the future positive event, just because it is future and not past. And if I compare a negative event in the future to an event in the past that is at least as negative, I'd always prefer the past negative event – because it's already past. So, if Louise Banks were future-biased, she would prefer her daughter to be alive in the future than in the past, rather than her daughter's existence being past, not future. But Banks is not future-biased, as she is temporally neutral, so it does not matter to her that her daughter's time alive is past, not future.

Is Dr Louise Banks strange? Or should we aspire to be like her? One argument against time-biases, and for temporal neutrality, is the *arbitrariness-argument* defended by Henry Sidgwick, John Rawls, and most recently Meghan Sullivan. The arbitrariness-argument states that a mere difference in temporal location is too arbitrary to provide grounds for a preference- or attitude-change.

---

[67] This is a version of future-bias consisting in absolute discounting. There is some evidence that future-bias takes the form of an absolute discount function and not a hyperbolic or exponential one. See Greene and Sullivan (2015), pp.961-962, and especially Sullivan (2017), pp. 49-50. This chapter will follow Sullivan in describing future-bias's discount function as absolute.

Just because two events occur at different distances of the present doesn't justify a preference between them. Therefore, near-bias is not justified. And just because an event is past rather than future, or future rather than past doesn't justify a change in attitude towards them. Therefore, future-bias is not justified. If the arbitrariness-argument is sound, then time-biases are impermissible, and temporal neutrality is vindicated.

So, should we be temporally neutral? I'll suggest that we don't have to, and that in some ways we shouldn't. The first half of my chapter is largely defensive: I first outline the arbitrariness-argument, which is supposed to show that time doesn't provide rational grounds for attitudes and preferences. I then summarise three problems with Sullivan's arbitrariness-principle, which in combination should lead us to reject Sullivan's argument.

In the second half of my chapter, I will try to sketch a positive suggestion on how we can provide rational grounds for future-bias: while it's conceivable that we could remember the future, we still cannot assume the perspective of a timeless person like Louise Banks. We must treat the past as something different because of how our agency and perception works – to perceive events, we need to view the world as sequential, as if time would pass. Time's passage does not have to be a metaphysical reality, like A-Theorists suggest. But without seeing the world through the lenses of passage, we wouldn't be able to make sense of events, and could not be conceived as agents interacting with the world. Therefore, as our perception of the past is necessarily different from our perception of present and future, our attitudes reflect

this as well, and there are rational grounds to be biased towards the future.

3.2 The Arbitrariness-Argument

Before we start, let's clarify what kind of preferences and attitudes we're talking about. Sullivan distinguishes two kinds of preferences and attitudes when it comes to practical rationality: preferences that are connected to choices and actions, and preferences that are not.[68] An example of the former kind of preference would be my preference for tea over coffee, leading me to choose a specific beverage at a café. However, not all sorts of preferences lead to actions – some are just reflections on my attitudes towards something, even if I cannot change it. Sullivan calls it *approbative rationality* – the rationality governing what we approve or disapprove of.[69] For example, I can disapprove of the second world war, even if this never leads me to a choice or action. I can also, as a Pythagoras-fan, prefer that the square root of 2 to be a rational number, although this is impossible.

Time-biases can be criticised as either kind of preferences. For example, they are vulnerable to pragmatic criticism: time-biases can lead me to choices that leave my life worse off overall. Sullivan calls this the *compensation-argument*, and I won't comment on it in this chapter. Sullivan thinks that the

[68] Sullivan (2018), p.3

[69] Sullivan (2018), p.4

arbitrariness-argument successfully rules out time-biases as part of approbative rationality – they may become action-guiding at some point, but we can criticise them independently of their leading to any choice. Let's see how successful the argument is.

The basic idea behind the arbitrariness-argument against time-biases is that a mere temporal difference between events is so arbitrary that it doesn't warrant a preference- or attitude-change. As Sidgwick says:

> "Hereafter as such is to be regarded neither less nor more than Now. [...] the mere difference of priority and posteriority in time is not a reasonable ground for having more regard for the consciousness of one moment than to that of another."[70]

And Rawls agrees:

> "The different temporal position of persons and generations does not in itself justify treating it differently. [...] There is no reason for the parties to give any weight to mere position in time."[71]

So, just because there is a difference in temporal location, position or order of events, goods and harms, this does not provide a reason for preferring one of them over the other. Why doesn't it? It may be instructive to compare preferences concerning temporal

---

[70] Sidgwick (2019), p. 380.

[71] Rawls (1971), p. 259.

location with preferences concerning spatial location first. Consider the following case by Sullivan:

> Detergents: My local grocery store stocks three kinds of detergent: Wisk, Surf, and Tide. Each is composed of the same cleaning agents, and I'm aware of this. Wisk and Surf are stocked on the same shelf at about waist height. Tide is one shelf above them at eye level. I'm indifferent between Wisk and Surf but prefer Tide to either of the other detergents.[72]

Would you criticise me for my behaviour, since the only difference between the detergents is location, which doesn't seem to justify a preference change? My preference for one detergent over the other seems to be supported only by a difference in locational properties, which are facts that are arbitrary, or not normatively relevant when it comes to detergents. Or in other words, if my mum asked me why I always bring home Tide and I'd answer that it's on the top shelf, she might not accept this as a good reason.

But if you'd criticise me for changing preferences because they are just based on spatial location alone, doesn't it also seem to be that you ought to criticise me if I change preference based on temporal location alone? If I prefer one event over the other just because of their difference in temporal location, that seems to be

---

[72] Sullivan (2018), p. 38. I amended the example to take out Sullivan's suggestion that Tide is more expensive than Wisk and Surf so that our intuitions are not distracted by pragmatic considerations.

as arbitrary as my preference in *detergents*. To illustrate, consider Future-Tuesdays:

> Future Tuesdays: A certain hedonist cares greatly about the quality of his future experiences. With one exception: he has Future-Tuesday-Indifference. Throughout every Tuesday he cares in the normal way about what is happening to him. But he never cares about possible pains or pleasures on a future Tuesday. Thus, he would choose a painful operation on the following Tuesday rather than a much less painful operation on the following Wednesday. This choice would not be the result of any false beliefs. The man knows that the operation will be much more painful if it is on Tuesday and agrees that it will be just as much him who will be suffering on Tuesday, and knows that Tuesday is merely part of a conventional calendar with an arbitrary name. Nor has he any other beliefs that might help to justify his indifference to pain on future Tuesdays. This indifference is a bare fact. When he is planning his future, it is simply true that he always prefers the prospect of great suffering on a Tuesday to the mildest pain on any other day.[73]

The hedonist's preference pattern, even if it wouldn't lead to a life with more pain to him, would clearly be strange to us – it's arbitrary to assign this importance to Tuesdays just because they are Tuesdays. And while Future-Tuesdays is a very obscure time-bias, we can apply the same thinking to more common time-biases like future-bias:

Arbitrariness-Argument against future-bias.[74]

---

[73] Shortened from Parfit (1984), pp. 123-4.

[74] Sullivan (2018), p. 108, Brink (2010), p. 4. Sullivan and Brink offer the same argument against near-bias. See Sullivan (2018), pp. 36-37.

(1) It's not rationally permissible to vary preferences according to arbitrary differences. (Arbitrariness-Principle)

(2) Being past rather than future is an arbitrary difference between events.

(3) If an agent is future-biased, her preferences are sensitive to events being past or future.

(4) Therefore, time-biases are not rationally permissible.

Sullivan thinks the best defence for (1), the arbitrariness-principle, comes from scenarios like the detergents-case or Future-Tuesday-Indifference. It seems very intuitive for us to criticise behaviour like this, and we ought to accept non-arbitrariness as a basic requirement of rationality. However, premise (1) can be challenged on several fronts, which in combination, to my eyes at least, already successfully undermine Sullivan's argument.

I'm just feeling it

Firstly, Lowry/Petersen argued that, if an agent is time-biased, this may be arbitrary because the preference isn't based on rational grounds, but this is insufficient to show that the preference is impermissible[75]. For rationality to rule out preferences, the preference must be based on *irrational* grounds. An irrational ground for a preference is a ground that provides a reason against having that preference, e.g. the fact that I have social anxiety provides me with a reason against a preference for conference dinners. However, mere differences in temporal

[75] Lowry and Petersen (2011), p.493.

location are neither rational nor irrational grounds – according to Lowry/Petersen, difference in temporal location is normatively neutral, neither providing a reason for nor against a preference.[76] Hence, it's not impermissible to have a time preference, and therefore, time-biases are rationally permissible.

This case can be bolstered further. According to Ruth Chang, "*if one is faced with a choice between two relevantly identical alternatives, 'feeling like it' can rationalise one's act of going for it'*[77]. Chang talks about the rationality of actions here, but as outlined above, rationality of actions is based on approbative preferences, so if "feeling like it" can constitute a reason to act, it can also constitute a reason for a preference. If we accept that desire can sometimes play a role in rationality, "Feeling like it" may serve as a tie-breaker when other independent reasons "run out"[78]. For example, if I have equal reasons for Surf, Wisk, and Tide in *Detergents*, *feeling the top-shelf* may generate a perfectly rational justification for my preference for Tide.

To make it clear that this is not an ad-hoc response, it might be helpful to illustrate it with a further example from the literature on rational requirements. Imagine yourself being lost on a trip in the Scottish Highlands, with your food running out. After several days, starving, you manage to carry yourself to a small town which has two pubs. Both pubs are open, are 10m apart from each other, and serve food at similar prices. You have equal

---

[76] Lowry and Petersen (2011), p. 494.

[77] Chang (2004), p. 82.

[78] Chang (2004), p. 83.

reason to choose either of the pubs, and you'd really rather not starve to death. Would we really criticise someone for being arbitrary by just going for the one on the left just because they're feeling like it?[79]

Similarly, if I'm faced with two future events that are equally good or bad in all rationally relevant aspects, "feeling the earlier one" can constitute a preference based on temporal location per se. If I face a trade-off between a past and a future event, and I have equal reason to prefer each, "feeling like it", can constitute a reason for future-bias.

Sullivan could reply that, while this may justify tie-breaker situations, it does not provide support for a systematic absolute discount function for past events: we are usually future-biased in a way so that we prefer bad past events over bad future events even if the past event is way worse. Lowry/Petersen and Chang can only defend a tie-breaker version of future-bias. However, this shouldn't really matter, as it is sufficient to reject a strict requirement to not be future-bias. Additionally, we are talking about approbative rationality and not about pragmatic gains: how beneficial or harmful an event is shouldn't matter for the type of argument Sullivan is advancing. Even if I am facing a trade-off between 10 hours of past pain and 1 hour of future-pain, I may have equal reason for both of them if we focus on

[79] This is the Scottish version of the classic mule in front of two haystacks-example.

approbative rationality alone and don't take into account which choice will leave me better off.

So, with Lowry/Petersen and Chang, premise (1) of the arbitrariness-argument can be challenged, since arbitrariness as such might not be ruled out by rationality requirements if "feeling like it" constitutes a class of reasons for attitudes if other reasons of rationality come to a tie.

Just be coherent

Secondly, one might just deny that the arbitrariness-argument as such is a thing. According to Sharon Street, what is reasonable for a rational agent to do is just whatever survives scrutiny. Street argues[80] that arbitrary preferences like the Detergents-Case aren't per se irrational – their irrationality can be shown only by pointing out a contradiction with other preferences and values the agent holds. If the agent is an *ideally coherent eccentric* with no contradiction in her belief- and desire-set, we could neither criticise her for Future-Tuesday-Indifference nor for discounting detergents based on location without taking a robust metaethical stance on objective values. If there are no normative facts independent of our beliefs and desires, Street would undermine the point of arbitrariness-arguments, since *anything* could justify preference changes as long as coherency is given from the agent's practical standpoint. And if we would provide a good explanation of how agents came to acquire these eccentric preference patterns, e.g. with an evolutionary story of how

[80] Street (2009), pp. 7-9.

Future-Tuesday indifference helped their ancestors survive, we might also find these cases a lot less counterintuitive.

Again, it is worth noting that while Street's position is radically anti-realist, she is not alone in thinking that rationality is mainly about coherence, and not about correctly responding to reasons.[81] If my beliefs are coherent, why would it be irrational for me to pick Tide from the upper shelf? And why would it be irrational for me to place more importance on the future than the past, as long as this does not contradict other beliefs I hold? I might not be responding to reasons, but why should that matter as long as I am being rational??

Sullivan, in response to this and to avoid infinite regresses about reasons, endorses externalism about prudential reasons in the sense that some facts just are normatively significant to agents, and some aren't, without the need for further justification.[82] This is also not uncommon, but opens up Sullivan's argument to several arguments against externalism about reasons, and makes her embrace of the arbitrariness-principle to defend temporal neutrality more difficult to accept.[83]

Actions and Choices

[81] See, for example, Broome (1999, 2005) for arguments in favour, and Kolodny (2005, 2007) for arguments against rationality as coherence.

[82] Sullivan (2018), p. 45.

[83] See Finlay andSchroeder (2017) for an overview.

In a recent paper, Kauppinen defends future-bias by rejecting Sullivan's suggestion of approbative rationality. He defends the so-called *Action Fixes Utility* principle.

> Only action fixes utility. If you act on the basis of assigning utility *u* to state of affairs S, rationality requires you to assign *u* to S whenever it is relevant to action or attitude, unless you gain new information about S.1 However, if you do not act on the basis of assigning *u* to S (nor have acted or ever will), it is rationally permissible to assign a different utility *u'* to S at different times without gaining new information about S.[84]

This principle basically says that, as long as my preferences do not lead me to act upon them, I can't really be criticised for having a certain preference pattern. So, for example, if I prefer Tide to Wisk and Surf, but never act upon this preference when doing my groceries, my mum really has no business criticising my preference pattern as irrational. And as long as I don't act upon my future-bias, why should Sullivan take issue with that?

The obvious response from Sullivan here is that it is still arbitrary to have this preference pattern without any rational basis. Kauppinen admits that this is not convincing to him, as very often, there is no further fundamental reasons to our preferences.[85] For example, my flatmate always wears her special Sunday-shirt on Sunday, even though the shirt isn't significantly

---

[84] Kauppinen (2018), pp. 240-241. Kauppinen restricts his principle further to hedonic cases only, which I believe to be unnecessary for our purposes.

[85] Kauppinen (2018), p. 244.

different from her other wardrobe, and Sundays don't hold any special meaning to her. So, what's the harm?[86]

Searching for Rational Grounds

I believe that if we consider all three arguments (feeling like it, being coherent, and not acting on it) together, it is already pretty clear that the arbitrariness-argument isn't off to a great start, even though we haven't even discussed premise (2). Considering that there are several, not ad-hoc concerns stemming from classical rationality theory about the arbitrariness-principle, I believe that we can reject Sullivan's argument on this basis alone.

However, I do think that Sullivan offers us a challenge to explore *why* exactly we should be future-biased. To illustrate this, let's pretend that (1) is not as shaky as it is. If (1) is accepted, it should be easy to see how the argument challenges the rationality of near-bias. If I arbitrarily prefer one event over the other merely because it is closer to the present, this mirrors the detergents-case, where I prefer something just based on a difference in location. So, premise (2) against near-bias seems robust, and since (3) follows from our definition of near-bias, the argument gets off the ground.

In the case of future-bias, this is not so clear. Premise

(2) Being past rather than future is an arbitrary difference between events.

[86] Special thanks to my favourite flatmate and her Sunday shirt.

of the arbitrariness-argument against future-bias isn't obviously mirroring the detergents-case. Whether "being past rather than future" is an arbitrary difference seems controversial, and there seem to be several asymmetries that we could appeal to, such as an asymmetry of control, asymmetry of possibility, and asymmetry of emotions. However, most of these, as Moller points out, at best explain why we have developed this bias – but it does not explain why future-bias is justified.[87] Just because we have developed in a way that we have a preference pattern tracking causation does not mean that this provides a justification for the preference pattern. We need a justification speaking in favour of future-bias. In the rest of the chapter, I will explore how we can use considerations from the metaphysics of time to meet Sullivan's challenge, to provide a positive reason for us to be future-biased.

3.3 A Short Look at Time

The most obvious explanation to (2) is that there is a metaphysical difference between past and future that constitutes a non-arbitrary difference in our attitudes. To explore what this means, we should look at what the metaphysical status of past and future is. There are generally two ways philosophers look at time.

Firstly, you may look at time as consisting in temporal periods we call "past", "present" and "future". The "past" is everything that lies behind us, the "future" is what lies ahead. The "present"

[87] Moller (2002), p. 79.

is what happens now, in this moment. These periods seem to be inherently different from each other, especially what is "now" seems to be a special moment in time. Only one time is "present" while all other periods of time are either "past" or "future". And what is "now" *constantly changes*, so what is "future" becomes the "present" and then falls behind us into the "past". This is what philosophers call the A-theory of time.

A-Theory There is an instant of time that is absolutely, irreducibly present.

Absolutely, irreducibly present means that there really is an objective moment in time that we can describe as "now", which cannot be explained away by referring to other properties of time, e.g. "earlier-than" or "later-than" relations. This contrasts with the so-called B-theory of time.

B-Theory No instant of time is absolutely, irreducibly present.

If you're a proponent of the B-theory, you'd think that what we call "past", "present" or "future" aren't objective descriptions of time, but rather something like metaphors of how we perceive things. Time can be completely captured in "earlier-than" and "later-than" relations between events, and there's no objective moment in time that could not be described as "earlier-than" or "later than" something. So, what we call "now" is just what we now see, but there is nothing metaphysically special about "now", it's like all other moments in time, "earlier" or "later" than other moments.

There are several versions of the A-theory of time. A common version of the A-theory is

Presentism There is an instant of time that is absolutely, irreducibly present, and that is the only instant of time that exists.[88]

According to presentism, "now" is all there is –past and future do not exist. What is "future" only comes into existence by becoming "present", and when it becomes "past", it ceases to exist. In other words, presentism combines the A-theory with the view that everything is temporary – things come into existence, and then cease from existence again. In contrast, if you combine the A-theory with eternalism, the view that everything always exists, regardless of past, present or future, you'll get

Moving Spotlight Everything always exists, and there is an instant of time that is absolutely, irreducibly present, changing from moment to moment.

This is the view Meghan Sullivan defends.[89] On this view, everything that is future already exists, and everything that is past is still existent. However, there is still a special moment in time, the "present" is the only time that is "now", and is distinct from "past" and "future", because it has this special glow of presentness, which moves along in time – hence the name of the theory. The last well-known A-theory is

Growing Block There is an instant of time that is absolutely, irreducibly present, and only the past and present exist.

---

[88] E.g. Bigelow (1996), pp. 35-36.

[89] Sullivan (2012), also Deasy (2015).

According to this view, things come into existence – from future to present – but they don't cease to exist, since everything that belongs to the present and past exists. Reality is like a growing block, where things keep coming into existence over time.

On the opposite side, the easiest explanation for the B-theory is that time is just another dimension – there is height, width, length, the three dimensions of space. And time is just like those, a fourth dimension, nothing else. It wouldn't make sense to assume that there is an objective "here" on height, so it doesn't make sense to assume an objective "now" on the time-scale. All objects already exist, and will exist – we can only see one part of the time-scale, but that does not affect the existence of past and future objects – they are just earlier or later to us at this moment, not inherently metaphysically different.

So, does that have anything to do with time-biases? The metaphysical asymmetry that would vindicate future-bias says that there is a non-arbitrary, metaphysical difference between past and future events. The truth of this claim varies with the theory of time we accept.

- On B-Theories, the claim seems false. There is no inherent metaphysical difference between past and future events that go beyond distance or difference in location.
- On Growing Block, the claim seems correct – future-events are different from past events, since future-events do not exist, while past events do.
- On Moving-Spotlight and Presentism, the truth of the claim seems unclear – on Moving-Spotlight, both past and

future exist equally, on Presentism both do not exist. So, you might think that makes the claim false, because there is no difference in metaphysical status in both. But you could also claim that the a-properties are different:

- on Moving-Spotlight, past events differ from future events in being already passed by the spotlight
- on Presentism, past events have ceased from existence while future events will come into existence.

Does that mean future-bias is permissible only to us if one of the A-Theories are true? Indeed, Hare uses the A-theory to justify time-biases via metaphysical asymmetry.[90] According to Hare, each A-theory provides an account of "all that there is"[91], so what the maximal state of affairs is like. And if we accept the A-theory, one feature of the maximal state of affairs is that present things are different from past and future things. So other things being equal, future negative events are just worse than past negative events[92].

Thereby, we arrive at a justification for future-bias:

---

[90] He argues that "If we accept four- dimensionalism, then [the problems of justifying time-biases] really are insoluble. With respect to grounding, past pains are just as real as future pains, so we have no grounds for thinking them less important simpliciter." Hare (2009), p. 16.

[91] Hare (2009), p.17.

[92] Hare (2009), p. 18.

> Harmony: When a time-biased person favours a scenario with less future pain, she thereby simply favours a better maximal state of affairs.[93]

Deng, discussing a defence of a past-future value asymmetry, argues similarly:

> "A process that is as metaphysically fundamental as A-theoretic passage is usually thought to be might be expected to have axiological effects as well. It could obliterate the intrinsic value of experiences as one after another relentlessly moves into the past, never to be experienced again. [...] According to A-theorists, there is something metaphysically important about the present, and time's passing involves a fundamental metaphysical change, for example, in which time is present. Why not think that this metaphysical change gives rise to a corresponding axiological one?"[94]

I'm not convinced that this is a strong defence. Even if the A-theory were true, it would still not be enough, in my eyes. The question remains as to why the metaphysical difference should be a non-arbitrary one. If past and future are different from each other, that's fine, but why does it matter normatively? The temporal asymmetry in that form seems to just consist in a mere repetition of the claim that the past is different from the future

---

[93] Hare (2009), p.10: Hare formulates Harmony for both near- and future-bias. I left out near-bias here because I'm solely looking at future-bias.

[94] Deng (2015), pp. 425-426.

– this does not entail that the past is different from the future in a normatively relevant way.

Hare replies to this as follows:

> "Would the metaphysics somehow make it seem plausible that future pain is in itself more significant than past pain? Well, [...], the metaphysics will not *explain* why future pains matter more than past pains. But maybe no explanation is needed. If the peacemaker has a strong (though perhaps defeasible) conviction that future pain is intrinsically worse than past pain, the metaphysics does nothing to undermine that conviction."[95]

But then we have arrived at the point at which we started – we have nothing to say about why future-bias isn't arbitrary and are subject to Sullivan's argument. If we want to defend future-bias, we need to do more than just to point out the metaphysical asymmetry. Just claiming that the past is different from the future does not get us far enough.

3.4 Parfit's Timeless Friend

We've seen that just pointing at the metaphysical asymmetry doesn't provide rational grounds for future-bias. However, metaphysics of time still tells us something useful. Another way of cashing out the difference between A-theory and B-theory is the question whether the passage of time is real. In this section, I show that future-bias is an attitude that responds to our perception of the passage of time.

---

[95] Hare (2009), p. 18.

What is time's passage? There are several explanations of what it means to say that time passes, some intuitive explanations are:

- Time's passage could mean that future events will come closer and closer to me while past events will move further away. We seem to be "moving" through time.[96]
- Time's passage could refer to "ontic becoming"[97]: When time passes, new things come into existence (and some might go out of existence)
- Time's passage could refer to A-properties: There are properties of "being present", "being past" and "being future", and changes in these properties is temporal passage.

Which of these explanations of what it means to say that time passes is most plausible is not necessary for us to establish. More relevantly, on all A-theories, time's passage is real. The reason for this is that the commitment to A-properties entails the passage of time. All relevant A-theories, from presentism, the moving spotlight to the growing block theory are committed to change[98] – some events happen that constitute change in the world, and this change expresses a "passing" of temporal properties from "future" to "present" to "past".

---

[96] Parfit (1984), p. 178, Prosser (2016), p. 25.

[97] Prosser (2016), p. 27.

[98] Prosser (2016), p. 9 mentions the possibility of an A-theory without time's passage, if there's only one time and no change. This is a solipsistic view, and will be ignored.

On presentism, time's passage is ontological in the sense that things come into and cease from existence when time passes. On the moving spotlight theory, there is passage in the sense that while everything always exists, some objects gain the property of being present and some lose this property again. And on the growing block theory, there is passage in the sense that things come into existence (like presentism), and then don't lose existence, but just the property of "presentness". Hence, on all A-theories, there is time's passage.

On the other hand, B-theorists deny that time's passage is real. The passage of time is merely an illusion, since there is nothing that becomes present and ceases to be present. But why believe this, as it sounds quite natural that time is passing? As Laurie Paul describes it, A-theorists argue for the reality of time's passage in the following way[99]:

(1) We have experiences of the nowness of events.
(2) We have experiences of passage and of change.
(3) The thesis that there are temporal properties of nowness and passage provides the only reasonable explanation of why we have these experiences.
(4) The thesis that there are temporal properties of nowness and passage provides the best explanation of why we have these experiences.
(5) Hence, there are temporal properties of nowness and passage.

[99] Paul (2010), pp. 338-339.

Paul herself denies (3) and (4) –providing a B-theoretical explanation of why we have experiences of passage and change by pointing out that our experiences come from our cognitive reaction towards successive change of properties in objects.[100]

Another B-Theorist, Prosser, denies (2) – we don't in fact experience time's passage. The argument goes as follows:[101] Imagine we are so technically advanced that we could build a machine that would detect the passage of time. The machine would have a light on it, and it would glow if and only if it detects the passage of time. However, regardless of whether the A-theory or the B-theory is true, the physical events described would be the same. So, whether the A-theory or the B-theory is true would make no difference in the outcome in the physical world. Hence, if the light of the machine would start glowing on the view of the A-theory, it also must glow on the B-theory. Therefore, the question whether time passes has no bearing on whether the light illuminates. Therefore, a machine detecting the passage of time is impossible. So, it's not possible to experience the passage of time.

Now, let me introduce you to Parfit's friend Timeless:[102]

> How Timeless Greets Good News: Timeless is in hospital for ten hours of painful operation. She wakes up without amnesia and can fully remember the operation.

---

[100] Paul (2010, p. 346.

[101] Prosser (2016), pp. 34-35.

[102] Case 2 from Parfit (1984), pp. 177-178.

> Parfit visits his friend on the day before and after the operation. On the day before, Timeless is distressed about the operation. On the day after, she is just as glum. "Why should I be relieved?", she asks. "Why is it better that my ordeal is in the past?"

Timeless' reaction is highly unusual. Is she making a mistake? Ought she to be relieved? Many of us would say yes – we would think that Timeless is reacting wrongly, maybe irrationally, because she is neglecting something.

Maybe you think that someone like Timeless is so bizarre that she's not possible anyways. But recall the beginning of the chapter, and assume that the real name of Parfit's friend is Louise Banks. Banks achieved a timeless perspective and manages to see how time really is by learning an alien language. This should be at least conceivable.

According to Parfit, one way of arguing that Banks is making a mistake is to appeal to the passage of time. Banks is failing to appreciate that time passes. This is the reason why she does not show future-bias. Both Parfit and Greene have characterised future-bias as an attitude responding to the passage of time.[103] Why is it? Let's take a closer look:

We feel relief when bad things become past because in a sense, they "go away" – if something bad, we anticipate it, when it is present, we perceive and experience it, and that perception and experience then "ends". So, we are future-biased because we feel

[103] Greene (Forthcoming), Parfit (1984).

bad things "passing" us, going from future to present and from present to past. This "feeling" of "passing" then is creating our future-bias. So, to develop Greene's point, if we take away the "feeling" of "passing" of events, we would also stop being future-biased.

If Banks would be future-biased, she would not have the same reaction. She would be relieved that the operation is over. Why would she be relieved? Because the event has passed and has moved away from her. If Banks would appreciate that time is passing, she would notice that the operation isn't present anymore. It has moved away from the present to the past, and thereby, she would react to it accordingly – in this case, since the operation is painful, she would be relieved that it's past. So, if Banks is future-biased, she would be because of time's passage. And if she'd appreciate the passage of time, she'd form an attitude reacting to the fact that the event moves away – she'd be future-biased. Therefore, time's passage is tracked by the attitude of future-bias.

So, future-bias tracks the passage of time. However, as you may have noticed, only A-theorists could make that argument. If we aren't accepting the A-theory, then the fact that time passes is no fact at all, but merely an illusion, and we could hardly blame someone for not falling for an illusion. If the B-theory is true, time does not pass in the external world, and Banks/Timeless should be praised for seeing reality and not falling for the illusion. On the B-theory, there would then be no explanation for future-bias, since there is nothing in reality the attitude responds to. So,

do we have to aspire to be like Louise Banks if the B-theory is true?

However, even if the A-theory is true, is Louise Banks making a mistake? Something is still missing – we have succeeded in tying future-bias to our sense that time is passing, but it is still not clear whether this justifies future-bias. The same problem that arose for Caspar Hare earlier on re-emerges – merely pointing at a metaphysical fact doesn't seem to show why this fact provides rational grounds. The point has been made by Moller:

> "The passage-view of time, even if correct, cannot in itself make any sense of our bias. [...] the reality of temporal becoming seems just irrelevant to the kind of justification we are searching for. Such a reality would do nothing to *justify* any asymmetry in our attitudes towards certain tensed facts, at most it could constitute an asymmetry our attitudes *track*."[104]

So, even if future-bias tracks a metaphysical fact like the passage of time – why is that a rational thing to do? Why is it not random or arbitrary to care more about the future?

So, we may have gained one step by tying future-bias to time's passage, but that does not suffice: If future-bias is grounded on time's passage, the question remains whether these grounds are rational – does the fact that the passage of time is real provide a reason to be future-biased? I'd like to continue where Moller's analysis ended and suggest firstly why future-bias is not arbitrary

[104] Moller (2002), p. 80.

as it is a psychological necessity, and then suggest a justificatory basis for future-bias. We've confirmed Moller's suggestion that future-bias *tracks* the passage of time – now, I'm going to explain why this is part of our rationality.

3.5 We Cannot Be Like Louise Banks

> *"Tell me one last thing," said Harry. "Is this real? Or has this been happening inside my head?" Dumbledore beamed at him, and his voice sounded loud and strong in Harry's ears even though the bright mist was descending again, obscuring his figure. "Of course it is happening inside your head, Harry, but why on earth should that mean that it is not real?"*

Recall that proponents of temporal neutrality view future-bias as irrational, and would like us to try and change, to not be future-biased anymore. We should be more like Louise Banks, and less like we are now. I will argue in this section that this demand does not get off the ground, as we cannot change towards becoming temporally neutral beings.

As we have already established that future-bias tracks the passage of time, I will now explain how passage of time is not a metaphysical fact of the external world but part of our perceptual apparatus – a way of understanding passage that is acceptable to a B-Theorist. In doing so, I will use an argument inspired by Kant's theory of time (without using Kantian vocabulary), and outline how there is a psychological necessity for humans, and human-like beings, to perceive passage of time. Since future-bias tracks the passage of time, it is psychologically impossible for us

to not be future-biased. Hence, proponents of the arbitrariness-arguments cannot demand a change in attitudes, and the arbitrariness-argument is blocked.

The A-theory and B-theory are metaphysical theories about the fundamental structure of the world. Their disagreement is about whether the passage of time is part of objective reality. An argument based on the reality of time's passage therefore depends on the truth or falsity of the A-theory – if we want to ground the permissibility of future-bias on the reality of time's passage as part of the fundamental structure of the world, this will only be possible if the B-theory is false.

The A-Theorists way, however, is not the only way to conceptualise the passage of time. Kant thinks that taking time's passage into account is necessary a priori because it's fundamental to all our perceptions and experiences. Kant describes time as a formal requirement for perception of objects and events[105]. As I cannot perceive objects without a sense of spatiality, I cannot perceive events without a sense of temporality – so by perceiving an event to happen, I necessarily "tense" the event, and set it into temporal relations[106] between now and when the event is happening. The event is "future" as I place it in a sequential ordering ahead of me now, and "past" due to it being ordered behind me.

---

[105] Kant (1889), A34.

[106] Kant (1889), B51.

I'll explain a bit more: When I look at an object, it is perceived by me as necessarily embedded in space and time. If I perceive an object, I cannot do so without the object being put into a spatial and temporal ordering. I am able to think of space and time without thinking of objects – e.g. I can think of empty space or times with no change. That tells us that time (and space) are a priori concepts that are not given to us by experience.[107] What I *cannot* do is see objects without putting them into some spatio-temporal ordering. I cannot perceive something without representing it in time and space.

When we represent an object in space, we represent it along length, width and height – for example, if I look at my copy of Kant's Critique, I'm representing it as being on my table, below the ceiling, etc. I cannot perceive it without representing the book in space. When it comes to time, we perceive objects over time as sequences of successions: when I look at my book at three different times, t_1 comes before t_2 comes before t_3..., and I notice that, as it is the same object, there is something "lingering on" with the object from t_1 to t_2 to t_3 – thereby giving us the idea of successive change of an object through it's relations between different temporal parts of my book.[108] At t_1, my copy of Kant's critique may be opened, at t_2, it may be closed again, and at t_3, it will have fallen to the floor. My perception of the book at three different times places it in a sequence of successive events – I automatically think, as a result of representing the book in a

[107] Rosenberg (2005), p. 69.

[108] Rosenberg (2005), p. 71-72.

(spatio-)temporal order, that changes over time happened to my book. So, in perceiving my book, I'm assuming that time passes, and thereby can see how changes in objects come to pass. Let us call these changes events.

This is not to say that time really passes as part of objective reality. It is rather that time's passage is part of our perceptual machinery – if we look at the world, we think of time as passing because this is part of how we perceive objects in the world. By placing objects into sequential ordering, we can see objects as changing, and can make sense of what events are. This sequential temporal ordering is our perception of the passage of time: by ordering the world into a sequence, we see things coming and passing, and thereby construct time's passage in our perception. Without this sense of passage, we would not be able to perceive events, as sequential ordering of objects constitute change in objects and thereby make us perceive them as events. Therefore, time's passage is a necessary condition for our perception of events.

How does that fit within A- and B-Theories? A-Theories think that time's passage is part of objective reality, while B-Theories deny this. Kant seems to have a more complicated answer to this: time's passage is real – but not as a part of the external, empirically accessible world, but as an a priori necessary condition of our perception of events.[109] According to Kant, we must perceive the world in a perspectival way, from a spatio-temporal point of view. In perceiving objects, we represent them

[109] Kant (1889), A34-B57.

in a temporal sequence, structured as passing from future to present and from present to past. That means that we cannot experience time's passage – whatever we do, we won't detect time's passage via empirical means, agreeing with the B-theory. However, time's passage at the same time is very real, agreeing partly with the A-theory: as a necessary component of our perceptual apparatus, it is part of the fundamental structure of the world via our perception – even if it is not part of the external world. In short, time's passage is only in our heads – but why should that mean that it's not real?

So, time's passage is part of our perceptual apparatus, and as future-bias tracks the passage of time, we are future-biased when we perceive time's passage. And as part of our perceptual apparatus, time's passage is a psychological necessity.

Why? Couldn't we also not perceive time as passing? Maybe after decades of meditation, Buddhist monks could, after realising the illusion of time's passage, finally free themselves from perceiving time as passing? I find this hard to believe: True, Buddhist meditation can achieve extraordinary feats, from raising body temperature to detaching the monk from their sense of self for a short time. However, breaking free from the psychological necessity of time's passage seems impossible to me, even if the self is successfully detached: even a detached person will still perceive the world through events and temporal sequences. And even if it might be metaphysically possible, ordinary human beings (including Buddhist monks) won't be able to achieve this level of detachment to perceive the world as timeless. So, in the relevant sense, time's passage is a psychological necessity.

As future-bias is tracking time's passage, and time's passage is a psychological necessity, it cannot be demanded from us to not be future-biased. If we are beings that are necessarily perceiving time as passing, and if future-bias follows this psychological mechanism in our head, it can also not be a demand of rationality to not be future-biased. This is similar to an appeal to the "ought implies can" principle advanced by Kant[110], and you might find this dubious, as "ought implies can" is usually applied to actions, not attitudes. But I think it is plausible enough to apply it to demands for attitude-change: why should an attitude be criticised if I have no psychological possibility to change it? Most attitudes we criticise – say, racist or sexist attitudes, overconfidence, and other types of irrational attitudes – we also demand a change by criticising them. And it seems perfectly possible to work on our racist attitudes, our sexism, our overconfidence – change might be difficult, but not psychologically impossible. The same cannot be said for time's passage and future-bias.

In summary, we cannot rid ourselves of future-bias, and therefore can also not be criticised for it via the arbitrariness-argument. This does not yet provide a rationale for being future-biased, but adds another obstacle to Sullivan's argument, as her argument

---

[110] "The action to which the "ought" applies must indeed be possible under natural conditions." / "Nun muss die Handlungs allerdings unter Naturbedingungen möglich sein, wenn auf sie das Sollen gerichtet ist." Kant (1889), A548/B578.

demands us to change an attitude that we cannot change. Therefore, it is permissible for us to be future-biased.

3.6 We Should Not Be Like Louise Banks

I will next explore a route to provide a rationale for future-bias by tying time's passage to our rational agency. I roughly follow these steps:

P 1 Future-bias tracks the passage of time.

P 2 Even if passage of time is not part of the outside world, it is constitutive of an agent's ability to perceive and experience events.

P 3 An agent's ability to perceive and experience events is a constitutive part of their agency.

C 1 Time's passage is a constitutive part of agency. (From P2 and P3)

C 2 Future-bias results from constitutive parts of agency. (From P1 and C1)

P 4 If an attitude is the result of constitutive parts of agency, the attitude is rationally grounded.

C 3 Future-bias is rationally grounded. (From C2 and P4)

By trying to connect future-bias to our agency, I aim to show that future-bias is based on rational grounds because it's a part of our agency itself. As part of our agency, we would lose something valuable without being future-biased: we would not be agents anymore. Hence, we should be future-biased. Let's look at each argument step in detail.

I have already shown something close to P1 above – if Louise Banks is future-biased, she'd perceive time's passage, and if Louise Banks perceives time's passage, she'd be future-biased. So, we have

P 1 Future-bias tracks the passage of time.

As we have already established P2:

P 2 Even if passage of time is not part of the outside world, it is constitutive of an agent's ability to perceive and experience events.

in the previous section with the help of Kant to provide an account of passage that is acceptable to both A- and B-Theorists, let's continue with P3.

P 3 An agent's ability to perceive and experience events is a constitutive part of their agency.

So, time's passage is real. But why should that matter? To show that something is normatively significant in a sense that it's not arbitrary to respond with a specific attitude to it, we need to show that the attitude is a result of some form of rational process. In short, I need to link time's passage to normative grounds. I will argue that time's passage, as part of our perceptual apparatus, matters because perception and experience of events constitute the possibility of rational agency.

A little side-story may be helpful here: Christine Korsgaard identifies the source of normativity within the agent herself. She describes normativity as a form of necessitation binding us to act – we must be agents because we are condemned to choices and

actions.[111] In order to have a reason to act, we must understand ourselves as unified agents – there is a formal requirement from our practical standpoint creating the normative drive in us. Because we are agents, we have reasons to act. Without us understanding ourselves as unified agents, we could not have reasons to act. Therefore, we have reason to "unify" or "constitute ourselves to a person.

Our argument is not about reasons to act, but our argument will take a similar shape. In order to form attitudes about events, we must be able to perceive events. Otherwise, we would not be able to form approving or disapproving attitudes about events. If I would not perceive events as a change, a temporal succession of objects over time, I would only see an object in space, with no temporal dimension. Kant thinks that this is not possible for us as reasoners. If Kant is right, then we cannot not perceive objects without time's passage, and therefore, time's passage is a necessary part of our rational process of forming attitudes about events. However, I don't think we need to go as far as to say that it's impossible for us not to see objects and events without temporal dimension. It is enough for us to consider what would happen in such a scenario. We would lose our agency.

Someone who does not see events in temporal order, someone who is not bound to the passage of time as a condition of her perceptual apparatus is Dr Louise Banks. Banks can see the world through a timeless perspective, where events do not follow each other successively – she can perceive all events simultaneously,

[111] Korsgaard (2009), 1.1.1-2.

perceiving past, present and future in an equal intensity. She "remembers" her daughter's death in the future as she remembers her daughters birth. She sees her first meeting with her husband as clearly as their divorce. In short, Louise Banks is not bound to Kant's necessary condition for the possibility of perception – she can perceive all objects and events of the world without using the lens of temporal succession.

Let's assume that this is possible. We can conceive, without obvious contradiction, that it would be possible for Louise Banks to achieve this perspective, be it by learning Heptapod B, or via gift from the gods. Kant would of course disagree, but let's assume that he's wrong.

In the short-story, and more drastically in the movie, Louise Banks robs herself from the ability to form attitudes about events in her life. She ceases to see the death of her daughter as something bad, and reacts to her husband divorcing her with indifference. Her husband is outraged that Banks, fully knowing that her daughter will die from a terrible disease, did not act otherwise in light of a clearly bad event. In the short-story and movie, her attitudes become different, she seems to not be able to form strong approvals or disapprovals towards events anymore. The more she sinks into a timeless perspective, the less she seems able to form attitudes about events in her life.

Why would Louise Banks develop in such a way? One possible explanation would be that her perception of all events simultaneously takes away her sense of change. Without a temporal sequence, events would be "just there", and no

successive ordering of events would take place, thereby robbing Banks of her drive to respond to events. I believe this is the correct explanation. Another possible explanation is that her perception of all times would make her feel everything at once, thereby rendering her numb due to the multitude of pains, pleasures, desires and other feelings and mental states being present at once. I also think this is a plausible explanation.

Recall, for a second example, Parfit's operations case:

> How Timeless Greets Good News: Timeless is in hospital for a painful operation. She wakes up without amnesia, and can fully remember the operation.
>
> Parfit visits his friend on the day before and after the operation. On the day before, Timeless is distressed about the operation. On the day after, she is just as glum. "Why should I be relieved?", she asks. "Why is it better that my ordeal is in the past?"

Would Timeless be able to form attitudes about her past or future operations? By assumption, she is distressed about the operation, and then does not cease to be distressed afterwards. However, if Parfit's friend would be truly timeless like Louise Banks, and would be able to perceive her operation before it takes place, as well as after it happens, on what basis would she be distressed? If the distress is caused by the impending pain, that should not be the case with Louise Banks, because the pain is not imminent or approaching in any sense – the pain can already be perceived. So, the distress cannot come from the approaching bad event, and Case 1 of Parfit's scenario would not be possible – if she does not

remember the pain of the operation itself, she can also not be distressed by it.

In Case 2, she would either constantly feel the same distress as during the operation due to the pain, or no distress at all – or she'd constantly feel the same level of painful distress, because without perceiving events as temporal sequence, she would not be capable of distinguishing between present and non-present events, or be unable to form attitudes about the events at all. Although the first option seems possible – her distress would come directly from the pain of the operation, it looks unlikely that someone can constantly feel the same level of distress about the same event, without becoming numb to it after a while. So, we arrive at the same two explanations as above – either, Banks loses her perception of events and does not develop attitudinal responses, or she becomes numb from the pain being constantly there. In either case, without seeing the world through temporal ordering, she would not be able to form a responding attitude to the event anymore.

In short, without time's passage in our perceptual apparatus, we would lose something of value. We would lose our ability to form attitudes towards events and thereby cease to be rational agents. Even if it is possible to see the world not as temporally ordered, we would thereby lose our agency. And this is why time's passage matters – time's passage is a constitutive component for our capacity of rational agency. Without perceiving events in temporal succession, we would not be able to form attitudes towards events. In other words, time's passage is a formal requirement for us to be rational agents. Without it being part of

our perceptual apparatus, we could not have attitudes. Since our agency is something of value, we therefore have reason to see the world through the lens of time's passage. In other words, we should not be like Louise Banks.

In summary, I have shown that future-bias is an attitude responding to time's passage (P1), time's passage is not part of the external world, but component of our perceptual apparatus (P2) and that our perception of events in a temporally ordered way plays a constitutive role in our ability to form attitudes towards events (P3). From here, we arrive at our first conclusion:

> C1 Time's passage is a constitutive part of agency. (From P2 and P3)

As a short side-note, this corresponds nicely with a view outlined by Simon Prosser: that indexicals are essential as they provide a perspective necessary for agency and that without an understanding that it is *me*, I cannot gain motivation and reason to act, even if I otherwise fully understand a proposition about the person that is me.[112] C1 is similar in the sense that without a temporal perspective, we would also not be able to act as rational agents. Also, since we have already discovered that future-bias is the responding attitude to the passage of time, we arrive at our next conclusion:

> C2 Future-bias results from constitutive parts of agency. (From P1 and C1)

[112] See Prosser (2015).

We now have established that future-bias is a part of our rational process forming attitudes and preferences about events. If I perceive the world as a sequence of events, which is necessary for me being an agent, then I will also be future-biased. But a last step needs defence before we can conclude that future-bias stems from rational grounds. Is everything that comes from us exercising our rational agency also rational? Let's look at:

> P 4 If an attitude is the result of constitutive parts of agency, the attitude is rationally grounded.

This may look like a strange premise to defend. If something is part of our rational agency, does our rationality approve of it? This might sound like a circular question. But to give some plausibility to P4, we should return to Korsgaards search for sources of normativity. As outlined above, Korsgaard thinks that we have reason to be agents in virtue of us being agents. This may also sound circular, but recall that Korsgaard thinks of us being agents as a necessity constituted by our agency. I cannot decide against being an agent, because it requires agency to decide like that. I can choose not to act, or act irresponsibly, but this requires me making a choice. In not acting, I am exercising my agency by choosing a certain course of how to direct myself. Thereby, I am condemned to agency: my agency is rationally grounded in the necessity of being an agent. In a similar sense, I believe we can defend attitudes resulting from our agency as rationally grounded. An attitude that results from a constitutive part of our agency is not arbitrary but built on rational grounds: our agency itself. Hence, an attitude that results from a

constitutive part of our agency is also rationally endorsed by our agency.

Now, we can finally arrive at the final step:

C3 Future-bias is rationally grounded. (From C2 and P4)

3.7 Conclusion

Recall the arbitrariness-argument against future-bias, as outlined by Sullivan:

Arbitrariness-Argument against future-bias[113]

(1) It's not rationally permissible to vary preferences according to arbitrary differences. (Arbitrariness-Principle)
(2) Being past rather than future is an arbitrary difference between events.
(3) If an agent is future-biased, her preferences are sensitive to events being past or future.
(4) Therefore, future-bias is not rationally permissible.

I've mentioned in the beginning that the argument can be resisted successfully on denying (1). But if my argument has successfully shown

C3 – Future-bias is rationally grounded

then (2) of the arbitrariness-argument against future-bias is also false. There is a non-arbitrary difference between past and future events: the passage of time, as part of our perceptual apparatus

[113] Sullivan (2018), p. 73, Brink (2010), p. 4.

ordering events into a temporal sequence. Without this, we would not be able to form attitudes and preferences towards events. We would become like Louise Banks, losing something of value: our temporal perspective that enables us to be rational agents. Therefore, the arbitrariness-argument against future-bias fails, and future-bias is rationally grounded.

At the same time, my argument does not support near-bias. Whether an event is closer to us in the future than another is not grounded in our perception of events. As Parfit describes it, the temporal asymmetry between past and future is constituted by the future moving towards the "now" and the past never again becoming "now" and drifting away from us. In the case of near-bias, both events in questions are future, so both will move towards us in a similar way:

> "Time's passage does not justify caring more about the near future, since, however distant future pains are, they *will* come within the scope of "now"".[114]

Hence, near-bias is ruled out by the arbitrariness-argument, but future-bias isn't.

But even if my argument does not succeed, I have shown in the previous section that Sullivan's arbitrariness-argument can be blocked by an appeal to ought-implies-can: time's passage is a psychological necessity, and as future-bias tracks the passage of time, we cannot demand ourselves to be temporally neutral.

---

[114] Parfit (1984), p. 180.

Either way, the past isn't arbitrary, and future-bias is rationally permissible.

Daniel Kahneman describes a bias as a flaw in the reflective mind, a failure of rationality.[115] Biases are a result of un-reflected, impulsive intuitions, which can sometimes be rooted out by putting effort into carefully reflecting upon and assessing the bias. Following this description, future-bias should not be considered a bias. Not only does it persist even after careful reflection and deliberation – it seems that if we get rid of future-bias, we would lose something of value. We would, if we'd follow Louise Banks, lose a feature of us that enables us to connect to the world, to its events and to other persons. Therefore, I suggest that we should stop calling the attitude "future-bias". We should call it "preference for the future" instead.

[115] Kahneman (2011), p. 48-49.

4 Can We Debunk the Rationality of Future-bias?

Since the dawn of time, philosophers have pondered how to teach philosophy to students. A particularly stressful method is exams, and at least since Arthur Prior, it has been recorded that the prospect of exams causes both teachers and students to feel dread, while the end of the exam period causes them to feel relief.

Feelings of dread and relief are so-called tensed attitudes, as the attitudes refer to time passing and events moving on.[116]

> Tensed Emotion: An emotion is tensed if it is directed at a past, present, or future event or state of affairs

These tensed attitudes are used by some philosophers as evidence for the rational permissibility of time-biases. If we are time-biased, it makes a difference for us when an event is taking place. For example, if Arthur Prior were near-biased, he'd prefer exam marking to be next month rather than next week – bad events should be further away in the future, and good things closer to the present. If Prior were future-biased, he'd want the exam markings to be in his past, rather than his future – bad things should be past, over and done with, while good things should be future, ahead of us.

---

[116] Maclaurin and Dyke (2002), p. 278.

It is quite normal for us to be time-biased, as the emotions connected to the biases are very common and natural to us. It's another question, however, whether it's rational for us to be time-biased. Recently, a lot of philosophers have argued that time-biases are not rationally permissible, and we should strive to be temporally neutral instead – whether an event is past, present or future should not as such matter in the events evaluation.

Proponents of temporal neutrality have offered three general ways of arguing against the permissibility of time-biases.[117] Their arguments aim to show that

a) There are no reasons in favour of being time-biased
b) There are reasons not to be time-biased
c) There is a debunking explanation that applies to the rationality of time-biases.

All three argument types have been advanced by Preston Greene and Megan Sullivan in their paper *Against Time-Biases*, as well as in Sullivan's book *Time-Biases: A Theory for Rational Planning*. She argues that a) nothing speaks in favour of being time-biased as time-biased preferences stand on arbitrary non-rational grounds, b) being time-biased will lead us to make choices that make us worse off than we would be if we were temporally neutral and c) there's something like a debunking story to tell about irrelevant influences shaping our time-biases and tensed attitudes.

[117] Moller (2002), p. 68. "permissibility of" added by me.

This chapter will mostly be concerned with c), whether there is a successful debunking explanation that shows that the basis of the formation of time-biases is unreliable and should therefore give us a reason to doubt the truth of our belief that time-biases are rational. Greene and Sullivan advance two attempts to shake the credibility of our beliefs that time-biases are rationally permissible: First, our tensed emotions have been subject to evolutionary influence, which makes us accept time-biases as rational, and second, the affective states that are part of our tensed emotions are an irrelevant influence on our rational process. If we use rational reflection to separate both evolutionary influence as well as separate affective state component from belief components in our tensed emotions, we will be inclined towards temporal neutrality.

I will argue that both debunking attempts fail, for similar reasons that debunking attempts against moral realism fail. In the first half of this chapter, I will outline an evolutionary debunking argument against time-biases, as well as a debunking argument based on emotions, and examine why both are unconvincing.

In the second half of the chapter, I will do the reverse and explore whether evolution or emotions can provide rational grounds to being time-biased. I will in particular explore whether future-bias's evolutionary advantages can provide rational justification, and whether our tensed emotions, like grief or nostalgia or relief, can provide rational grounds for attitude patterns like time-biases.

In summary, not only do both evolutionary and emotional debunking arguments fail, adaption to control and tensed emotions might in some cases provide us with rational grounds to be time-biased, and we can answer both a)-style and c)-style arguments from proponents of temporal neutrality.

4.1 Debunking Time-Biases?

We should firstly clarify what Sullivan and Greene actually want to propose in terms of evolutionary influences on time-biases, and whether it actually is an evolutionary debunking approach a la Sharon Street with moral realism.[118] Generally, debunking arguments are attempts to show that the source of our beliefs is not reliable by providing an explanation how your beliefs are very likely to be mistaken and why you've been led to believe them, e.g. by showing how irrelevant influences have contaminated your belief forming process. As a result, you cannot rationally maintain your beliefs anymore.

Let's look at what Sullivan is saying in her book:

> "The evolutionary account of time-biases gives us an error theory for why we've been tempted to discount the past and distant future, even if it isn't prudentially rational to have such attitudes. [...] We are susceptible of time-biased preferences because we have strongly temporally asymmetric emotions. There emotions are adaptive, which

[118] See Street (2006). Her argument in short poses a dilemma for moral realists who both accept that there are robust moral facts as well as evolutionary influence to our moral beliefs.

> offers an explanation for why we have them. And for the kinds of scenarios our evolutionary ancestors needed to reason about, they were highly successful. Indeed, for times when we are not capable of sustained deliberation about our reasons, it isn't such a bad idea to rely on these emotions."[119]

Sullivan here talks of giving an error theory why we are time-biased, even though we shouldn't be. I believe her approach is more like a debunking attempt, which, is not directly targeted at the rationality of time-biases, I think.[120] In their paper, Greene and Sullivan first provide systematic arguments why time-biases are irrational. After presenting their arguments, they assume that, even if their arguments are successful, time-biases might still seem so rational that these intuitions are taken as evidence to dismiss their arguments – regardless of how well their

---

[119] Sullivan (2018), pp. 90-91.

[120] To clarify the difference between error theory and debunking a bit, this is what I understand an error theory as:

> Error Theory: To provide an error theory about X is to explain how none of the claims in a discourse about X are true.

For example, an error theorist about moral realism explains how none of the sentence in the discourse about moral facts are true. See Daly and Liggins (2015), p. 209. Contrast this with

> Debunking Explanation: To provide a debunking explanation of X is to explain how irrelevant factors led us to our likely mistaken beliefs about X.

Here, the focus is not on sentences being false, but that our beliefs are likely to be prone to error, and that we should be sceptical of their truth as a result. See Vavova (2014).

arbitrariness arguments and compensation arguments work, they have to be wrong because time-biases feel so right. So, to respond to this dialectic situation, they provide us with an explanation ("error theory") that aims to convince us why our intuitive response is mistaken, or led astray, so that we are more likely to accept their arguments in favour of temporal neutrality. In short, they respond to the following way of thinking:

a) There is a systematic argument that time-biases are irrational.
b) My intuitions tell me very strongly that time-biases are rational.
c) If my intuitions tell me very strongly that X is rational, I can take them as evidence for X, even in the face of a systematic arguments that X is irrational.
d) If I have evidence for X's rationality that holds even in the face of a systematic argument that X is irrational, I can maintain my belief that X is rational.
e) Therefore, I can maintain my belief that time-biases are rational.

Sullivan and Greene's target is c): by providing either an error theory or a debunking explanation of our intuitions behind time-biases, they want us to accept their systematic arguments, however counterintuitive they may seem at first glance. We can roughly reconstruct Sullivan's argument like this:

Soft Debunking

(1) If my intuitions on X have been subject to irrelevant influences, I should not take them as evidence for or against X.

(2) My intuitions about time-biases have been influenced by tensed emotions.

(3) Tensed emotions are irrelevant influences on my intuitions.

(4) Therefore, my intuitions about time-biases have been subject to irrelevant influences.

(5) Therefore, my intuitions about time-biases should not be taken as evidence for the rationality of time-biases.

If we take the recent debate about debunking of moral realism into account, this argument could be pushed further to establish that not only should we not use our intuitions on time-biases as evidence, the fact that they've been subject to irrelevant influences like evolution should lead us to revise our beliefs in the rationality of time-biases:

Hard Debunking

(6) If you have good reason to think that your belief is mistaken, then you cannot rationally maintain it.

(7) If my belief that time-biases are rational are based on intuitions that have been subject to irrelevant influences, I have good reason to think that my belief is mistaken.

(8) My belief that time-biases are rational are based on intuitions that have been subject to irrelevant influences.

(9) We cannot rationally maintain our belief that time-biases are rational.[121]

I do think that Greene and Sullivan argue for (5) at the end of their paper, and Sullivan does the same in her book. I believe that they don't explicitly go further towards the more aggressive debunking attempt to argue for (9), but I also think that some of what they say hints at hard debunking, as they clearly are quite sceptical of irrelevant influences to our rational deliberation process, regardless of whether it's evolution or emotion.

In any case, if Greene and Sullivan succeed in establishing soft debunking, they can rest easy, as we now have a much easier time to accept their arguments that time-biases are irrational due to arbitrariness and pragmatic loss. And if anyone would like to go further, hard debunking would establish an independent argument against time-biases, and Moller's

c) There is a debunking explanation that applies to time-biases

is established. I will argue that it's irrelevant which form of debunking explanation Sullivan and Greene want to provide, because they fail to establish

(3) Tensed emotions are irrelevant influences on my intuitions.

---

[121] Heavily inspired by Vavova (2014), p. 91. Also, see Street (2006, 2015) and Vavova (2015).

Greene and Sullivan provide two explanations why tensed emotions are contaminating our intuitions, evolutionary forces, and that emotions themselves are non-rational. I will argue that both fail, and hence, their debunking attempts on time-biases can be rebuffed.

4.2 Evolutionary Debunking

So, let's first unpack what Sullivan said to brew a proper evolutionary debunking argument step by step. Let's look at the first premise:

(1) If my intuitions on X have been subject to irrelevant influences, I should not take them as evidence for or against X.

Even if this sounds somewhat plausible on first glance, note that this first premise is already very controversial, as there is a big risk of overgeneralising and ruling out using any intuitions as evidence for anything. After all, all our intuitions on moral, rational or even metaphysical cases in philosophy have likely been causally contaminated by irrelevant influences and worries about irrelevant influences should be distinct from general scepticism. Vavova proposes that irrelevant influences are a distinct worry and do not collapse into old-fashioned scepticism if the worry is based on us having a good reason to think that we are likely to be mistaken.[122] So, we should revise (1) a wee bit:

[122] See Vavova (2018), pp. 144-145.

(1*) If I have good reason to think that my intuitions on X are mistaken because of being subject to irrelevant influences, I should not take them as evidence for or against X.

The difference between (1) and (1*) is the difference between irrelevant influence and epistemically problematic influence. Consider the classic example of G. A. Cohen, who chose Oxford over Harvard for graduate school. Oxford philosophers tend to accept the analytic/synthetic distinction, while Harvard philosophers tend not to, even though everyone studied the same arguments. But location of your study should be irrelevant to the truth of a philosophical claim, so should Cohen worry about how he came to believe the analytic/synthetic distinction? According to (1), he should, but with (1*), he can relax a bit, until he finds a good reason to think that he was led astray, e.g. Oxford folk trying to secretly manipulate him.[123]

Note, however, that this pushes the argument closer to *Hard Debunking,* as outlined above, as we need to think that there is a good reason for us to doubt. This also means that we could straightaway push for a stronger conclusion, and not only not take our intuitions not as evidence, but doubt the rationality of time-biases altogether, based on the debunking argument. We also need some kind of auxiliary principle that resembles (7) from above,

[123] See Vavova (2018), or Cohen's classic example from 2000, p. 18.

(7) If my intuitions on X have been subject to irrelevant influences, I have good reason to think that my intuitions are mistaken.

to make our conclusion at (5) valid. Now, do we have a good reason to think our intuitions are mistaken based on irrelevant influences? This depends on whether we can establish the following argument steps:

(1) My intuitions about time-biases have been influenced by tensed emotions.

I would like to grant (2) to Sullivan, who argues that we come to believe that time-biases are rational preferences because of our tensed emotions. As Sullivan says, we do genuinely believe this, because are tempted to believe it from those emotions that are temporally asymmetric, like grief, dread and anticipation. We believe that our time-biases are results of our rational deliberation processes because of those emotions. While this could be resisted, I find it plausible enough, and many authors, including Parfit and Greene have commented on how tensed emotions lend plausibility to time-bias. The trickier bit will be to show that the influence of tensed emotions is an *irrelevant* one.

Here, evolution comes in: Our belief that time-biases are rational preferences are not resulting from our rational deliberation process, as the tensed emotions responsible for our belief are adaptive and do not track normative truth – they track survival and fitness, not rational sustained deliberation. Now, we need a plausible story about how evolution has influenced our beliefs that time-biases are rational. Sullivan provides those:

Firstly, emotions associated with bias towards the near are a heuristic for tracking probabilities. Our ancestors have evolved to place priority to the near future over the far future as they faced more immediate threats and challenges that did not involve the necessity for long term planning. So, tensed emotions like dread and anticipation for closer events rather than events far away in the future track the uncertainty involved in far-future events and the certainty of near future events and the need to decide about them first. When Prior dreads the exams next week much more than those in a month, it is because he evolved to feel this, as it was advantageous for his ancestors to focus on closer threats and challenges – even if the exams in a month are much worse, he will focus his attention on those to mark in a week. Hence, near-biased emotions were evolutionary advantageous for our ancestors.[124]

Secondly, emotions associated with bias towards the future are a control-heuristic. We have evolved to care much more about the future rather than the past because of the direction of causation and our ability to influence events. We can never change the past, but can influence the future, so our tensed emotions like grief, nostalgia or relief track this, as it is much more evolutionarily beneficial to focus on challenges in the future instead of crying over spilt milk. When Prior has finished marking exams and, feeling intense relief, shouts "Thank Goodness that's over!", he feels so because he evolved to not focus his care and attention on

[124] Sullivan (2018), p. 89.

past threats and challenges anymore, and his relief reflects that. So, future-biased emotions were evolutionary advantageous.[125]

A few clarifications: "rational" here means either it being beneficial, something that is good for you in virtue of its effects, or as intrinsically appropriate or fitting to have. Some attitude being rational means that that it speaks in favour of having time-biases, either in the sense of making your life go well, or in the sense of being a fitting response to a certain situation.

So, in short, we now arrived at

(1) Tensed emotions are irrelevant influences on my intuitions.

as evolution has pushed our tensed emotions in a way that have tracked survival and fitness for our ancestors. Premise

(4) Therefore, my intuitions about time-biases have been subject to irrelevant influences.

follows from (2) and (3), and

(5) Therefore, my intuitions about time-biases should not be taken as evidence for the rationality of time-biases.

follows from (1), (1,5) and (4). Sullivan's argument is complete, and hopefully faithfully reconstructed, and we can dispel the intuitive force behind time-biases.

[125] Sullivan (2018), pp. 89-90.

4.3 Problems with Evolutionary Debunking

Regardless of whether you would like to push further towards *hard debunking*, and establish that evolutionary influence on its own constitutes a reason to doubt the rationality of time-biases, or whether you would like to stay on *soft debunking*, only going so far to say that we should not use intuitions contaminated by evolution as evidence, I believe the argument fails. I share the scepticism on evolutionary debunking arguments expressed by Vavova against evolutionary debunking of moral realism, and I believe her take applies to evolutionary debunking of time-biases just as much.

Vavova outlines several arguments against the debunker, but I will focus on one of her arguments that applies most clearly to the debunking attempt against time-biases, regardless of soft or hard. Vavova explains that an evolutionary debunking attempt is overgeneralising and does not limit itself appropriately on one specific target.[126] The debunker is always at risk of debunking not only the target beliefs, but the entire class of these beliefs, sometimes collapsing into general scepticism (which is uninteresting) or undermining enough to make the debunking argument self-defeating.

Think about it like this: If you outline a debunking attempt to show that our moral beliefs should be subject to scepticism, why should that not go for all evaluative beliefs? We can build an argument in the same way for all rational and epistemic beliefs,

[126] Vavova (2014), p. 90.

thereby generalising the argument to a much stronger sceptical attack on our belief systems. But if we debunk all our evaluative beliefs, the debunker undermines her own argument. For example, a premise like

> We have good reason to think that our moral beliefs are mistaken.

is undermined as we can't assume to have a good reason to be mistaken about our beliefs anymore – the debunker cannot provide us with a good reason to think that we are mistaken because all reasons are at stake.[127]

Compare this with the debunking challenge against time-biases. The challenge aims to undermine the intuition that time-biases are rationally justified, which is a much narrower target than debunking moral realism. However, the debunker will struggle to keep it as narrow as she wants, because the argument will target more than she wants it to target.

Firstly, the evolutionary challenge can be extended to all emotions, not just tensed ones – it is quite obvious to assume that if evolution has influenced tensed emotions, then it also has influenced other emotions – so everything in the realm of rationality with a connection to emotions is potentially infected by evolutionary influence. But it seems ridiculous to assume that my emotional reaction to pain cannot ground the rationality of pain-avoidance. So, the evolutionary challenge targets too much already. More on emotions later.

---

[127] Vavova (2014), p. 88.

Secondly, the evolution challenge can be extended to all kinds of intuitions, not only those justifying the rationality of time-biases: if some intuitions about our rationality are at stake from evolutionary influence, why not all of them? Why not the entire process of rational deliberation?

Here might be a possible evolutionary debunking attempt for all intuitions concerning rational justification:

(1*) If I have good reason to think that my intuitions on X are mistaken because of being subject to irrelevant influences, I should not take them as evidence for or against X.

(7) If my intuitions on X have been subject to irrelevant influences, I have good reason to think that my intuitions are mistaken.

(2) My intuitions about rationality have been influenced by tensed emotions.

(3) Tensed emotions are irrelevant influences on my intuitions.

(4) Therefore, my intuitions about rationality have been subject to irrelevant influences.

(5) Therefore, my intuitions about rationality should not be taken as evidence in rational justification.

This argument would threaten temporal neutrality, as all our rational requirements are at stake. Of course, Sullivan might point out several issues here. She might, for example, say that rationality and adaptive advantage overlap, but there are enough

cases where this might not be the case – to use a crude example, we have evolved to be racist and xenophobic, which at some point might have been advantageous for your ancestors to stick to those familiar to them, but that might not make *your* life go well. And again, from this point, we can again generalise the argument further to target intuitions concerning all evaluative beliefs, including epistemic ones, making the argument self-defeating.

Sullivan could also object that influence from evolution comes in degrees: while time-biases and the intuitions behind them have been subject to a lot of evolutionary influence, other beliefs, intuitions and features of our rationality may have been less or not influenced by evolution. Sullivan could say that we can tell a convincing story about evolutionary influence on tensed emotions and time-biases, but that's not the case for temporal neutrality or rational reflection in general.

The problem with this kind of response is that telling a convincing story about evolutionary influence is an arbitrary criterion that does not serve to shield the entire process of rational deliberation. We can very well tell a story about how our rationality process gave us evolutionary advantages, thereby upholding a general debunking of all our intuitions in rational deliberation. If Sullivan thinks that evolution has influenced a part of our rational deliberation process, why should some parts of it be unaffected or immune? Hence, the evolutionary debunking attempt of time-

biases ends up threatening our rational reflection process generally and is thereby already blocked.[128]

I believe this to be decisive already. However, a second problem for an evolutionary debunking attempt is that you don't seem to necessarily need evolution to do so. In both the article by Green and Sullivan, as well as Sullivan's book, the authors express scepticism about emotions, feelings and affects generally contaminating rational reflection. If this is the case, evolution simply might be an unnecessary, overly complicated component in the debunking attempt, and should be dropped anyway. So, let's set evolution aside for a moment to have a look at emotions and rationality.

4.4 Are Emotions Bad Influences?

If we look at an earlier article Sullivan wrote together with Greene, their general vibe seems to suggest that we shouldn't trust our mere feelings when it comes to rationality because they are irrelevant to rationality.

> "[...] if such relief is understood only as a "feeling," then it seems appropriate to view this feeling not as rational or irrational, but rather as a nonrational response to one's situation."[129]

Or, even more clearly:

---

[128] For more arguments against evolutionary debunking, see Vavova (2014).

[129] Greene and Sullivan (2015), p. 967.

> "Without these affect-laden mental states distorting our preferences we are inclined toward complete temporal neutrality."[130]

So, Greene and Sullivan seem be uneasy about feelings when it comes to evidence, justification and rationality. Instead, they'd like a rational agent to engage in reflection, weeding out irrelevant disturbances such as feelings we just evolved to have.

This leads to similar problems we have already encountered with our evolutionary debunking attempt: why is it that rational reflection can reliably weed out feelings without the process being affected by them itself? If affect-laden mental states distort our preferences, why do they not distort our reflection ability? How do we limit the irrelevant influences, be it evolution or feelings, to only affect one part of our belief and preference systems, but not others, like our capacity to reflect? After all, it is quite likely that feelings have affected our capacity in rational reflection, and as we know from Kahneman, we are very often less rationally reflective than we think we are. A lot of the times our intuitive base system makes a decision and our rational reflection process just follows or justifies what our base system said.[131]

---

[130] Greene and Sullivan (2015), p. 970.

[131] This is better as system one vs system two: System one is quick, intuitive and impulsive, system two is slow, deliberative and energy intensive. We very often use system one and system two simply follows suit and constructs a story around why the decision made by system one was a rational one. See Kahneman (2011) for more detail.

Even if Greene and Sullivan can find a way of showing how our rational reflection process is not contaminated by affect-ladden states, another problem remains: it's not mere feelings, but tensed emotions like grief, nostalgia, relief, that make us inclined to accept time-biases as rational. Emotions are much more complex than mere feelings, and here, it's clearly emotions connected to time-biases, not just feelings pushing us. But why we should be sceptical of emotions? Why should emotions be non-rational? This again depends on your view of what rationality is supposed to be, and what you take emotions to be. However, regardless of how you view this debate, the claim that emotions per se are non-rational without rational reflection is most certainly false.

The Rationality of Emotions

In her book, Sullivan's sceptical take on the rationality of emotions is a bit more detailed, distinguishes between emotions and feelings, and seems to involve a general sense of emotions having affective components that are not really our own. Even though she doesn't go into much detail, she outlines that even if we do not take the classic Enlightenment stance of emotions being sharply distinguishable from beliefs, "emotions are rational or irrational to the extent that they have rational or irrational preferences and beliefs as components."[132] But if we focus on the affective part of emotions like relief, joy, grief and so on, they really are non-rational. She adds that, the more mature we get, the more capable we are of controlling these emotions and we can

[132] Sullivan (2018), p. 89.

separate them from our beliefs about time-biases to become more temporally neutral.

That's a more nuanced claim and addresses tensed emotions as complex mental states with different components, and not as pure affective states, but Sullivan is still taking a controversial stance here.

There are generally two camps in the debate on what emotions really are. One faction is called the so-called cognitive theory of emotions, or the propositional attitude view, which holds that emotions are a kind of mental state that expresses a proposition or a judgment about something.[133] For example, if I feel anger, I am expressing my stance at something that has wronged me or someone close to me. The other faction is the non-cognitivist or affect-based school of emotions, according to which emotions are not mental states that express propositions, but automated feelings that simply arise within us in response to situations.[134] My anger does not express a judgement about the world, I just happen to be in a situation that gives rise to my anger.

Given these two camps, Green and Sullivan seem to belong to the latter camp, as they treat emotions as something that is affect-based or at least "affect-laden", something that we neither have any control over, nor can we assess it in any way with our mind, as emotions escape the realm of rationality.

---

[133] E.g. Nussbaum (2004), Solomon (1988).

[134] E.g. Griffith (1997).

However, even on an affect-based view, this seems to be way too quick. Even if emotions are mere affects, there's still three ways of assessing emotions in terms of rationality: an emotion can be assessed on whether it's beneficial, fitting or warranted.

> Beneficial: An emotion is rational if it is beneficial in terms of bettering an agent or furthering her ends.[135]

Prior's relief after exam marking is rational as it makes him appreciate his own research time more and makes him use it better.

> Fitting: An emotion is rational if the situation in which it is felt is relevantly similar to the scenarios that are paradigm cases of situations in which the emotion is felt.[136]

Prior's relief after exam marking is rational as it is just like when his exam marking last year was finally done.

> Warranted: An emotion is rational if an agent has sufficient evidence that the emotional response is fitting to a situation.[137]

---

[135] Kerr (2014), p .50.

[136] Kerr (2014), p. 53. This one might be a bit controversial, as there may be several ways of an emotion being fitting. For example, one might think that an emotion fits if it corresponds to the appropriate or correct situation. However, this might be an unhelpful definition as it leaves too many questions about what qualifies as appropriate or correct. Seeing fittingness through paradigm cases of emotion does not presuppose a substantive view on appropriateness for emotions.

[137] Kerr (2014), p. 46.

Prior's relief after exam marking is rational as he just handed the marked exams to the office and concludes that marking now is over.

So even if Prior's relief is nothing cognitive or propositional, we can still assess it and make it subject to rational critique: is it beneficial for Prior to feel relief? Is his relief warranted, or is it misplaced, as the marking is not really over? Is his relief fitting to a scenario that is similar to last semester, or does he unfittingly feel it in the middle of marking?

With these forms of assessment, Sullivan would probably say that relief in these cases are connected to beliefs that make the relief rationally assessible. For example, we need a connection to our beliefs about our ends for an emotion to be beneficial, we need a belief about similarly relevant paradigm scenarios to judge whether relief is fitting, and we need beliefs about evidence we hold to judge whether belief is warranted.

However, it seems odd to say that we should assess the affective state of relief, grief, nostalgia and other tensed emotions without their connections to beliefs, as this would likely render tensed emotions incomprehensible.

For example, if Prior feels dread about incoming exams, the emotion of dread only makes sense in combination of Prior's belief about the temporal location of the exams, as well as a directed "target object" of the emotion. Taking it this way would reduce dread to a feeling that would not accurately capture its nature as a tensed emotion anymore. Hence, assessing tensed emotions without their associated beliefs is not possible.

Note that saying this does not commit us to the cognitivist or propositional attitude camp of emotions, as this camp holds that emotions themselves are propositional mental states. We are here only pointing out that separating the belief-components from tensed emotions make them incomprehensible, the emotion itself can still be a mere affect. It's just hard or impossible to talk about a complex emotion like grief without the associated beliefs about losing someone or something.

But if we assess relief, nostalgia or dread without separating their affective states from their belief components, there is no reason to think that they are non-rational, as they can clearly be assessed through all three ways of examining the rationality of emotions. Tensed emotions can be fitting, warranted, and beneficial, and are therefore not non-rational, and there is no reason for us to think that tensed emotions cause our deliberation process to get off-track from rationality.

So, where does Greene's and Sullivan's scepticism about emotions come from? One source of their anxiety might be the thought that, if you accept an affect-based theory of emotions, emotions are something we cannot direct, control, or influence. Emotions are just something that overcomes us, that isn't part of our agency.

However, if that's the source of their scepticism, Greene's and Sullivan's worries can be calmed: we can and do direct, influence and control our emotions quite regularly.[138] We can regulate

[138] Kerr (2014), pp. 95-96.

whether we have certain emotions, when and for how long we have an emotion, or how to express my emotions. I can direct my emotions into different directions, I can decide to expose myself to situations that will trigger certain emotional responses, and I can focus my attention to certain thoughts that will emphasise or weaken my emotions. Of course, Greene and Sullivan would say that regulating our emotions is part of our rational reflection, not the emotions themselves – but it is certainly not true that we do not have any control about them.

Rational Reflection Without Emotions

However, even if we grant that tensed emotions are non-rational and that we need to separately assess them with rational reflection to determine their rational status, the problem mentioned earlier remains: what is it that makes the process of rational reflection immune from emotional and/or evolutionary influence?

As Street outlines, rational reflection always assesses our evaluative judgement in light of other evaluative judgements we hold.[139] If I rationally assess a belief or a judgement of mine, I am not evaluating its merit on its own, but have to start examining my belief in light of other beliefs I hold that might provide reasons for or against it.

But if my beliefs are contaminated by irrelevant influences, regardless of whether we're talking affective states or evolutionary forces, rational reflection is contaminated as well –

[139] Street (2006), p. 124.

even if we grant the purity of the process of rational reflection as such, the process still examines our beliefs and judgements in light of our other beliefs, which are contaminated by irrelevant influences.[140]

To tie this back to time-biases and tensed emotions: if Sullivan argues that our emotions are misleading us about the rationality of time-biases, the process of freeing us from their undue influence is contaminated by emotions just as much. If it is my tensed emotions that wrongly push me towards thinking time-biases are rational, why should the other beliefs that are supported by rational deliberation not be influenced by emotions as well, as Sullivan admits that there is non-rational emotional influence on my belief system?

For example, even if we admit that rational reflection process as such is pure, we still evaluate beliefs about time-biases by examining them in reference towards other beliefs that could be influenced by emotions: Sullivan's inclination towards temporal neutrality could be distorted by an affective desire towards being at ease, an emotion of tranquillity. But if we treat emotions as an irrelevant influence, we should not allow this either – rational

---

[140] Another way of putting this is with Kahneman (2011): We have two ways of thinking, one quick, affective (system 1) to respond quickly and effectively to arising situations and one slow, deliberative (system 2) that evaluates situations more carefully. However, system 2 is mostly not used, and if it is used, it mostly relies on base information from system 1, so that a pure system 2 response is not possible.

reflection just does not provide a clean way of assessing our beliefs and judgements free of emotional influence.

So, in summary, Greene and Sullivan's arguments about emotions, even if we accept an affect-based view on what emotions are (which is highly controversial and probably inadequate for complex tensed emotions like grief or nostalgia), we will still be able to rationally assess tensed emotions as fitting, warranted, or beneficial. And even if all that I have said about emotions is wrong, and they are really a non-rational distortion, Sullivan still owes us an answer why they would not contaminate rational reflection as well.

4.5 Thank Goodness that the Debunking Debate is Over

Let's draw an intermediate conclusion: Debunking fails. To recap, Greene and Sullivan are trying to block the following way of thinking:

a) There is a systematic argument that time-biases are irrational.
b) My intuitions tell me very strongly that time-biases are rational.
c) If my intuitions tell me very strongly that X is rational, I can take them as evidence against systematic arguments that X is irrational.
d) Therefore, I have evidence against a systematic argument that time-biases are irrational.

Their arguments try to block c) and establish that their systematic arguments against time-biases should be taken as decisive evidence for the irrationality of time-biases. For this, they need to debunk the intuitions leading us to believe that time-biases are rational. Unfortunately, as I have argued, an evolutionary debunking attempt against time-biases is set to fail. At the same time, emotions do not establish a successful debunking of time-biases either. So, we can still use our intuitions as evidence for the rationality of time-biases.

In the rest of the chapter, I would like to go further and explore whether appeals to evolution or emotions can actually provide justification for the rationality of time-biases. The general idea is this: firstly, if it's evolutionary advantageous for us to have developed time-biases, doesn't that at least sometimes make it rational to be time-biased? And secondly, if time-biases are grounded in our emotional life, doesn't that make time-biases a non-arbitrary, helpful preference pattern?

4.6 Might Evolution Actually Help Future-Bias?

The basic idea is that evolutionary advantage not only not undermines time-biases, but actually makes it appropriate for us to be time-biased. To sketch out the idea, let's take a step back and examine the so-called "Thank Goodness that's over" argument.

Evolution-Responses to Prior

To recap: Tensed attitudes are emotional responses that make references to time's passage – or in other words, a tensed attitude

can only be understood within a reference framework of past, present and future. Recall Prior's exam case from the beginning, and imagine that there is no such thing as events passing from future to present and then to the past: Without time actually passing, the future event would not "come towards you", so why would you dread the "incoming" exams? And without exams being past, over and done with, what are you relieved about?

This is a puzzle posed by Arthur Prior to defenders of the B-theory, or the tenseless theory of time. The B-theory states that there is no objective moment that is "now", and that the distinction between past, present and future does not describe objective reality. For the b-theorist, there are *no tensed properties* in the world. In other words, a b-theorist does not believe that things are genuinely past, present or future – things are just earlier or later than other things. The opposing position is the A-theory, according to which there are tensed properties – so-called a-properties – in the world. Some events have the property of future-ness, some the property of now-ness, some are past. Also, this property shifts – things lose future-ness and become now, and then become past. This is called the passage of time, to which an a-theorist is committed, and which a b-theorist denies.

While there was a lot of debate about which facts are referred to under the B-theory, there is a second, more hidden challenge:

According to Maclaurin and Dyke, the b-theorist must answer the following Prior-style question[141]:

> If there's no tensed facts we are thankful for, why are tensed emotional responses such as dread and relief appropriate?

What does it mean for a tensed emotional response to be appropriate if there are no tensed facts to be thankful for?

This challenge was recently made by Cockburn on behalf of Prior, and essentially asks the b-theorist to explain why we assign different significance to different times – and should the b-theorist not be able to do so, they must ask us to radically revise our tensed attitudes to events.

If the B-theory is true, then past, present and future events are equally real and existent. This, according to Cockburn, commits the b-theorist to the equal significance of past, present and future events – and if this is the case, there is no explanation for our emotional responses being sensitive to different times.

> Imagine that you're marking exams right now. It is painful, and you'd like it to stop.

Cockburn suggests that the fact that the pain happening right now gives it a special status – it gives us reasons to act, to avoid the pain, etc. However, if the B-theory is true, "right now" is not a property pain can have. "Marking exams right now" just means that the utterance happens at the same time as the marking. Just

---

[141] Maclaurin and Dyke (2002), p. 278.

as "Marking exams right here" does not refer to any property the pain can have, and therefore does not assign a special status to the suffering, "right now" can also not assign a special status to suffering.

Neither Maclaurin and Dyke nor I are sure what Cockburn exactly means by "special status", and his examples to support the argument. I think it's easier if we just cut all that "special status" talk - the upshot seems to be this:

Temporal Chauvinism[142]

- An emotional response is appropriate if it corresponds to a property (of an event, object or state of affairs) that warrants that response.
- If the property is tensed, then a tensed emotional response is appropriate.
- If the property is not tensed, then a tensed emotional response is not appropriate.

In short, it is tensed properties that make tensed emotional responses appropriate. E.g. the "pastness" of the marking of exams is what warrants relief.

If temporal chauvinism is correct, the b-theorist cannot answer the second challenge and is forced into a radical revisionist position – why grief about your loved ones if their past existence is just as real as their present non-existence?

---

[142] Term by Maclaurin and Dyke (2002), p. 282.

Here's where Maclaurin's and Dyke's response comes in: Tensed emotional responses can and do appropriately correspond to tenseless facts – we can explain tensed emotions with tenseless facts about our evolutionary history. Maclaurin and Dyke argue that our tensed emotions are either adapted for evolutionary advantage themselves, or plausible consequences of other behaviour or capacities that are adapted.[143] The authors engage in what they call "Darwinian Reverse Engineering", which seems to consist in telling an evolutionary story with a "high degree of plausibility", but without providing empirical evidence for it.[144]

Here's the evolutionary story for why we feel differently about past and future, which is similar to the one Sullivan tells for her debunking story: We have evolved to dread future pains and feel relief about past pains to track a control-asymmetry. We can sometimes control the future, but never the past, so it is advantageous to adapt the emotional responses in that way. If we dread future pains, we focus our actions on things we can affect and change, and if we feel relief about pain being past, we don't cry over spilled milk.[145] Also, Maclaurin and Dyke suggest that relief and grief constitute learning processes about what we ought to avoid, and what to value.

Hence, the feeling a tensed emotion is warranted by a tenseless fact – by our adaption to a control-asymmetry, or by adaption of a learning process, or even by being a mere side effect of other

---

[143] Maclaurin and Dyke (2002), p. 283.

[144] Maclaurin and Dyke, p. 284.

[145] Maclaurin and Dyke, p. 285.

capacities we evolved to have. Even though it might not always be the best response, there is nothing inappropriate or inexplicable about a tensed emotion corresponding to a tenseless fact. Therefore, temporal chauvinism is false, and tensed emotions can be appropriate even if there are no tensed facts.

Note that, for Maclaurin and Dyke, facts about evolution can make tensed emotional responses appropriate – which seems in stark contrast to Greene and Sullivan's claim that the evolutionary story behind tensed emotions should cause scepticism about using them as evidence for rational permissibility. If evolutionary facts can make tensed emotional responses appropriate, why can't we use tensed emotions to justify our preferences or behaviour because of their evolutionary origin?

Evolution as Justification?

So, inspired from Maclaurin and Dyke, we could attempt to build an evolutionary justification for time-biases: Time-biases are appropriate attitude patterns because they are based on facts about our evolutionary fitness. As tensed emotions are appropriate because they are based on facts about evolution, time-biases are appropriate based on their evolutionary advantage to us. The argument could roughly look like this:

(1) An attitude pattern is rational if it tends to make my life go well.
(2) If an attitude is tracking survival, then it tends to make my life go well.
(3) Time-biases track survival.

(4) Therefore, time-biases are rational.

While there might be a legitimate amount of overlap between evolutionary fitness and what makes my life go well, so that a debunking argument can be blocked, this is obviously a very questionable argument and leads me to think that evolution doesn't give us reason either for or against time-biases. Firstly, note that (3) is probably false in at least some cases when it comes to near-bias: while prioritising close future events over the far future might have been evolutionarily advantageous as an uncertainty tracker, nowadays we have much better means to do so, and near-bias very often does not make my life go well. And (2) seems to be questionable as well: not everything that is evolutionarily advantageous might be good for me – there's a need for a further argument why everything contributing towards my survival is also good for me.

This argument might be more plausible for future-bias, as tracking the direction of causation and asymmetry of control might be advantageous to us: why cry over spilled milk when you can focus on things you can actually change? However, a problem appears: evolution doesn't actually seem to do the work here. You can argue in favour of future-bias based on the control-asymmetry without using evolution at all:

Control-Argument for Future-bias

(1) Caring about things we can change and not caring about things we cannot change is good for us.
(2) CONTROL-ASYMMETRY: We cannot change the past but can change the future.

(3) Future-bias is an attitude pattern sensitive to control asymmetry.

(4) Future-bias is good for us.

So, evolution is not actually needed to make this argument – where future-bias comes from seems to be irrelevant in the question whether it is good for us or not. The argument from control-asymmetry might provide a reason for thinking that future-bias is rationally justified, but it does so without the need to outline evolutionary influence. Hence, we shouldn't use evolution to seek justification for time-biases – whether time-biases are justified or not seems to be independent from arguments based on evolution.

Should we accept the control argument? I believe so. A recent study has found that future-bias is indeed sensitive to ability to control what happens and what doesn't: if participants could affect the past, they would be less future-bias than they would be if they couldn't.[146] This supports the claim that future-bias isn't "evolutionarily hard-wired" and that the preference pattern is instead reasons-responsive.

A possible objection here could be that, contrary to popular wisdom, we can actually affect the past and (2) is false. Dorsey, for example, argus that we can improve the well-being of our past selves by fulfilling achievements for them – even if these do not affect our current or future well-being, completing projects our

[146] Latham et al. (2020). The study also found, however, that people remained future-biased to a degree even if they could affect the past, leaving conceptual space for further investigation.

past selves held dear will be an achievement for them, thereby adding to their well-being.[147]

However, even if we grant that this is possible (which is not a given and subject to controversy), the more plausible explanation for this phenomenon would be that fulfilling projects adds not to the well-being of my past self, but to my temporally extended well-being, such as my life-time well-being. The reason why we think that fulfilling projects is good is usually not tied down to momentary well-being but attributed to well-being of the temporally extended kind.[148] So, we don't actually benefit our past selves, but our temporally extended well-being. Dorsey replies to this that goods are always to some extend "temporally localised", and that even a-temporal value has some tie to a moment in time.[149] This is obviously true, but that rejoinder seems to misread the objection: what is questioned isn't whether the added value is temporally localised, but whether the beneficiary of that value is our past self – and it is more plausible that fulfilling a project adds an achievement to us as a temporally extended person, rather than our past self at a given time.

To sum up: evolution does not play a role in justifying future-bias. Instead, we should treat future-bias as a preference pattern fitting our ability to control and influence present and future events, and our inability to affect the past. This is both intuitively

---

[147] See Dorsey (2018), pp. 1906-1908.

[148] See for example, Hurka (1996).

[149] Dorsey (2018), p. 1908.

plausible, as well as backed up by empirical research. Hence, future-bias is a rational preference pattern.

4.7 Emotions as Rational Grounds for Justification

My aim is to provide rational grounds on which the permissibility to be time-biased could be build. What this means is to find a reason that speaks in favour of being time-biased. I will explore two ways of doing so, one that works for both schools of emotions, and one that will only work if you think that emotions are a kind of value judgement.

The first argument will be via dynamic prudence:[150] an agent's emotional patterns are prudent to her if it furthers her future well-being. If that is the case, the emotional patterns provide a rational basis for choice as a means to uphold her prudential emotional pattern. The general idea is that patterns of tensed emotions tend to contribute to an agent's well-being, and therefore provide rational grounds to choose attitudes and options that uphold that pattern. So, if a person's time-biases contribute towards a pattern of tensed emotions, that person has a rational basis for being time-biased.

The second argument will outline that our tensed emotions are value-judgements. According to this view, emotions are statements about the world that indicate not only how we perceive things, but also how much we value something, and whether something is harmful or beneficial to us. These value

[150] See Kerr (2014), pp. 84-88.

judgements sometimes are more fundamental than knowledge. In this way, tensed emotions tell us something about what is valuable to us and what harms us. And if time-biases are based on tensed emotions, tensed emotions can be used as evidence for why time-biases are rationally justified.

1 Dynamic Prudence

Dynamic Imprudence is a way Alison Kerr tries to capture a specific form of emotions going wrong that goes beyond the emotion being unfitting, unwarranted, or harmful in a certain particular instance. Kerr looks at patterns of emotions that either promote or diminish an agent's well-being, practical endeavours or goals. If a pattern of emotional responses diminishes an agent's well-being, and the agent does not take steps to regulate her emotions, given the evidence she has for the pattern to diminish her well-being and practical endeavours, the agent is imprudent. In Kerr's words, "*Imprudence occurs when an agent's emotions are getting in the way of her relevant interests or harming her well-being, she has information about this fact, but still fails to take steps to regulate her emotions properly.*"[151]

For my purpose, I will look at the reverse: when is an agent dynamically prudent when it comes to emotions? Given Kerr's account of imprudence, we can suggest a similar account for dynamic emotional prudence:

> A is prudent in respect to A's pattern of emotions if A possesses evidence that her pattern of emotions further her

[151] Kerr (2014), p. 91.

practical endeavours or well-being and succeeds in upholding her pattern of emotions.

Take anger as an example: my anger about oppressive behaviour towards me in particular instances and situations might be warranted, but not beneficial to my interests and well-being. I might be outraged at a type of oppressive behaviour directed at me, which is warranted because it's the right kind of response to a situation, but it will be "counter-productive", in Amia Srinivasan's words, to my goals and aims, e.g. ending the oppressive behaviours I am angry about. In other words, the rationality of my anger comes into conflict: rational in terms of warrant, irrational in terms of benefit.

Srinivasan suggests that we are looking at two different, possibly incommensurable goods here: one is the epistemic value of getting the right response to a situation, or to appreciate the situation in the correct way, while the other one is the good of reaching and fulfilling my goals.[152] Both are valuable in their own ways.

However, if we look at it from a perspective of dynamic prudence, we might be able to resolve the conflict: my anger in a certain situation might not be beneficial to me in that instance as a token. But my anger as a pattern across different scenarios where I am consistently being subject to the same kind of oppressive behaviour might further my goals, aims and well-being, by motivating me to act against oppressive structures, by making

[152] Srinivasan (2018), p. 19.

me able to recognize and respond better to oppression, and by building resilience against oppressive behaviour.

In this way, my anger can be rational, even if anger in a particular instance might be counter-productive. And if I act in ways that upholds my anger, e.g. regulating it in a way that makes it less destructive and more sustained, more directed and focused, because I have evidence that my anger pattern is furthering my ends, I act in a prudential manner. And actions, choices and attitudes I affirm in order to sustain my anger have a rational basis because they are prudent with respect to my pattern of anger.

We can say the same about tensed emotions and time-biases. Tensed emotions, e.g. like grief after a loss may sometimes not feel beneficial in a particular instance (except relief of course). Your grief might be warranted in response to losing someone you love, but it might render you unable to deal with your daily tasks. However, seen over a pattern of emotional responses, your grief may contribute to your well-being in the sense of helping you to coming to terms with your loss, honouring your commitments to the person you lost, and taking time for yourself for remembrance. In that sense, grieving may contribute to your well-being, and you as an agent nurturing your grief and taking steps to feel it actively, are acting prudentially.

The same goes for relief, which is less complicated than grief: A pattern of relief after bad events passing might be beneficial to you in the sense that it helps your practical endeavours of focusing on events that are ahead of you, which you can influence

and control, and to not hold yourself up too long with terrible things like exam marking that have passed you. In other words, patterns of relief help you to appreciate the control-asymmetry we have talked about earlier, and steps to sustain your emotional pattern are prudent.

Future-bias is an attitude that uphold and sustain your patterns of tensed emotions. If you are future-biased, you will sustain your relief after something bad has passed. If you are future-biased, you will prefer your loved ones to be future rather than past, and good times to be ahead of you rather than behind you. In other words, when we respond to events in a future-biased way, we are contributing to patterns of tensed emotions that are contributing to our goals and well-being. In that way, being future-bias is prudent, and has a rational basis in tensed emotions.

I believe this line of argument is more likely to succeed for future-bias than near-bias, as near-bias is not necessarily based on the same tensed emotions as future-bias, and it might be less obvious whether it will be prudent or imprudent to be near-biased. For my purposes, vindicating the prudence of future-bias is sufficient to provide us with a reason not to be temporally neutral.

2 Emotions as Value Judgements

According to the propositional attitude theory of emotions, emotions are mental states that say something about the world. If I feel an emotion, I express something about the world, and how I stand in relation to that. My anger describes a certain situation, and how I am dissatisfied with or feel harmed by that

situation. In other words, emotions can indicate what we value and what matters to us.

Annette Baier, in her essay *Feelings That Matter* cites an example of a young man on trial for stabbing his mother to death.[153] He does not remember doing it, and hence does not know whether he did it, but says that he has the "guilty sort of feeling", so he must have done something. In this admittedly extreme case, his emotion of guilt do not only indicate value, but his emotions are a value judgement about what he did that precedes him knowing that he did it.

Take a less extreme example: many people of colour are subject to micro-aggressions, brief and short, but common interactions that do not constitute a racist incident, but constantly communicate hostility and prejudice towards people of colour. People of colour sometimes cannot conceptualise the harm or wrong done to them, but nevertheless feel angry, exhausted and alienated from these interactions. In this case too, emotions form a value judgement that precedes knowledge about the harms and wrongs at play.

Often, this anger serves as evidence that some beliefs and attitudes are rationally justified. For example, a person of colour might be take her anger as evidence that racism exists in her workplace, even in absence of explicit racist incidents. Or, she might take her anger as evidence for her attitude to minimise interactions with certain people that trigger this anger. In this

[153] Baier (2004), p. 200.

case, we see that an emotion that indicates values and harms can justify beliefs and attitudes.

So, what does that tell us about tensed emotions? Grief, nostalgia and relief are all tensed emotions that are widely believed to express a value judgement too.[154] With grief, we indicate how valuable our loved one was to us, nostalgia expresses that something at the moment and in the future is lacking something past events had. Tensed emotions express values, and if time-biases are based on tensed emotions, we shouldn't be sceptical of using tensed emotions as evidence for the rational permissibility of future-bias.

4.8 Wrong Kind of Reason?

Proponents of temporal neutrality could of course object to both control-asymmetry as well as emotional prudence as rational grounds for future-bias. One objection could be that this is the wrong kind of reason: it's merely a pragmatic reason to sometimes be future-biased, but it doesn't show how future-bias as such has rational grounds. This links to the so-called arbitrariness-objection against future-bias: just because there are pragmatic gains, that doesn't make it less arbitrary for me to care about the future but not the past.

To illustrate, compare it with a situation where we you know that, rationally, you ought to be temporally neutral. However, I offer you 10€ if you're future-biased. Would that be a proper reason for future-bias being rational in your case? Or would that

[154] E.g. Baier (2004), p. 206.

be a *wrong kind of reason*, in the sense that this kind of "cheap" pragmatic concern does not actually show whether your preference patterns or attitudes are good or bad?

Compare this with the control-asymmetry first: does future-bias tracking what we can and cannot control amount to a reason of the wrong kind, similar to the 10€? In the sense that focusing on what I can change and not on what I cannot change, and 10€ would both be "good for me", they are similar. However, the "good for me" differs, as with the control-asymmetry. There is an inherent link between what future-bias tracks and how it provides a reason for me to be future-biased, while the 10€ have no bearings on whether future-bias is actually a good thing or not. In other words, it's both appropriate and beneficial for me to be future-biased based on the control-asymmetry, but future-bias only happens to be beneficial to me to get 10€.

With emotional prudence, I would argue similarly: yes, it is a kind of pragmatic gain, leaving you better off, but emotional prudence does say something about whether future-bias as such is good, not only providing a pragmatic gain that does not say anything about the attitude in question: in the case of emotional prudence, future-bias is beneficial to me in virtue of the attitude pattern being good, in contrast to future-bias being good because I happen to get 10€ out of it. So, the control-asymmetry and emotional prudence as justification do not constitute a wrong kind of reason for future-bias.

4.9 Conclusion

So where are we now? I have explored two ways emotions can serve as rational grounds for future-bias: future-bias is rationally prudent as it contributes to a pattern of tensed emotions that furthers our well-being, and tensed emotions can be used as evidence for future-bias's permissibility, as tensed emotions are value judgements indicating what matters to us.

What about the control-asymmetry? It seems that an attitude-pattern that tracks the control-asymmetry will make us focus on things we can change, and discount things we cannot change, leading us to better decision-making. This might be especially useful in avoiding so-called sunk-cost fallacies, where we tend to stick to choices we've already invested in, even if the investment is lost either way, as it is past, and we don't want the choice anymore. This is of course limited to providing a rationale for future-bias, not for near-bias.

Recall the three argument-types against time-biases:

a) There are no reasons in favour of being time-biased
b) There are reasons not to be time-biased
c) There is a debunking explanation that applies to time-biases.

If emotional grounding and the control-asymmetry successfully provide us reasons to be future-biased, then we have not only deflected c) in terms of debunking, but also have disproven a): we have reason to be future-biased, as it is better for us to care about the future and not to care about the past.

So, not only could we deflect scepticism about emotions being a intrusive influence on our rationality, emotions can serve as a basis for us to justify our beliefs and preferences. So, the next time you feel relief after marking a mountain of exams, maybe that says something about what you take to be good or bad.

5 What's Better for You: Future-Bias or Temporal Neutrality?

Suppose you are a member of the generation of so-called "millennials". As a "millennial", you have a preference for owning a house at some point in the future. To buy one, you need to save a lot of money, since housing prices in your area aren't exactly cheap. So, you start saving. However, you also have a preference for delicious avocado toast – and even though your preference for avocado toast is a lot weaker than your preference for owning a house, avocado toasts can be bought in the very near future, while your future house is far, far away. So, even though you want to own a house more than eating avocado toast, you keep buying avocado toast – which in the end makes you unable to save enough money for a house.[155] Irrational millennial, you.

In other words, you and other millennials discount the objects of far-future preference relative to those preferred in the near-future. As a result, your life goes less well than it could have. Therefore, you should be temporally neutral, making smaller sacrifices now when they'll be compensated by greater gains later, so that your

[155] Credit to this argument goes to Real Estate Mogul and Avocado-Expert Tim Gurner: https://www.theguardian.com/lifeandstyle/2017/may/15/australian-millionaire-millennials-avocado-toast-house

whole life goes as well as possible. This is known as the compensation-argument for temporal neutrality[156].

More precisely, this argument challenges the rationality of

> Near-Bias: An agent is near-biased iff for two exclusive future events E_1 and E_2,
>
> - where E_2 is at least as positive as E_1, the agent prefers E_1 because it would occur earlier to now than E_2, or
> - where E_2 is at least as negative as E_1, the agent prefers E_2 because it would occur later to now than E_1.

and supports the rationality of

> Future Neutrality: An agent is future-neutral iff for two exclusive future events E_1 and E_2, with E_2 being later and at least equally good as E_1, the agent is indifferent between E_1 and E_2.

Let's grant that the compensation-argument succeeds in showing near-bias to be irrational, and that we should be future-neutral.

Recently, Sullivan and Dougherty have attempted to extend the reach of the compensation argument to also challenge the rationality of future-bias.[157] They argue that

---

[156] At least by Brink (2010) and Sullivan (2018).

[157] Sullivan (2018), Dougherty (2011).

Future-bias: An agent is future-biased iff for two exclusive events E_1 and E_2, with E_1 being in the past,

- where E_1 is at least as positive as E_2, the agent prefers E_2 to E_1 because E_1 is in the past and E_2 is not, or
- where E_1 is at least as negative as E_2, the agent prefers E_1 to E_2 because E_1 is in the past and E_2 is not.

is not rationally permitted and that we are rationally required to be past-future-neutral:

Figure 1: Andersen (2016)

Past-Future Neutrality: An agent is temporally neutral iff for two exclusive events E_1 and E_2, with E_1 being in the past and equally good or bad as E_2, the agent is indifferent between E_1 and E_2.

Their arguments aim to show that, if we are future-biased, our lives will be worse over all – even when it comes to the past, it "pays" to be temporally neutral. Therefore, we should be entirely temporally neutral not just with respect to future events.

This chapter will argue that Sullivan's and Dougherty's creative arguments are unconvincing. To show that the irrationality of future-bias is not shown by either argument, I will proceed in three steps. Firstly, I will show that Sullivan's and Dougherty's argument rely on diachronic norms governing over a pattern of choices over time, and do not criticise individual choices at a

time. However, if you are sceptical of diachronic norms, because you think it's unfair to be criticised for something you don't actually choose, e.g. ending up in a diachronic tragedy, Sullivan's and Dougherty's arguments fail. Secondly, I will argue that, even if there are diachronic norms of rationality, the arguments by Sullivan and Dougherty still fail to establish future-bias's irrationality, as they do not reveal any actual inconsistency, but merely show the exploitability of two attitudes held together. As this is dangerously overgeneralising, we should reject their arguments. Thirdly, even if that is not the case, Sullivan's and Dougherty's arguments only establish that in cases of pragmatic loss, future-bias is irrational, but do not show the general irrationality of future-bias. As an illustration, their argument can backfire, as one can easily imaginea temporally neutral person making decisions ending in diachronic tragedy: a person who is temporally neutral will be subject to sunk cost fallacies, which are much more commonly accepted as failures of rationality and occur far more often than pragmatic loss based on future-bias. While this does not establish the general irrationality of temporal neutrality, it suffices to show that future-bias is almost always rationally permissible, and in a lot of cases, better for you than temporal neutrality.

You're still not allowed to be near-biased though, especially if you're a millennial like me.

5.1 The Compensation-Argument Against Near-bias

Let's start with the Compensation-Argument against near-bias in more detail first. Sullivan states the argument as following: [158]

Compensation-Argument against Near-Bias

(1) A rational agent prefers her life to go forward as well as possible.[159]

(2) If you are near-biased, you will choose earlier lesser goods over later greater goods just because of them being earlier.

(3) Your life would go better if you chose the later, greater good over the lesser, nearer good.

(4) Therefore, if you're near-biased, your life will not go forward as well as possible.

(5) Hence, a rational agent would not be near-biased.

This argument is not strictly valid, as there is the possibility of an agent always encountering better, earlier goods, so that a near-biased person would choose the exact same goods as a temporally neutral person. In this case, the near-biased person's life would not go worse than the temporally neutral person's one. But as that's quite unlikely, let's just ignore that.

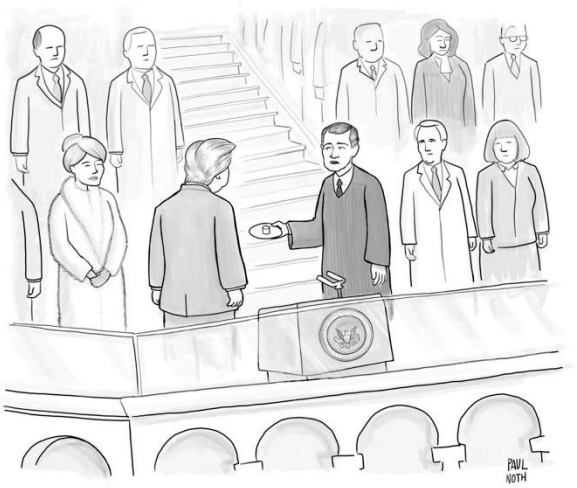
---

[158] Sullivan (2018), p. 11 calls it Life-Saving Argument.

[159] "Life going forward" is borrowed from Sullivan herself, as the so-called success-principle. If you feel that this stacks the deck against temporal neutrality, think of it as

(1) A rational agent prefers her life to be as good as possible.

Notice that the argument does not imply that discounting future goods based on uncertainty is irrational, it's only pure temporal discounting (discounting based on temporal distance as such) that is criticised. Notice also that the argument does not presuppose a specific account of goodness, or what makes a life go well. Whether you believe goodness is pleasure, desire-satisfaction, achievement, friendship, or a combination of those, you can accept the argument.[160]



"You can eat the one marshmallow right now, or, if you wait fifteen minutes, I'll give you two marshmallows and swear you in as President of the United States."

Figure 2: Noth (2017)

Also, according to premise (1) of the argument being rational consists in something more substantive than just being coherent. You might be an *ideally coherent eccentric* who prefers their life not to go best.[161] Let's grant the substantive conception of rationality presupposed in (1).

It's clear why near-bias would lead you to (3) and to a worse life overall. Imagine that you always pick earlier avocado toasts instead of later house-buying. You will end up with a life much worse than if you had picked later house-buying over earlier avocado toasts. Plausibly, on any conception of the good life that one would want to subscribe to, an abundance of avocado toasts now scores lower than owning property later. As a result, the

[160] Sullivan (2018), p. 12.

[161] See Sullivan (2018), pp. 12-13, but also Street (2009), p. 5 for a full discussion.

compensation-argument successfully shows near-bias to be irrational. Let's now turn to future-bias, which might be more difficult to argue against.

5.2 Pain-Pumps and Cookies Against Future-Bias

Things are more complicated with future-bias. First, it's not obvious how future-bias would lead you to pick worse options over better ones, or lead to a worse life overall. After all, future-bias is not obviously connected to making decisions between events, but rather with having preferences or attitudes about them. For example, if you wake up in hospital with no memories, you could prefer that you have already been operated yesterday over having a shorter, less painful operation tomorrow – here, there's no decision for you to actually make, you only hold attitudes with no obvious link to actions, as you can't choose the past operation.[162] However, Dougherty and Sullivan have suggested ways of connecting future-bias to decision-guiding that leaves you worse off. Let's start with Dougherty:

Pain-Pumps against Future-Bias

Dougherty argues that an agent who is both future-biased and risk-averse will make decisions that leave her worse off overall. Dougherty amends Parfit's operation case to show this.

[162] See Parfit (1984), p. 186 for the classic Past and Future Operations Case.

> Operations 2 On Monday, you are admitted to a hospital. You are told that you will have one of two courses of operations, but you are not told which. If you get the earlier course, you'll have a painful *four-hour* operation on Tuesday and a painful *one-hour* operation on Thursday. If you have the late course, you'll have just one painful *three-hour* operation on Thursday. You have a fifty-fifty chance of getting each course. After any operation, you'll have amnesia, so that you won't remember any operations you had recently. There's a calendar next to your bed so that you always know what day it is.[163]

On Monday, you'd prefer the later course, since you'd suffer through two hours less. Now, assume that you are *risk-averse*. This implies that if you are faced with a gamble between a better and a worse option, you'd like the risk of getting the worse option to be lower, other things being equal. So, you'd actually be willing to lower the overall expected value of the gamble to reduce the risk of getting the worse option.[164] Thus on Monday, you'd like the risk of going through the operation course with the most pain (the early course) to be lower. Imagine now that I offer you help in form of a pill:

> Early Help If you have the early course, then the pill will reduce the time of pain you'd experience on Thursday by

[163] Dougherty (2011), pp. 526-528.

[164] Dougherty (2011), p. 525.

> 29 minutes. If you have the late course, the pill will increase the duration of pain on Thursday by 31 minutes.

If you are risk-averse, you'd accept the pill, because even if your overall expected pain is increased by one minute, it decreases the risk between the two options by reducing the gap of 5 hours of pain for the first course versus 3 hours for the second to 4:31 for the first course versus 3:30 for the second.

You wake up on Wednesday. By assumption, you don't know whether you had an operation yesterday. Also, you are future-biased, so you're facing another gamble – now on Wednesday, you'd prefer to be on the earlier course, since the first operation would already be *over*, and you would have only a very short operation ahead of you instead of a long one. Since you are risk-averse, you'd like to make the very short operation longer and the long operation shorter to reduce the gamble. What luck! I've got a second pill for you:

> Late Help If you have the early course, the pill increases the time of pain on Thursday by 30 minutes. If you have the late course, the pill decreases the duration of pain by 30 minutes.

Since you are risk-averse and future-biased, you would accept the second pill too, since it would reduce the gap between the two outcomes you face.[165]

[165] Dougherty (2011), p. 528.

	Effect of Early Help	Effect of Late Help	Overall Effect
Early Course	-29 min of pain	30 min	1 min
Late Course	31 min	-30 min	1 min

Regardless of which course you undergo; you will experience one additional minute of pain. Since you are risk-averse and future-biased, you won't refuse the pills, leading you to an overall worse life. So, assuming risk-aversion is rationally permissible, future-bias should not be. Here's the argument spelled out:

Pain-Pumps against Future-Bias

(1) Risk-Aversion is rationally permissible.

(2) A rational agent prefers her life to go forward as well as possible.

(3) If you are risk-averse and future-biased, you will choose lesser future goods over greater past goods just because they are in the past.

(4) Your life would go better if you chose the greater good over the lesser good.

(5) Therefore, a rational agent would not be future-biased.

With Dougherty's pain-pumps, we can put forward (2) and show that sometimes, future-bias in combination with risk-aversion leads to a worse life. Therefore, you ought to reject future-bias. Do we have to accept risk-aversion? Sullivan casts doubt on the assumption that risk-aversion works in that way when it comes

to pain.[166] But even if we don't accept risk-aversion for pain, Sullivan suggests another argument against future-bias.

Cookies against Future-Bias

Sullivan's argument is similar to Dougherty's. Instead of risk-aversion, she states that an agent who accepts both future-bias and regret-avoidance will behave irrationally. What is regret? Regret, so Sullivan, is a preference about your past behaviour: you regret something if you prefer that you had done otherwise.[167] Since at least sometimes, you can foresee that your future preferences will change over time, you will know that with certain options you pick now, you will face regret later on. If you can foresee that, it seems permissible to accept regret-avoidance.

> Weak No Regrets[168] If I have full information about the effects of the options available to me, then it is *permissible* for me to avoid the options I know I will regret choosing over the one that I won't regret choosing.

Sullivan calls it weak because it is sufficient for her purposes to only assume the permissibility of regret-avoidance, not the requirement of it, in contrast to authors like Bratman, who thinks you are *required* to avoid an option you know you will regret choosing.

---

[166] Greene and Sullivan (2015), pp. 955-956.

[167] Greene and Sullivan (2015), p. 957.

[168] From both Greene and Sullivan (2015), p. 958, as well as Sullivan (2018), p. 61 with different names.

> Cookies: Let's say that I offer you cookies. You can either have two cookies at once, or one cookie at some point in the future. My cookies are *very* delicious, so the answer seems clear – you should pick two cookies immediately. However, you are future-biased. Let's say you are absolutely future-biased – you don't assign any value to the past at all. Now, if you'd choose to have two cookies, you'd prefer to have the one cookie later, since the one cookie would still be in the future. After the time of the later cookie has passed, you will become indifferent to either choice, since it's in the past. Until that time, however, you can expect to regret your choice of two cookies, while you wouldn't have to face regret when choosing only one cookie. Therefore, you are permitted to choose one cookie over two cookies, which is irrational.[169]

If you indeed are absolutely future-biased, the case could even be more extreme – even if the choice would be between 10 cookies now or a crumble later, an agent with both future-bias and regret-avoidance would always have to choose the latter option, leading to less cookies in your life. This, however, is irrational. So the compensation-argument can get off the ground.[170]

Cookies against Future-bias

(1) Regret-Avoidance is rationally permissible.

---

[169] Greene and Sullivan (2015), p. 961.

[170] Greene and Sullivan (2015), p. 965, Sullivan (2018),p. 64.

(2) A rational agent prefers her life to go forward as well as possible.

(3) If you are regret-averse and future-biased, you will choose lesser future goods over greater past goods just because they are in the past.

(4) Your life would go better if you chose the greater good over the lesser good.

(5) Therefore, a rational agent would not be future-biased.

Should we accept regret-avoidance? As with risk-aversion, these principles seem to be pretty intuitive. So, should future-bias be rejected? This will depend on how Weak No Regrets will be spelled out, in terms of what regret means, what "full information about the effects" means, and what kind of permissibility we are operating on. Let's start with what regret means.

Dorsey has some doubts about Weak No Regrets and how we should interpret regret: He points out that regret, as used by by Greene and Sullivan, can mean various things.[171]

> Agent-Regret: I agent-regret an action if the action was irrational, wrong or otherwise displaying a salient normative failure.
>
> Preference-Regret: I preference-regret an action on the basis that the action led to a state of affair that I currently disprefer.

[171] Dorsey (2016), pp. 15-17.

Dorsey thinks that only agent-regret is plausible as an interpretation of Weak No Regrets, as preference-regret could violate temporal neutrality, but that I wouldn't feel agent-regret if I ate the two cookies while future-biased, as I am maximising cookies for my future-self while making the decision.

While I agree with Dorsey's argument, I feel that Greene and Sullivan can be defended here: you might still think that, even if you don't feel agent-regret, you're still allowed to avoid acting in ways that will later result in preference-regret, and it would still be incompatible with future-bias being rational, even if it's not compatible with temporal neutrality. However, there are more problems with Weak No Regrets.

For example, what does "full information about the effects" mean? Does it mean full information about all possible future outcomes and their effects, or full information about how much and why I will regret my options? If it's the former, then all bets and uncertainty is excluded, and Weak No Regrets seems plausible enough. But as we're never going to be omniscient, Weak No Regrets won't be helpful at all as a heuristic for rational decision-making. Also, if full information goes this far, it's likely that Sullivan's cookie case is going to be a lot less plausible: I will know at T1 (when I can choose two cookies) that after T2 (when I would've gotten the one cookie), I will become indifferent and not feel any regret anymore.

With the help of a reflection principle for preferences, we can turn the argument:

> Reflection for Preferences: If you know now that at a later time you will prefer an option or be indifferent to an option, and you know that you won't be in a worse epistemic or evaluative position at that time, then you should prefer that option or be indifferent to that option now.

This is a variation of reflection for beliefs:

> Reflection: $P_0\ (A|P_1(A)=r)=r$

If reflection for preferences is plausible, and I think under circumstances of full information it does look quite plausible, then it follows that I should be indifferent about the cookies now because I know I will be indifferent about my choice at T2. Hence, I wouldn't regret anything after T1, as I know that later on after T2, I will be indifferent. Of course, I should also be indifferent now, before T1. But Greene and Sullivan's argument risks grinding to a halt here, as I wouldn't automatically choose one cookie over two anymore. So, Weak No Regrets with full information about all effects and outcomes is not only a very restricted heuristic, but also threatens to cancel Greene's and Sullivan's own argument.

If "full information about the effects" does not mean omniscience, but only that I have full knowledge about which choices will cause how much regret and why, then this problem disappears, as Reflection will no longer yield indifference right now or after T1. Note that this might be a move that is attractive independently from the problem above, as Weak No Regrets is pitched as a useful rationality heuristic, and the principle entailing

omniscience would make it too idealised to be applicable to everyday situations. However, this opens Weak No Regret to other counterexamples:

> The Cable Guy:[172] The cable guy is coming to your flat, but you don't know exactly when: he said he will arrive tomorrow between 8:00 and 16:00, so you must wait all day. Alan Hájek offers to keep you company while you wait and suggests a bet to make things more interesting to you: you divide the waiting time into two four-hour intervals, and bet on whether the cable guy will arrive in the morning (between 8:00 and 12:00) or in the afternoon (12:00 and 16:00). The winner gets cookies.

So, what's the problem? At first, you think there's none, there's no reason to prefer one interval over the other, you can just bet. However, if you are regret-averse even on Weakest No Regret, this implies that you ought to bet on the afternoon. Let's say you are regret-averse and bet on the morning. When morning approaches, and it becomes 8:05 and 8:10, and the cable guy isn't there yet, you notice that the probability for your bet to win has fallen, and you come to regret your decision to bet on morning. The further the morning progresses without the cable guy arriving, the more regret you will feel. So, even if there is a 90% chance of the cable guy arriving in the morning, it seems permissible for you to bet on the afternoon just to avoid regret.

---

[172] Hájek (2005). The case was originally levelled against van Fraasen's Reflection principle. That it applies here shows that Weak No Regrets somewhere relies on a similar diachronic norm.

Hence, if full information does not entail knowledge about all outcomes, Alan Hájek will probably get the cookies.

So, either way, Greene and Sullivan will run into problems and need to spell out what they mean by "full information about the effects". However, Weak No Regrets, even if cleaned up, has more problems still. Consider

> Sophie's Choice: On the train to Auschwitz, Sophie is forced by the guards to choose between her two children. One will live, one will die. If she does not choose, Sophie and both children will all be killed at once.

Sophie is required to choose an option that she will profoundly, utterly regret. It is not permissible for her to not choose an option, even if not choosing will lead to no regret, as she will be dead. Even if you think it is permissible for her to choose her own death, it seems clearly impermissible to condemn both of her children to death to save herself from regret. Hence, Weak No Regrets is false.

You might think that this is unfair. Weak No Regrets is a rationality heuristic and is not designed to deal with moral dilemmas. Sophie's Choice seems too extreme. However, there are weaker counter-examples:

> Acquiring Moral Expertise: If I choose to read Peter Singer's Famine, Affluence and Morality for the first time, I will become aware of my moral duties towards the world's poorest but will regret reading it due to my guilt

and the weight of my moral obligations.[173] If I choose to reread Harry Potter instead, I won't regret anything.

Here again, it seems that, excluding other factors, I am not permitted to choose the option that would lead to no regrets over the option that will lead me to a lot of regret, assuming that I have reasons or a duty to acquire moral expertise. And in this case, we're not in a dilemma.

Hence, Weak No Regret just seems false. Even if they manage to avoid Dorsey's objections, it's fairly easy to come up with counterexamples to their principle, much easier than to think of counterexamples to the permissibility of future-bias. But let's pretend I'm wrong and that Weak No Regrets, or any variation of it, is not obviously false, just for the sake of the argument, and continue to explore the compensation-argument.

In the rest of the chapter, I will advance three central points against the arguments offered by Dougherty, Sullivan and Greene:

1. All three authors rely on diachronic norms to establish the irrationality of future-bias. Diachronic norms are a point of controversy. If you don't believe in them, or if you find it unfair to be criticised for ending up in a tragic sequence of decisions, even if every decision as such is rational, then the arguments for future-bias's irrationality fail.

---

[173] This is not made up: several first-year students told me explicitly that they regret learning about Singer's principle, not world poverty, which they were aware of before.

2. Even if there are diachronic norms that govern patterns of choices, it seems that Dougherty, Sullivan and Greene's arguments do not actually show an inconsistency in our attitudes, but merely reveal the exploitability of holding two attitudes together in certain scenarios. As pragmatic loss only shows irrationality if they uncover incoherence within our beliefs and attitudes, and mere criticism of exploitability risks overgeneralising to just any combination of attitudes, the arguments should be rejected.
3. Even if the first two points don't stand, and the authors succeed in their arguments, they do not establish the general impermissibility of future-bias, but only the impermissibility of future-bias in a limited set of cases. At the same time, future-bias is not only permissible, but temporal neutrality might be irrational in a whole range of cases where we are at risk of committing sunk cost fallacies.

5.3 Are Tragedies Irrational?

Let's go back to Dougherty's argument first: if you are risk-averse and future-biased, you will behave in ways that will leave you worse off overall. His example was an operations-case where you took two pills, each of them warranted by your risk-aversion, but *put together in a sequence*, they will leave you worse off. At any given time, you would prefer not to take both pills together because they'd increase your pain. But at the times where you're given the choice, you prefer to take each pill – leading you to a

diachronic tragedy.[174] A diachronic tragedy is a sequence of actions you don't want to perform, but end up performing anyways because your preferences and attitudes at that time lead you to perform each individual action of the sequence.[175]

(1) At any given time, you don't want to take both pills.

(2) On Monday, you want to take Early-Help.

(3) On Wednesday, you want to take Late-Help.

(4) Although you don't want to, you take both pills.

This represents a diachronic tragedy or a diachronic Dutch book: at t1 I have the option UP or DOWN, and at t2 I have the option

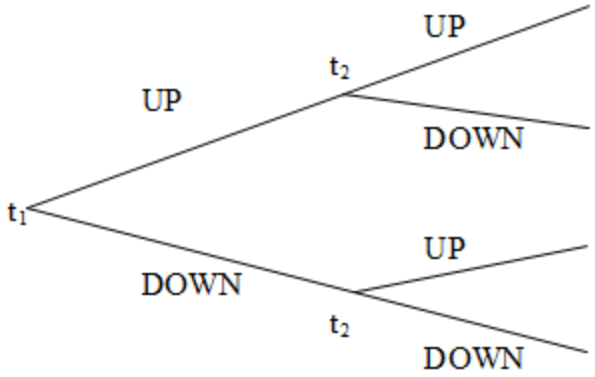


Figure 3: Hedden (2015: 429)

UP or DOWN again. At all times, I prefer (DOWN, DOWN) over (UP, UP), but at t1 I prefer UP over DOWN and at t2 I also prefer UP over DOWN. Therefore, I end up with (UP, UP).[176]

Intuitively, we want to avoid behaviour like this. But should we be criticised for ending up in a tragedy where we wouldn't choose

[174] Hedden (2015), p. 5.

[175] Hedden (2015), p. 2.

[176] Hedden (2015), p. 429.

otherwise? Note again, that this structure of argument is not needed when it comes to near-bias. What is the additional component needed in this argument to challenge future-bias?

To avoid diachronic tragedies such as Dougherty's pain-pumps, it has been suggested that we ought to follow certain rules or norms that help us govern our behaviour over time. Such a norm is called a diachronic norm. A norm is diachronic if it requires the agent to have attitudes which fit together over time. For instance, a diachronic norm would be violated if at t_1 I intend to eat an avocado toast at t_2, and yet, without changing reasons for my intention, I abandon my intention to do so at t_2.[177] So a diachronic norm constrains me in requiring me to have a consistent pattern of attitudes over time. In short:

> Diachronic Norms: What attitudes I ought to have at a given time directly depends on what other attitudes I have at other times.

This contrasts with

> Synchronic Norm: What attitudes I ought to have at a given time directly depends on what other attitudes I have at this time, but not on what attitudes I have at other times.

Take *means-end coherence:* If you at a time intend to buy a house, and at the same time believe that not eating avocado toast is necessary for buying a house, you ought to intend to not eat

[177] Bratman (2012), p. 79.

avocado toast. Means-end coherence is synchronic: it coordinates how my attitudes hang together at a given time, but *not* cross-temporally. Diachronic norms can be understood as a temporally extended version of synchronic norms.

Dougherty's argument tries to show that if a set of attitudes produces pragmatic costs in a cross-temporal case, this shows the irrationality of that set of attitudes.[178] calls this the No-Way-Out argument[179]:

The No Way Out Argument

(1) A set of attitudes is irrational if there are cases where no matter what you do, you will have done something that, given those attitudes, you rationally ought not have done.

(2) Tragic attitudes are a set of attitudes that will, in some cases, no matter what you do, lead you to do something that, given those attitudes, you rationally ought not have done.

(3) Therefore, Tragic Attitudes are irrational.

This explains Dougherty's case against future-bias. If I hold a set of attitudes (future-bias + risk-aversion), this will lead me to do something that I rationally ought not to have done. Therefore, this set of attitudes is irrational, and since risk-aversion is rationally permitted, future-bias is irrational.

However, Hedden criticises this line of reasoning by challenging (2): To get (2) going, we need the crucial assumption that the

[178] Hedden (2015), p. 433.

[179] Hedden (2015), p. 433.

rational ought can be applied to a sequence of actions over time and not only particular action at a given time.[180] We need to assume not only that we rationally ought or ought not to Early Help, or that we rationally ought or ought not to Late Help – we have to assume that we rationally ought or ought not to (Early Help, Late Help). Hedden argues that our options for choice at a time consist in only what we are able to decide at that time – if we cannot decide about it at that time, then it is not an option for choice, and we cannot be criticised for it. Therefore, the rational ought can only be applied to a specific choice of action at a time, and cannot be applied to a sequence of actions over time – to do so would be a category mistake.

If we follow Hedden's line of reasoning, the No-Way-Out argument loses (2), and tragic attitudes do not yield irrationality. Therefore, holding future-bias and risk-aversion together might be tragic sometimes and lead to pragmatic costs, but aren't irrational, because the rational ought does not apply to a sequence of actions like (Early Help, Late Help). You can be criticised for Early Help, and you can be criticised for Late Help, but to apply the rational ought to both actions combined in a sequence would be a mistake.

This explains a big part of why Dougherty's argument does not reduce my credence in my belief that future-bias is rationally permissible. It would be much more convincing if Dougherty had shown how acting after future-bias at one given time only would lead to pragmatic costs – then I could be criticised for choosing

[180] Hedden (2015), p. 434.

something that is bad for me. But in the pain-pumps-case, I am criticised for ending up in a sequence, rather than for my choices at a time. That feels harsh. In other words,

(3) If you are risk-averse and future-biased, you will choose lesser future goods over greater past goods just because they are in the past.

of the compensation-argument might be false, as I don't at any time choose lesser future goods over greater past goods – whenever I choose, I choose the greater good, it so just happens that the sequence of my choices leaves me worse off.

To illustrate this point further, consider an illustration by Arntzenius, Elga and Hawthorne[181]: let's say you're not the one choosing the pills, but you get me and Dougherty advising you. As we also have other commitments, we split the task: I advise you for Early Help, while Dougherty advises you for Late Help. However, none of us can communicate with each other and and cannot influence the other person's advice. Given that Dougherty and I both have your best interests at heart and want to avoid any risks for you, and also prefer that your ordeal is in the past behind you, both me and Dougherty will advise you to take the pill. So, you end up with one minute more pain than you would have without us. Should you fire us as advisers?

The upshot is not that we are bad advisers, but that the ability to coordinate and causally influence future choices with our present choice. The problem is not that Dougherty and I are

[181] Arntzenius, Elga and Hawthorne (2004), p. 268.

future-biased, but that we lack the ability to bind each other's decisions – and similarly, you as a patient choosing both pills are not irrational because of future-bias, rather you at that moment lack the ability to coordinate the first pill with the second. Arntzenius, Elga and Hawthorne state it in the following way:

> Rational individuals who lack the capacity to bind themselves are liable to be punished, not for their irrationality, but for their inability to self-bind.[182]

Being unable to self-bind in a situation like this does not make you irrational, as it is not up to you, but to circumstances that you cannot make your present choice influence your future choice: if you were offered both pills at once, you would refuse at least one, and if you would be able to bind yourself to a commitment of only taking one pill, you would also be able to avoid Dougherty's Dutch book. But the lack of such an ability does not make you irrational, or as Arntzenius, Elga and Hawthorne put it: "Some agents who are led to ruin this way are perfectly rational. It is just that certain situations exploit rational agents who are unable to self-bind."[183]

Hence, we should not be criticised for being part of diachronic tragedies. And it's worth noting again that in order to criticise near-bias, we do not even need to get near a diachronic tragedy.

Does it also say something about Greene and Sullivan's cookie-argument? Recall that the cookie-example does not involve a

---

[182] Arntzenius, Elga and Hawthorne (2004), p. 269.

[183] Arntzenius, Elga and Hawthorne (2004), p. 269.

sequence of actions over time, hence it's not a case of diachronic tragedy. This, I believe, is an advantage Greene and Sullivan's argument has over Dougherty's, even though, as I have shown earlier, their additional premise appears to be false:

> Weak No Regrets If I have full information about the effects the options available to me have, then it is *permissible* for me to avoid the options I know that I will regret choosing.

Pretending that it isn't false, let's look closer: Sullivan's and Greene's norm seems not diachronic, since it refers only to your beliefs you *now* have about future effects of your *current* options, and hence seems more plausible than Dougherty's argument at first glance.

However, a closer look will reveal Weak No Regrets as a diachronic norm that makes my attitudes depend on each other over time:

> Weak Regret Avoidance: If I know at t_0 that I regret A at t_1, then I am permitted to not choose A.

What this principle relies on is a variant of the so-called reflection principle that requires you to update your beliefs according to your future beliefs. In other words, what you should believe now should depend on what you reasonably think to believe in the future. The variation here is that it is not your future beliefs but your future preferences that your current preferences should be dependent on. This, if we read preferences as a kind of desire, resembles what Elizabeth Harman calls Reflection for Desires:

> Reflection for Desires: If a person reasonably believes that in the future she will reasonably prefer that p not be true, and she reasonably believes that she won't be in a worse epistemic or evaluative position at that time, then she should now prefer that p not be true.[184]

In short, Weak No Regret is diachronic because it makes your current preferences depend on what you know what your future preferences will be like. And that makes Weak No Regret subject to criticism against diachronic norms, and the reflection principle in general. Hedden has counterexamples against the reflection principle that would apply,[185] but I will focus on the case Harman herself gives, as it directly applies to Reflection for Desires.[186] Her counterexample-recipe goes like this:

> An agent ought to perform action A, and it will be better in every way the agent cares about if she performs A, but she knows that if she performs A, she will reasonably regret doing A.

If cases like this are conceivable, Reflection for Desires is false. Cases like this are conceivable:

> Teenage Mother: A 14-year old is pregnant and considering whether or not to have the child. She knows, if she gives birth to the child now, it will be harder for her to get a good education, live a fulfilling life and be a good

---

[184] Harman (2009), p.187, I have reversed so that it resembles regret, not gladness.

[185] Hedden (2015), p. 463.

[186] Harman (2009), p. 193.

> mother to her child. However, if she does not have the child, she knows that she will reasonably regret it, even if it will be better overall in every way.[187]

Hence, Reflection for Desires is false, as the 14-year old shouldn't now prefer to have a child now, even though she will later have reasonable grounds for regret. Weak No Regrets is weaker than Reflection for Desires, as it requires full information and only asks for permissibility to avoid regret, but this case still puts pressure on the principle: assume the Teenage Mother will not feel regret if she has the child, as she fully devotes to her child, but will feel regret when deciding against the child, even if it is the better choice in every way. As already stated in the section before, Weak No Regrets seems false.

5.4 What's Irrational About Losing?

So far, so good for defenders of future-bias's rationality. However, you might not go as far as Hedden in your scepticism about diachronic norms – some diachronic norms might be perfectly justifiable, to criticise tendencies over time. For example, you might hold the belief that climate change is the most urgent problem humanity faces, and that we should reduce carbon emissions to stop it. Then, you get invited to a conference, and as one flight does not make a difference to climate change, you take it. Then, a second conference accepts you, and again, you fly, as one flight does not make a difference (this is oversimplified

[187] Harman (2009), p. 181.

and false, you do make a difference).[188] Same for the third and fourth conference this year. Each singular instance does not make a difference, but in combination, you contribute massively to climate change. So, shouldn't there be a diachronic norm that governs all your flights over time, requiring you to make your attitude towards one flight dependent on other, later flights? Your preference for both flying and reducing carbon emissions seems inconsistent, even if a single choice at a time seems permissible. So, a norm governing over several choices at several times seems helpful to avoid this behaviour.

What I will try to show next is that both Greene's and Sullivan's and Dougherty's argument lack something that the above example has: their arguments do not show an actual inconsistency within our belief system, but merely pragmatic costs of holding two attitudes together. The same is the case for Dougherty's argument – merely pointing at some pragmatic costs connected to your belief system is not enough. This is not the case for the flying/reducing emissions example; here you have a diachronic case of clear inconsistency in your preference for flying and reducing carbon emissions at the same time.

[188]This is oversimplified and false, as I have stated it. Your individual emissions do make a difference, as they are either affecting the outcome directly by pushing emissions closer to catastrophic thresholds (see Broome 2018) or are likely to be a triggering case for airlines to book more flights (see Kagan 2011).

Let's start with the cookies again: the reason why I am not convinced is that regret-avoidance and future-bias aren't actually inconsistent.

Compare

> Weak Regret Avoidance: If I know at t_0 that I regret A at t_1, then I am permitted to not choose A.

again to the already mentioned principle

> Reflection: $P_0\ (A|P_1(A)=r)=r$

Reflection is a principle that asks you to match your current credences about A to what your future beliefs about A will be. For example, if you're very sure that you will feel love for avocado toast after you made one, that gives you a reason to feel the love for avocado toast now. Reflection is stronger than Regret Avoidance, but according to both, my attitude at t_1 ought to have an impact on my attitude at t_0.

There is a wide range of counter-examples against Reflection, as sometimes it seems that following Reflection is irrational, and sometimes not being reflective seems rationally permitted. Arguments in favour of Reflection are very often Dutch Book cases, structurally similar to Sullivan's cookie case, where not being reflective results in a pragmatic loss. However, Christensen is puzzled why this doesn't seem to convince him, and I believe that his explanation of why Dutch Book arguments don't work to support Reflection can help us explain why Sullivan's Cookie-

Argument does not convince me. Let's start with a slight variation of the Cookie-Case, inspired by Christensen.[189]

> Biscuits: Suppose that you're married. Suppose also that you are future-biased and regret-averse. Now I, a clever biscuit-bookie, offer you (not your partner) a choice. On the condition that neither of you is allowed to share, either you get two biscuits at t_1, or your partner gets one biscuit later at t_2. You, your partner and I know all future consequences of our choices. How do you choose?

If you're future-biased and regret-averse, you'd pick the one biscuit for your partner, right? After all, you're future-biased, and after you have had your two biscuits, the event would be discounted. But then, after you've eaten your biscuits, you'd rather your partner have one biscuit in the future, rather than you having had two biscuits in the past. So, you'd regret your choice. Since Weak No Regrets is true, you are allowed to avoid options that lead you to regret. Hence, you'd choose one biscuit for your partner over two biscuits for yourself, and you'd live happily ever after.

Did you do anything wrong? I don't think so, it seems that this behaviour is acceptable and should not be criticised. Maybe you think we should hold on to a marital ideal like marriage solidarity, and you'd be wrong to not bring two biscuits home, but only one. However, that seems like dubious relationship advice, so in

[189] Christensen (1991), pp. 242-246.

absence of an independent reason why you shouldn't give your partner the biscuit, it seems perfectly permissible to do so.

Why does *Biscuits* seem much less problematic than *Cookies*? As you might have noted, the cases are structurally identical except for the fact that the baked goods are split up between two people. Does the feature of being interpersonal completely change the case? I don't think so, it seems to me that if your behaviour in *Biscuits* is permissible, then it should also be in *Cookies*, as there are no strong, relevant differences between the cases. Even if you think that there is a significant difference between interpersonal and intrapersonal matters, in a case involving biscuits and another person that is not you but very, very close to you, things should be so similar that the interpersonal vs intrapersonal shouldn't matter – you're not asked to sacrifice your life or someone else's. Hence, *Biscuits* and *Cookies* should be treated similarly, and then *Cookies* seems to be as permissible as *Biscuits* is. So, Sullivan's argument fails to demonstrate the irrationality of future-bias, even without us denying Weak No Regrets.

The reason for this, I believe, is that Weak No Regrets is not actually inconsistent with Future-Bias. Even synchronically, when I both hold the principle of Weak No Regrets and am future-biased at the same time, it doesn't seem to result in a clash within me.

Let's look closer: inconsistencies usually happen with beliefs and principles. Here, Weak no Regrets should conflict with the permissibility of future-bias. Compare this with me holding the beliefs P and ~P at the same time – that seems to be an

incoherency I should avoid. However, believing P and Q at the same time, even if exploitable by a very clever bookie, seems not to be an actual incoherency. So, why should it be irrational to both believe Weak no Regrets and permissibility of future-bias at the same time?

Maybe it's about attitudes? Let's say I am both annoyed and not annoyed with my partner. This is possible, and happens a lot probably, but that still seems somewhat more incoherent than having the attitudes of being regret-averse and future-biased at the same time. With annoyed/not annoyed, you could immediately ask me to clarify, to make sure that I am not making a mistake, or that I'm not somehow confused, as I could be annoyed about a certain aspect and not annoyed about others. That is not obvious with regret-aversion and future-bias – you'd have to build a complicated, far-fetched thought experiment, which, as a result, would only lead to pragmatic costs, but not an actual inconsistency..

We can apply the same trick to Dougherty's argument.:

> Assume it's not only me undergoing the operation, but me and my partner. I'm being operated on Tuesday, my partner on Thursday. Assume that I care for my partner to some extend and vice versa, and we have a 50% chance for each course.

	Tuesday	Thursday
Early Course	4 hours of pain (me)	1 hour of pain (partner)

Late Course	0 hours of pain (me)	3 hours of pain (partner)

Assume that I am future-biased and risk-averse. Dougherty comes along and offers two pills, Early Help, which reduces my pain on Tuesday on the early course by -29 minutes, but increases it on the Late Course by 31. He also gives me a second pill, Late Help, which I can give to my partner, to increase their pain on the early course by 30, and decrease their pain by -30 on the late course.

	Effect of Early Help	Effect of Late Help	Overall Marriage Effect
Early Course	-29 (me) min of pain	30 (partner)	1
Late Course	31 (me)	-30 (partner)	1

Since I am risk-averse, I will take Early Help. As I am future-biased, after Tuesday, my pain is discounted. I also offer Late Help to my partner, since it also reduces her risk. Regardless of what happens, together we will always end up with one additional minute of pain. Am I irrational?

The case is analogous to Dougherty's pain-pumps, with the only difference being that the operations are split up between two people. In this case, as with Sullivan, it seems perfectly fine to take both pills, even though we together end up with more pain.

Maybe the unconvincingness here comes, again, from the point that there's no real inconsistency between risk-aversion and future-bias. Just pointing at pragmatic costs isn't enough, if it's perfectly fine for two different people to jointly suffer these consequences.

The upshot is that pragmatic costs like those Sullivan and Dougherty point out are significant only insofar as they are a symptom of an underlying inconsistency.[190] Vulnerability towards pragmatic costs are, when it comes to holding beliefs and attitudes, not a problem per se – without a deeper, independent reason, pragmatic costs don't tell us a lot about rationality or irrationality of attitudes. The deeper independent reason is this: pragmatic costs show irrationality when they point towards a systematic incoherence within our sets of beliefs and attitudes. This is the case for the classical synchronic Dutch Book cases, which show epistemic inconsistencies that are defective in themselves - but not for pain-pumps and cookies, as they merely show the exploitability of holding certain attitudes in combination. Or as Christensen puts it, "Dutch book vulnerability is philosophically significant because it reveals a certain inconsistency in some systems of beliefs, an inconsistency which in itself constitutes an epistemic defect."[191] This is not the case for cookies and pain-pumps, and that is why the arguments are unconvincing, even if showing pragmatic exploitability.

---

[190] Christensen (1991), p. 238.

[191] Christensen (1991), p. 239.

To further illustrate, consider the following: if we took arguments like those from Sullivan and Dougherty too seriously, we risk running an overgeneralised argument of ruling out holding any attitudes and beliefs together that might in combination result in being exploited. For example, my belief that I ought to help others in need, and my attitude of trusting other people's word might lead me to being exploited quite easily even without a very clever bookie. That does not show that it is irrational of me to believe that I should help others or that my attitude of trusting what other people say is not rationally permissible. If we extend rational permissibility this far, a lot of things will end up being irrational.

To shortly summarise, I have tried to explain why Sullivan's and Dougherty's arguments fail:

Compensation Argument against Future-Bias

(1) Risk-Aversion/Regret-Aversion is rationally permissible.
(2) A rational agent prefers her life to go forward as well as possible.
(3) If you are risk-averse/regret-averse and future-biased, you will choose lesser future goods over greater past goods just because these are in the past.
(4) Your life would go better if you chose the greater good over the lesser good.
(5) Therefore, a rational agent would not be future-biased.

Firstly, Dougherty's argument relies on a diachronic tragedy, which, as I have argued, we shouldn't be criticised or blamed for, since we should be criticised only for our choices, and we do not

choose tragedies, so (3) seems false. Secondly, Sullivan's and Dougherty's arguments point towards pragmatic costs of holding future-bias together with either risk-aversion or regret-avoidance but fail to show an actual inconsistency between the attitudes. Hence, (5) just simply does not follow from the premises, and the argument is invalid, as an agent can still be risk-averse/regret-averse and future-biased at the same time without being irrational.

5.5 Sunk Costs and Temporal Neutrality

The compensation-argument failed to establish that future-bias is not rationally permissible because it leads you to a worse life overall. But even if it did succeed, the arguments by Dougherty, Greene and Sullivan would only show the impermissibility of future-bias in cases where I risk pragmatic loss – and the same can be said for temporal neutrality. In the rest of the chapter, I would like to return the favour to the friends of temporal neutrality, and propose that, in a lot of cases, it "pays" to be future-biased, and temporal neutrality leaves us worse off. The main idea is to put forward very common sunk cost fallacies via permissible preference shifts.

To reiterate, we are challenging the rationality of

> Past Neutrality: An agent is temporally neutral iff for two exclusive events E_1 and E_2, with E_1 being in the past and equally good or bad as E_2, the agent is indifferent between E_1 and E_2.

Assume the following:

> Preference change: In absence of independent reasons against it, it is permissible to change your preferences.

I believe this to be plausible enough. Should nothing speak against it, e.g. unjustified beliefs, irrationality, or you making a grave mistake about something that ends up making you unhappy, it should be perfectly acceptable to change your preference about something.

Now consider the following case:

> Concerts: Say that at t_1, I liked Linkin Park and had a very strong desire to see them. I can now buy a ticket to see Linkin Park at t_2, some time in the future. Since I really want to see them, I buy a ticket. However, between buying the ticket and the concert, between t_1 and t_2, I listen to folk music for the first time, and am so amazed that I change my taste – I now like folk music and dislike Linkin Park.
>
> Now, at t_2 (on the evening of the concert), I also discover that there's a folk session around the corner for free. However, because I'm temporally neutral, I treat my past desires as equally important as my present desires. Since my past desire to see Linkin Park was stronger than my current desire to go to the folk session, the former outweighs the latter, given that I'm temporally neutral. So, I go to the Linkin Park concert, and suffer through it, wishing I hadn't bought the ticket.

Here's the tragedy: At t_1, I choose buying ticket over not buying ticket. Because of temporal neutrality, at t_2, I choose Linkin park over folk concert. But overall, I'd prefer <buying ticket, folk concert> over <buying ticket, Linkin Park concert> because of

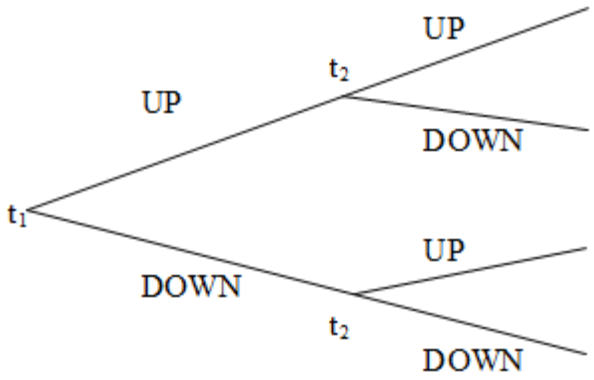


Figure 4: Hedden (2015)

my change in taste. So, given that taste change is justified, temporal neutrality leads to pragmatic loss.

This resembles the structure of diachronic tragedies we outlined above. Given the assumptions, each choice as such is , rationally justified, but combined into a sequence, we end up where we don't want to be.

A crucial assumption here is what is implied by Temporal Neutrality according to Sullivan: If I am temporally neutral, I am taking into account my past reasons, desires and preferences as much as my present and future ones. Without this, the argument would attack a straw-man. Luckily, Sullivan is rather explicit about this:

> "Temporally neutral agents reason differently than their time-biased counterparts. For instance, temporally neutral

agents sometimes take their past preferences into account when deciding whether to complete projects."[192]

So, we can put forward a generalised argument against all variants of Temporal Neutrality that include past preferences and past reasons to act:

Sunk Cost Argument against Temporal Neutrality:

(1) In absence of independent reasons against it, changing my preferences is rationally permissible.
(2) A rational agent wants her life to go forward as best as possible.
(3) My life will be worse off overall if I give equal concern to my past as to my present and future and change my preferences.
(4) Giving equal concern to my past as to my present and future leaves my life worse off overall.
(5) Hence, giving equal concern to my past as to my present and future is irrational.

Premise (1) and (2) follow from our assumptions. (3) is shown by Concerts – if I both change my preferences and give equal concern to my past than to my present and future, I will end up in a diachronic tragedy. (4) follows from (1) and (3), and (5) from (2) and (4). There we go, we've got a compensation-argument against temporal neutrality.

[192] Sullivan (2018), p. 95.

This is a standard example of the sunk cost fallacy, tailored for temporal neutrality. Sullivan has anticipated this challenge and prepared an answer with entertaining cases to this.[193] Consider

> Dragon Sagas: George has spent years working on a series of seven fantasy books called The Dragon Sagas. He is nearly finished with the final volume. He has always wanted to finish it: he did not want to die without explaining the fate of his fictional kingdom. He learns he only has a few days to live. And because he has such a short time left, he doesn't predict that finishing the final book will matter either way for his life going forward. But he reasons that he prefers to finish because he ought to tie up this last "loose end" in his career.

and compare it with

> Rice Cooker: A few months ago, Peter purchased a rice cooker. At the time he preferred to cook more rice – he was planning to eat healthier and save money. But he rarely used it. Now it takes up valuable space on his kitchen counter. Tonight, Peter just barely prefers ordering pizza to cooking rice. But he reasons that because he made an investment in the appliance in the past, this tips the scales in favor of staying home and cooking rice.

Sullivan thinks that in Dragon Sagas, there's reason to stay true to sunk costs, while in Rice Cooker, there isn't. What sets these cases apart, according to Sullivan is that Dragon Saga is

[193] Sullivan (2018), pp. 96-97.

"Honouring Past Preferences", while Rice Cooker constitutes a sunk cost fallacy. What is distinctive to sunk costs is that sunk costs honour a past preference the agent now regrets having, while in a proper honouring case, the agent does not regret having had that preference. George, even though he does not have the preference of finishing his saga currently, also does not regret it now to have had it in the past. Peter however, does not have the preference for rice, and also regrets to have ever preferred to have rice. Hence, George has reason to finish his book, and Peter has no reason to eat rice.

The problem with Sullivan's response is that it doesn't answer Concerts-cases. In Concerts, I do not regret having preferred Linkin Park – listening to the band made me very happy at that time, it just doesn't do anything for me anymore. Since I am temporally neutral, I also do not regret buying the ticket for Linkin Park, since at that time, that did fulfil my desire to finally see Linkin Park live. What I regret is rather ending up in a diachronic tragedy – hence, although I am honouring past preferences properly, I still end up worse off overall.

To be clear, I do not think at all that this is a knock-down argument against temporal neutrality. We can again note that this does not show an actual inconsistency between temporal neutrality and preference shifts as such, it merely points towards pragmatic costs without an independent reason of epistemic incoherence. Hence, this does not establish the general irrationality of temporal neutrality.

However, sunk cost fallacies are a very common phenomenon and happen both in our private lives as well as in complex planning cases, where we do not want to admit to have failed at planning correctly, and ignore evidence and reasons for changing our preferences as a result.[194]

In comparison, cases like Dougherty's pain pumps are very rare, and cases like Sullivan's cookie-fallacy are easily outweighed by other pragmatic gains, such as ongoing satisfaction or fond memories. A thought experiment being unrealistic should not lead to it being dismissed. However, if we compare an unrealistic thought experiment with a very realistic one that argues in the opposite direction, that may affect our credences differently. Therefore, as sunk costs are realistic and very common, while Dougherty's and Sullivan's cases are quite removed, I believe that the compensation-argument carries a lot more weight against temporal neutrality than against future-bias.

Future-bias can serve as an insurance policy against the sunk cost fallacy – if we discount our past events, reasons, desires and preferences, and focus on our present and future, we will not fall prey to overestimating our past investments, and take present and future into account properly. Hence, future-bias is not only not irrational – it might be good for you.

[194] For more on the sunk cost and planning fallacy, see Kahnemann (2012).

6 How to be Future-Biased

6.1 Introduction

Cillian Murphy always dreamt of being a rock star. He started writing songs when he was 10, and dismissed his English teacher's encouragement to pursue acting, as being a rock star is a lot cooler. In his late teens and early twenties, he actively pursued a career as a musician, playing in a band with his brother, both growing up obsessed with the Beatles. This led to them being offered a record deal for five albums, and Murphy's dream of becoming a rock star being realised.

However, when he was offered the record deal, Murphy had a change of heart: after seeing the film 'A Clockwork Orange', he wanted to become an actor. Does Cillian Murphy's past life events, reasons and preferences for becoming a rock star give him reasons to choose the record deal, even if he currently does not want it?

Set aside for a moment whether it is possible for reasons to be generated from desires at all, and ignore how dodgy the record deal was in reality, to focus on this question: would it make a difference whether Cillian Murphy was temporally neutral or biased towards the future?

Meghan Sullivan, a proponent of temporal neutrality, thinks it does.[195] According to her, temporally neutral agents reason

[195] Sullivan (2018), p. 151.

differently to time-biased agents. The former, she says, take into account their past preferences and choices as much as they consider their anticipated future preferences and choices for their current decision. The latter do not. So, applied to Cillian Murphy, Sullivan proposes:

a) If Murphy is temporally neutral, then he can take his past preferences and choices into account.

Mirroring this claim is:

b) If Murphy is future-biased, then he cannot take his past preferences and choices into account.

Note that b) is not the negation of a), merely a mirror claim, as Murphy could not be future-biased and still not take his past preferences into account. What I'm interested in, however, is whether temporal neutrality and future-bias make a difference for Murphy in his choice here, so I will examine a) and b). If we only have a), then Murphy could not be temporally neutral and still take his past preferences and choices as a reason to choose. If we only have b), Murphy could not be future-biased and still ignore his past preferences and choices.

Even though I disagree with Sullivan on temporal neutrality, and think that future-bias is rationally permissible, or sometimes even rationally required, I will defend her claim a), and aim to establish b), so that it makes a difference in how we treat our past in our decisions if we are temporally neutral or future-biased. The more important claim for me is b) - I will defend the claim that being future-biased means that your past preferences and reasons are not taken into account for present and future considerations. So,

assuming that Cillian Murphy is future-biased, he should not take his past rock star preferences as reasons to choose – and become an amazing actor instead.

More generally, the aim of this chapter will be to establish a more comprehensive principle explaining why we are future-biased in particular instances, e.g. why we prefer pains to be past rather than future, by explaining how future-bias means that we disconnect from our past in a way that present and future reasons for choices and preferences will always be trumped by present and future reasons.

I will start by briefly distinguishing two different ways of being temporally neutral or future-biased, and showing where the intuitive appeal behind these versions comes from. Then I will explain different ways of conceptualising the discounting mechanism behind future-bias and argue that b) is the best explanation of it. I will then show how accepting b) and treating future-bias as rationally permissible would undermine some forms of moral theories.

6.2 What is Temporal Neutrality supposed to be?

Brink describes temporal neutrality as the requirement that "agents attach no normative significance per se to temporal location of benefits and harms within someone's life and demands equal concerns for that person's life".[196] This definition is a bit vague and open to interpretation, and also seems quite

[196] Brink (2010), p. 1.

comprehensive, in the sense that being temporally neutral demands quite a lot from an agent: not only should I not prefer a future event over a past event, but I should also give equal concern to the past parts of my life, including preferences, decisions, and choices in my past.

This version of temporal neutrality would demand that Cillian Murphy takes his past seriously in a quite comprehensive way: his past preference to become a Rockstar counts as much as his present preference not to, even as much as his future (anticipated) preference to become an actor.

> Comprehensive Temporal Neutrality: Do not prefer an event over another only because of its temporal location and give equal concern to all parts of your life.

This quite demanding version of temporal neutrality is not only endorsed by Brink, but also by Sullivan and Dougherty, who see temporal neutrality as a requirement to ensure extended agency of a person over time.[197] However, the comprehensive version is not strictly necessary to rule out being near-biased or future-biased:

> Minimal Temporal Neutrality: Do not prefer an event over another only based on its temporal location.

In this version of temporal neutrality, Cillian Murphy would not be allowed to prefer goods closer in time to goods in the far future, just because they are closer to the present. He would also not be

[197] See Sullivan (2018) and Dougherty (2015).

allowed to prefer his pains to be past rather than future. However, it would be completely acceptable for him to ignore his past preferences, such as his preference to be a rock-star, or that all his life events so far pointed towards him becoming a musician. A more minimal version of temporal neutrality would just mean: Don't be time-biased, while being silent on whether past preferences, choices and actions should play a role in your decision-making.

As most authors are working with the comprehensive version of temporal neutrality, I will follow their lead: if I say temporal neutrality, it will mean comprehensive neutrality from this point on. However, before we move on, do keep in mind that comprehensive temporal neutrality is much more demanding, and needs more arguments in favour of it than minimal temporal neutrality – with the latter it's enough to argue that time-biases are bad or irrational, e.g. by pointing out how time-biases are arbitrary or lead you to pragmatic loss, while the former may need a more developed argument. Let's look at what it means to be future-biased, bearing in mind that there are two versions of temporal neutrality to correspond to.

6.3 What's Future-Bias again?

Let's summarize briefly what future-bias is. Most people prefer bad things to be over. If you think back to a painful appointment at the dentist or to an awful social event, you are glad that it's over and done with. At the same time, you'd like good things to be ahead of you. You'd prefer your vacation or a good concert to

be in your future rather than in your past, already gone. If you have this sort of preference, you are future-biased:

> Minimal Future-Bias: An agent is future-biased iff for two exclusive events E_1 and E_2, with E_1 being in the past,
>
> - where E_1 is at least as positive as E_2, the agent prefers E_2 to E_1 because E_1 is in the past and E_2 is not, or
> - where E_1 is at least as negative as E_2, the agent prefers E_1 to E_2 because E_1 is in the past and E_2 is not.

If a person is future-biased, she prefers bad things to be past and good things to be future. Note that this definition of future-bias is quite minimal: it only refers to events, their positive and negative evaluative properties, and their temporal location, and thereby mirrors *minimal temporal neutrality*, which rules out this kind of future-bias. Of course, with *comprehensive temporal neutrality*, you should also not be future-biased, but you are required to more than that. Is there a form of future-bias that mirrors *comprehensive temporal neutrality*?

> Comprehensive Future-bias: An agent is future-biased iff she discounts her past parts of life against her present and future parts.

With this kind of future-bias, an agent would discount not only events, but anything belonging to past parts of her life, including preferences, choices, actions. For example, Cillian Murphy, if future-biased in a comprehensive way, would discount his past preferences to become a rock-star against his present and (anticipated) future preferences of becoming an actor. So, like with *comprehensive temporal neutrality*, *comprehensive future-*

bias also makes a difference in how Cillian Murphy should make his choice.

So, how should we read future-bias? Note that it does not depend on what kind of temporal neutrality you find yourself drawn to: both minimal and comprehensive temporal neutrality rule out both kinds of future-bias. Most authors in the debate only examine the appeal behind minimal temporal neutrality, which as a preference pattern seems to be quite natural – so natural that most would think it to be permissible to have this attitude. The permissibility of minimal future-bias is classically motivated by Parfit's Past and Future Operations case:

> My Past and Future Operations: I am in some hospital, to have a safe, but painful surgery. Because the operation is so painful, patients are afterwards made to forget it.
>
> I have just woken up. I cannot remember going to sleep. I ask my nurse if it has been decided when my operation is to be, and how long it must take. She says that she knows the facts about both me and another patient, but that she cannot remember which facts apply to whom. I may be the patient who had his operation yesterday, lasting ten hours. I may also be the patient who will have a short operation later today. I either suffered for ten hours yesterday, or will suffer for one hour later today.

> It is clear to me which I prefer to be true. If I learn that the first is true, I shall be greatly relieved.[198]

This case shows that most of us would prefer bad things like pain to be past rather than future, even if the past pains are worse than the future pains. So, considering Parfit's operations-case, future-bias seems so natural that it ought to be permissible. Dorsey[199] states that it's so natural that we care more about events ahead of us, that we want bad things to be in the past that future-bias may be considered a brute feature of our practical rationality. So, anyone who would want to show that minimal future-bias is rationally impermissible would have to carry a quite heavy burden of proof, since the attitude's permissibility is so intuitively appealing.

With comprehensive future-bias, things might be different: you might think it quite plausible to prefer bad pains to future pains. But discounting your past preferences and reasons against future preferences and reasons is a step further, as you might be less ready to exclude parts of your past from your decision-making. Cillian Murphy might have the exact same reaction if he were in the operations case, but that does not automatically mean that he would discount his past preferences of becoming a rock-star.

But just like defenders of temporal neutrality tend to defend the comprehensive, not the minimal version, I would like to suggest that we move on with a conception of future-bias that includes

---

[198] Shortened from Parfit (1984), pp. 165-166.

[199] Dorsey (2016), p. 5.

not only events, but all parts of life, including preferences, reasons, choices of agents.

Reasons for this partly mirror the appeal of a broader conception of temporal neutrality: if you're temporally neutral about benefits and harms of events, why not be temporally neutral about other aspects of your life too? It would be arbitrary just to single out pain and pleasure to be temporally neutral about, so why not be temporally neutral about all parts of life? In a way, temporal neutrality about benefits and harms of events can be seen as a part of a broader principle of comprehensive neutrality. Similarly for future-bias: If we're future-biased about benefits and harms of events, why not appeal to a broader version of future-bias that encompasses preferences and choices too? This way, we have a more systematic and less arbitrary way of being future-biased.

I earlier mentioned that arguments for minimal temporal neutrality - consisting usually in arguments that time-biases are irrational, e.g. by pointing out arbitrariness or pragmatic loss – are not enough to establish comprehensive temporal neutrality. In contrast to that, I believe that other arguments in favour of minimal future-bias do support comprehensive future-bias. For example, considerations about how future-bias can help us avoid sunk cost fallacies apply to a more comprehensive future-bias that includes past preferences and choices. A sunk cost fallacy happens when you stick to a course of action solely because of past investment in that course of action, but there is no expected benefit to your course of action in the future. Comprehensive temporal neutrality commits you to this form of fallacy, as you need to include your past preferences, reasons and choices into

your deliberation with the same weight as present and anticipated future preferences, reasons and choices. Mirroring this, comprehensive future-bias can help us avoid this fallacy by discounting past against present and future preferences, reasons and choices in our decision-making.

Another argument that supports comprehensive future-bias is asymmetry of control: we can appeal to the fact that we can influence the present and future, but not the past as justification for the rationality of future-bias. For example, Murphy can somewhat influence whether his future preference will still be to be an actor, e.g. by minimising predicted burn-out, choosing roles that he will find fulfilling, limit his exposure to toxic Hollywood circles. But Murphy cannot influence his past preferences for being a musician. As we should give more weight to what we can influence over what we cannot, Murphy should give much more weight to his future over his past preferences.

And finally, appeals to our tensed emotions tend not only to support minimal future-bias, but a broader conception of discounting the past: we don't only feel relief, grief, anxiety and nostalgia in terms of beneficial or harmful events – we grieve about past choices we made, we feel relieved that some of our past preferences are gone, and nostalgic for values and convictions we used to have in the past. So, if appeals to tensed emotions support a temporal value asymmetry, they do support a comprehensive, not minimal one.

So, in summary, concerns about arbitrariness, sunk costs, tensed emotions and control asymmetry support us going forward with

comprehensive future-bias, including past events, preferences and reasons, not with minimal future-bias, which is limited to past events and how beneficial/harmful they are. I am not saying that these arguments are all absolutely convincing and knock-down arguments against temporal neutrality, but if these concerns carry any weight, they support a comprehensive form of future-bias, not a minimal one. Going forward, if I say future-biased, I mean comprehensive future-bias.

Let's now look closer at what it means to be comprehensively future-biased. Another way philosophers and economists have used to describe future-bias is as a discount function – the value of an event is decreased if it's past. Sullivan argues that there is reason to think that future-bias is not an exponential or hyperbolic discount function, where the value of an event would decrease the further it goes into the past, but *absolute*.

If future-bias is absolute, we assign no or almost no value to a past event – as soon as it is past, it's just gone. If future-bias weren't absolute, we'd have a function with three variables:

(1) The temporal distance between now and the past event.

(2) The value I'd assign to the event if it were present.

(3) A discount-function.

So, let's say that I have a discount rate such that every 7 days, the value of an event halves. My painful operation I just had is a 10/10 now on the pain-scale, but after a week, I'd rate it 5/10, after two weeks, 2,5/10 and so on. In this case, my future-bias could be represented as an exponential discount function. If future-bias would be a non-absolute discount function, my

evaluation of the past event would be sensitive to the temporal distance to the past event. I'd map the value of the past event to how much time has passed. However as Parfit's Past and Future Operations Case shows, I do not. I just care about the past event being past. It would be absurd to hope for a future operation instead of a past one, regardless of how far away the past one is. As soon as an operation is past, I will always prefer it to a future operation. Hence, future-bias is absolute.[200]

So, to summarize: future-bias is absolute and shouldn't be represented as a non-absolute discount function sensitive to temporal distance towards the present. However, we should clarify further what it means to be absolutely discounting. In what follows, I will argue that we should be sceptical about future-bias being represented as a discount-function as such, and that we should go for a less value-based explanation of the preference pattern.

6.4 What's the Best Explanation for Future-Bias?

Future-bias is a preference pattern. How that preference pattern behaves should be explained clearly. I will argue that future-bias is not best explained by

> *Absolute Discounting*: If an agent is future-biased, the value of her past events, decisions and preferences are decreased to almost zero, such that the value drops to an amount that will always be outweighed by any positive

[200] Sullivan (2018), pp. 49-50.

future event or will always outweigh any negative future event.

which is the explanation most authors in the debate currently use, and is better explained by

> *Disconnecting*: If an agent is future-biased, her past events, decisions and preferences cease to generate reasons for preferring, choosing and acting now.

First consider why we should read future-bias as *Absolute Discounting*. *Absolute Discounting* explains Parfit's operations case well, and generally captures most hedonic cases where we discount past pain, e.g. when you're in your child's violin concert, the screeching just won't stop, and you just want it behind you. In hedonic cases, it seems that you would prefer pain to be past rather than future at a cost that is extremely high: you'd trade 1 hour of future pain against 10 hours of past pain, in Parfit's case.

Another reason why authors like to use *Absolute Discounting* to explain future-bias is that, as explained above, they generally believe future-bias to only apply to hedonic cases involving pain and pleasure. When it comes to non-hedonic goods, many authors believe that the intuitive appeal behind future-bias's rationality disappears.[201] *Absolute Discounting* would be odd in non-hedonic cases (e.g. reducing the value of your past achievements to almost zero?), but since future-bias is only applicable to hedonic cases, we should use *Absolute Discounting*.

---

[201] E.g. Sullivan (2018), Brink (2010), Hurka (1996), Dougherty (2015).

However, there are several problems with that conclusion. Firstly, if we look at empirical data, it might not actually be the case that we discount absolutely, even if we don't discount hyperbolically or exponentially. As Fernandes helpfully summarises the empirical findings, people usually do give past events a significant amount of value, compared to future events, as well as only discounting past pain to a certain point: while 92-92% prefer past pain to equal future pain, and future pleasure to equal past pleasure, 53-54% switch their preference if the amount of past pain or pleasure is doubled. At the same time, people report to be indifferent between 5.6 hours of past pain and 2 hours of future pain. This seems to show that the amount of past pain is still significant.[202]

However, this does not show that Sullivan is completely on the wrong track – empirical studies on what people report might insufficiently isolate the phenomenon of future-bias from other considerations. For example, participants being asked whether they'd prefer 5.6 hours of past pain or 2 hours of future pain might not understand that they won't have memories of the 5.6 hours, or not realise what it means to "choose" past pains over future pain. Sullivan can still appeal to Parfit's Past and Future Operations, where the case is cleaner, as the patient won't remember the past pain, and the past pain might have actually occurred. So, we should not dismiss *Absolute Discounting* on the basis of empirical data alone, as Sullivan proposes a version of

[202] As summarized by Fernandes (2019), p. 7. For the full studies, see Caruso et al. (2008), Caruso et al. (2018).

future-bias that people may hold if empirical studies account for factors contaminating intuitions on future-bias.

Secondly, *Absolute Discounting* leads to strange questions about how low or to which extent past pain is discounted. If past pain's value is reduced to "almost zero", there should be some amount of future pain that is so ridiculously low that the past pain still outweighs it. Think about the nurse telling you that you either had 100 hours of pain yesterday or getting an injection tomorrow – less than 10 seconds of pain (if you're not afraid of injections, you won't even have that much dread). If 100 hours of pain gets discounted to "almost" zero, it seems possible that some low amount of future pain would be preferable to the past pain, even with absolute discounting.

While this is a question that does need answering – is it really rational to prefer having had 100 hours of pain in your past rather than 10 seconds of pain ahead of you? – I think proponents of future-bias's rationality here should be steadfast and commit that it isn't irrational to have that preference. Bear in mind that you won't remember the actual pain, that your past pain will not affect your future in any traumatic way. In that case, it would not seem so weird for a person to choose 100 hours of past pain over 10 seconds of future pain. Of course, you could interpret the "almost zero" in absolute discounting as a value that is infinitesimally small, so that there is no amount of future pain that would be outweighed by it, and thereby get the right result. But *Absolute Discounting* does still lead to confusion here if we pair an infinitesimally small amount of discounted past pain and an equally infinitesimally small amount of predicted future pain

against each other – would a future-biased person throw a coin in this case? This, I think, does not really fit a steadfast position about discounted past pains that intuitively makes sense, and points towards treating discounting in future-bias in a different way.

What we could do to avoid all this confusion is to amend Absolute Discounting in a way to avoid both problems:

> *Lexical Discounting*: If an agent is future-biased, the value of her past events, decisions and preferences becomes a kind of value that will always be outranked by the value of present and future events, decisions and preferences it is compared with.

This form of lexical ordering of different kinds of value is most well-known through Mill's hedonism, where the crude pleasure-maximisation suggested by Bentham is replaced by Mill's suggestion that higher order pleasures, e.g. intellectual stimulation and learning by watching Cillian Murphy in The Wind That Shakes The Barley, will always outrank lower pleasures, e.g. crude satisfaction by watching Cillian Murphy in Peaky Blinders. Generally, lexical ordering of value takes the form that there are types of good G1 and G2 such that any small amount of G1 is better than any large amount of G2. Of course, the question arises about whether it is arbitrary to draw a sharp line between two goods G1 and G2 like this – where's the line between higher pleasures like The Wind that Shakes the Barley, and lower pleasures like Peaky Blinders?

Fortunately for our purposes, this problem does not apply for future-bias: neither is the line between higher and lower order values arbitrary – after an event has become past – nor is it as implausible as Mill's suggestion. This is because it captures how future-bias is absolute, rather than gradual in its discount function. At the same time, we can still make sense of why folk still assign value to the past, even when future-biased: people still assign a significant amount of value to their past, but that is a different kind of value than the value of past events. The value of past events might still be there for you if you're future-biased, you might still be able to assign a number to it and rank it, but it ceases to be relevant in a certain way in comparison to present and future value. The value of present and future events will have lexical priority over that of past events, so that present and future value will always outrank past value.

But in what sense does present and future value "outrank" past value? How do we make sense of the lexical ordering beyond saying that present and future value is more significant than past value?

Let's apply this to Cillian Murphy's situation: if Murphy is future-biased, and Lexical Discounting explains what it means to be future-biased, Murphy would still see value in his past. His past events, decisions and preferences of wanting to become a Rockstar still matter to him, just in a different way, and his present and anticipated future concerns always trump them. So, we could say, if we apply Murphy's future-bias to his decision whether he should pursue his music career or not, his past *ceases to be reason-generating*.

If applied to past and future pains, we could say the same. We might still be able to rank our past pains and assign significant value to them, but when it comes to an actual choice, our past pains cease to be reason-generating, while the higher order pains in present and future give us reasons to act, choose and prefer. So, we arrive at

> *Disconnecting*: If an agent is future-biased, her past events, decisions and preferences cease to generate reasons for preferring, choosing and acting now.

Our past events just cease to be reason-generating, and if we are future-biased, our past pain does not provide us with any reasons for preferring or acting anything anymore. The future pain, however, does give us reason to prefer it not to be the case, and reasons to try and avoid that pain. So, if you are future-biased, you disconnect from your past events in a way that they no longer give you any reason for choosing or preferring anything now, even if it is a hundred hours of past pain. The only thing left giving you a reason for a preference is the future injection – which you'd rather not have, even if you could still assign a value to your past pains. In short, *Disconnecting* explains how we can make sense of the lexical priority of present and future over past events, choices and preferences. So, in summary, if we move to address the problems appearing with Sullivan's *Absolute Discounting* explanation of future-bias, we quite naturally arrive at *Disconnecting*.

Sullivan and others might still object that *Absolute Discounting* still best explains future-bias, as it fits best with hedonic vs non-

hedonic cases: future-bias seems to only apply to hedonic goods, but not to non-hedonic ones. However, as I have argued in a previous chapter, future-bias does apply to non-hedonic cases, and it seems equally rational to be future-biased in cases of achievement, friendship or life and death. Assuming this is correct, we need an explanation for future-bias that works on both hedonic and non-hedonic cases. Here, *Disconnecting* works better than *Absolute Discounting*: if you look at a past big achievement in your life and compare it to a smaller achievement in the future, it might not necessarily make sense to speak of discounting your past achievement if you're asked which one you'd prefer. It might not be the case that the past achievement is worth almost zero to you. It might not even be the case that your past achievement is worth less to you now. The thing is just that you care differently about your past than about your future: your past achievement now doesn't generate a reason for preference like a future achievement would, even if in the past, your achievement did provide you with a reason for preference. Another example would be past friendships vs future friendships: it might not be that you discount your past, it simply ceases to be reason-generating for you. Hence, given that future-bias applies to non-hedonic cases, we are better served with *Disconnecting* rather than *Absolute Discounting*.

And finally, Sullivan, one of the main proponents of temporal neutrality, also thinks so. In her book, she describes temporally neutral agents as being different from time-biased (read future-biased) agents: temporally neutral agents take their past events, preferences and choices into account in making decisions. For

them, past preferences and choices can license present and future choices just as much as anticipated future preferences and choices can.[203] Hence, the difference between a temporally neutral agent and a future-biased agent is that the former does not disconnect from past events, preferences and choices, while future-biased agents do. As Sullivan treats temporal neutrality as a way of being connected to your past choices and preferences, future-bias should be treated correspondingly, as *Disconnecting*. Past preferences and choices are not taken into account for current and future considerations if I am future-biased.[204]

So, if we treat future-bias as disconnecting, we can say something about Cillian Murphy's situation:

a) If Murphy is temporally neutral, then he can take his past preferences and choices into account.

Following Sullivan, if Murphy is temporally neutral, then he can take his past reasons into account – it is not irrational to stick to his plan to become a rock star just because of my past reasons and no other reasons. If there are reasons speaking against him signing the record deal, they can be outweighed by his past reasons.

a) If Murphy is future-biased, then he cannot take his past preferences and choices into account.

If Murphy is future-biased, the past decisions would cease to be reason-generating for his decision now. If he then decides to sign

---

[203] Sullivan (2018), p. 151.

[204] Sullivan (2018), p. 131.

the record deal, even if he had no reasons to, he acts arbitrarily. If he decides to sign the record deal based on his past preferences to become a rock star even if there's reasons speaking against signing (e.g. because the record label is exploitative), he will act irrationally.

So in summary, if we take problems with *Absolute Discounting* seriously, and amend it to *Lexical Discounting*, where a future-biased person gives lexical priority to present and future value over past value, we will arrive at *Disconnecting* when we apply this view to a decision Cillian Murphy has to make. And if we treat future-bias as disconnecting from our past, not as discounting, being future-biased or temporally neutral does make a difference in how Cillian Murphy should decide. Luckily for all of us, he decided to become an actor.

6.5 Implications for Some Moral Theories

In the rest of the chapter, I will outline how understanding future-bias as both comprehensive, including preferences, choices and reasons in discounting, and lexical, assigning different priority to past and future value, will have implications for some moral theories. I will explain how treating future-bias as Disconnecting, and as rationally permissible, can undermine views that rely on past events, choices and preferences to justify certain choices.

Take for instance moral theories that not only consider what is ahead of you, but your whole life in moral evaluation. An example is

> Whole Life Egalitarianism: When distributing goods across different lives, it is required to take into account an entire lifetime and not only segments of a life.

For illustration, consider an example by Derek Parfit:

> A doctor has two patients feeling pain. Patient Anna's suffering is not as severe as the suffering of Bertha, but Anna has suffered much more than Bertha in her past. The doctor can only help one patient, and the treatment would relieve more suffering right now if it were given to Bertha.[205]

A whole life egalitarian would be committed to not only consider who is currently suffering the most, but also who has suffered, and how much through their entire lives. Anna might be in less pain right now than Bertha, but her past suffering might make her more deserving of treatment.

Another example comes from McKerlie:

> Imagine a Society where there is great inequality at all times. In the beginning, Antoinette is 100 times richer than Bob, but after ten years, Antoinette and Bob switch places, and now Bob is 100 times richer than Antoinette. After another ten years, Antoinette and Bob switch places again, and so on.[206]

---

[205] Parfit (1986), pp. 869-70.

[206] McKerlie (1989), p. 479.

Assuming Antoinette and Bob live equally long, at the end of their lives, a whole life egalitarian would say that they were equal, even though there was great inequality at any given time during their lives.

You might not find this very plausible or might think that this is somewhat uncharitable to Whole Life views. That might be so, but what is important here is not the specific kind of Whole Life view in question, or how they could react to these kind of cases, but rather the mechanism behind a Whole Life view. If you are a Whole Life egalitarian, you consider all aspects a life, including its past parts that make a difference to well-being, such as past events, decisions and preferences you might have had if they make a difference in well-being, and they can give you reasons to decide in a certain direction. Anna's past suffering might give you a reason to think her more deserving, Antoinette might justify current inequality by pointing out that Bob was having all the cake in the palace during the last ten years, and so on.

You might see already where I am going with this, but before that, a short caveat: Future-bias is intrapersonal. An agent is allowed to disconnect from her own past but dismissing other people's past might not be allowed. The doctor treating Anna and Bertha being future-biased and deciding that Anna's past is disconnected is different to Anna being future-biased and deciding to disconnect from her past suffering.

There might be cases where someone can be future-biased on someone else's behalf. Let's take a look at a variation of Parfit's operation case by offered by Brink[207]:

> Past and Future Pains of Others. You receive a message about your daughter who lives in another country. The message says that your daughter had an accident that injured her greatly, and that she will suffer great pain in an operation. This depresses you. But then you receive another message, telling you that the earlier message was delayed, and your daughter already suffered through the operation. Do you feel relief that it is already over?

In this case, you may be future-biased for your daughter by virtue of caring for her. And in the same way, a doctor, after listening to Anna explaining to him that her past suffering is over and done with, can be future-biased on her behalf if she is future-biased. However, if Anna is not future-biased, it seems doubtful that the doctor is allowed to be future-biased for her, as this might wrongly evaluate the moral situation. The doctor's action, in a way, would be morally paternalistic, because she is forcing her evaluation onto Anna, even while Anna may be fully rational.

So, future-bias does not apply to the doctor case if Anna is not future-biased. But let's assume she is – is the doctor still allowed to prioritise Anna over Bertha because of Anna's past suffering, given that the doctor is a whole life egalitarian?

---

[207] Parfit (1984), pp. 181-182.

If Anna says, "My past suffering doesn't matter to me, I don't care whether I get preferential treatment based on that, just look at momentary suffering", would the doctor be allowed to say "Nope, I'm still prioritising Anna, she suffered so much, but it's the first time Bertha ever suffered"? I think this is something Whole Life egalitarians might not be willing to say, for similar reasons as above – the doctor's action would be morally paternalistic, because she fails to take another person's view seriously, even though the other person is fully informed and rational.. For this reason, we should be careful to not apply future-bias too quickly in interpersonal cases.

However, what should be of more interest here is the intrapersonal scenario. Imagine Anna is indeed future-biased. She says "I am future-biased, I prefer my suffering to be past rather than future, but I still want preferential treatment because of my past. Doctor, please treat me, forget about Bertha." Is Anna making a conceptual mistake?

If we understand future-bias not only as a preference for suffering to be past rather than future, but that it also involves past preferences, choices, and events being no longer reason-giving – which I have argued for above – then I think she does. Her reason for wanting preferential treatment is not based on current suffering but past suffering, and that reason should be undercut by her not taking past events as reason-generating anymore. In other words, Anna prefers pain to be past rather than future because past events do not enter current considerations anymore. But if that is the case, then why would she want preferential treatment based on past suffering?

You might think that if we think back to *Lexical Discounting*, there's no reason why Anna is making a mistake: even if she prefers her pain to be past rather than future, she can still assign value to her past pains. She still thinks her past pains were really, really bad, even though she's glad that it's behind her. So, why can't she appeal to her past pains when future-biased?

The reason why she can't is that she would give lexical priority to momentary pain over any amount of past pains. If Anna knows that Bertha is in more pain right now, and more momentary suffering would be prevented with the doctor treating Bertha, Anna's past pains, even if they were immense and she would give a lot of negative value to them, would never outrank them if Anna were future-biased, as present and future pains dominate past pains.

So, as long as Anna knows that Bertha right now suffers more than Anna, and Anna is future-biased, the momentary suffering will take priority over Anna's past suffering, regardless of how big the past suffering was. Or in other words, Anna's and Bertha's momentary suffering give them reasons for preferring treatment, while their past suffering doesn't. Since Bertha right now suffers more than Anna, Bertha should be treated.

You might think that it's not entirely true that, if future-biased, past pains can never be reason-generating. To see that, let's amend the Doctor case:

> A doctor has two patients feeling pain. Patient Anna's suffering is exactly as severe as the suffering of Bertha, but

> Anna has suffered much more than Bertha in her past. The doctor can only help one patient.

If it were only momentary well-being that mattered, the doctor would flip a coin. Let's say again that Anna is future-biased. Can she appeal to her past suffering to convince the doctor to treat her? I think yes – in this case, it seems that even if Anna discounts her past, even if momentary well-being takes absolute priority over past well-being, Anna's past suffering can act as a tie-breaker, and provide a reason for preferential treatment. So, past suffering can provide a decisive reason for acting if present and future reasons run out. Let's amend *Disconnecting* to include this:

> *Disconnecting**: If an agent is future-biased, her reasons for preferring, choosing and acting now generated by past events, decisions and preferences will always be trumped by reasons generated on present and future events, decisions and preferences.

So, reasons generated from Anna's past are lower order reasons in comparison to reasons based on her present and future, which take lexical priority. If the higher order reasons run out, so to speak, Anna's past suffering generates reasons for acting, preferring and choosing.

However, *Disconnecting** will still undermine most whole life views, as this exception is limited to a tie in momentary well-being. If Anna's momentary suffering is lower than Bertha's, and Anna is future-biased, Bertha's momentary suffering generates

reason for treatment that outranks any reason based on Anna's past.

Hence, if agents whose lives we're looking at are future-biased, whole life views end up being undermined – if Anna is future-biased, then why should she consider her whole life? However, if Anna is temporally neutral, whole life views can still stand.

Let's consider McKerlie's case again: If we look at society in terms of equality, it seems that if just one person in that society were future-biased, the entire position-shifting society would not seem just anymore. For example, if at switch time, Antoinette was future-biased, she would immediately object to how society is arranged, as she has poverty in front of her, and her rich past doesn't matter anymore. The same would go for Bob: if he were future-biased (and cared about equality), he'd find society unjust, as Bob's past poverty wouldn't matter to him anymore, but currently being 100 times richer than Antoinette would appear unjust.

So, if the agents in question are future-biased, then Whole Life views seem to end up being undermined. The cases discussed are a bit simplistic, but it is not difficult to see how this can be applied to a range of egalitarian and prioritarian views that demand attention to a life as a whole. So, being future-biased does not only mean that Cillian Murphy can stop caring about his past preferences to become a rock star – it also means that we need to revise some of the more ambitious moral theories we have.

6.6 Do We Really Disconnect?

Most Whole Life egalitarians, and in general, moral philosophers who believe in lifetime well-being, will object to the argument above. One objection whole life egalitarians will pursue is to deny that future-bias really means disconnecting from all past parts of life. Future-bias, so the objection could go, in no way entails that we are disconnecting our entire past from our moral reasoning. So, just because I'd rather have my pains in the past rather than future doesn't mean that past pains can't generate reasons just as important as present or future reasons.

One way of pushing this objection comes from Dorsey, who rejects a "strong present and future-bias" and defends a more hybrid view of future-bias: according to him, it is rational for us to be biased towards present and future pleasures and pains, but when it comes to project-oriented goods, we ought to be temporally neutral.[208] So, with Dorsey, whole life views could have their cake and eat it – we can be future-biased, but still care about the past in a non-instrumental, morally significant way. Before we start, note that Dorsey presupposes that we in fact can benefit our past selves – something I do not agree with. But let's assume that this is true for the sake of the argument.

Dorsey follows a two-pronged strategy to show that a hybrid view is a better, more appealing explanation of future-bias than a disconnecting view. Firstly, he attacks the disconnecting view as intuitively implausible, as it makes it impossible for us to benefit

[208] Dorsey (2018), p. 1916.

our past selves. For example, if you have the mutually exclusive choice between completing a project your past selves spent a lot of effort and energy on but your current self has no reason at all to complete, and a "Tootsie Roll", benefitting your present self.[209] Faced with a choice between a very small present benefit and a very large past benefit, a future-biased person whose past reasons are trumped by present and future reasons would be rational to choose the Tootsie Roll over completing her project. But surely, says Dorsey, it's not very intuitive to say that it would be irrational for a future-biased person to choose a past benefit over a future one, or at least "someone who completed his (past) life's work would not be treated as imprudent". Therefore, the strong interpretation of future-bias should be rejected.

I admit that I simply do not share Dorsey's intuition here – if your present and future selves have no reason at all to complete the project, it does seem more prudent to me to go for the smaller present benefit. I also think that, if we apply this thinking to more life-changing cases like Cillian Murphy, most people would agree that Murphy shouldn't have chosen past benefits (becoming a Rockstar) over present and future ones, even if the present and future ones are smaller (getting a slight chance of being an actor). But Dorsey bolsters his argument with an interpersonal case:

> Albert and Joan: Albert is a physicist who worked for decades to explain a particular observed phenomenon

---

[209] Dorsey (2018), p. 1913. If you, like me, aren't US American, a Tootsie Roll is a mildly chocolate-flavoured taffy-like candy that has been manufactured in the United States since 1907, says Wikipedia.

> given standard physical theory. He made substantial progress, training several talented physicists along the way. Unfortunately, perhaps a year's worth of dedicated work away, he loses interest in his research. Instead, he takes immense pleasure from reading detective novels.
>
> Knowing his past dedication, a student of Albert's, Joan, takes up the problem and completes the work Albert had begun, using the tools and theoretical apparatus Albert had constructed during his lifetime. When asked why she had taken up his project, which was related, but not identical to her own, Joan responds: ''For Albert's sake.''[210]

Dorsey says that "it seems right to say that taking up this project is or could be done for Albert's own sake" or that "Joan plausibly benefits Albert."[211] So, if past benefits are trumped by present and future benefits, it would be wrong of Joan to complete Albert's project instead of getting him new crime novels. But surely, so Dorsey, this is not the case – we wouldn't criticise Joan for wanting to benefit past Albert so much and thereby forgoing some small present benefits for him.

Again, I don't share Dorsey's intuition on this case. Additionally, Dorsey introduces an interpersonal case to argue against an intrapersonal attitude – we've already discussed with Anna and Bertha that we shouldn't be future-biased on another person's behalf if the other person isn't future-biased. Dorsey says that,

---

[210] Dorsey (2018), p. 1906 and 1914. I put together two versions of Dorsey's case, and I'm skipping the interpersonal case with Jennifer from p. 1913.

[211] Dorsey (2018), p.1913.

even if that's the case, it would just be "jarringly implausible" if it would be prudentially wrong for me to benefit my past self over my present or future self, but that it wouldn't be wrong in terms of beneficence for someone else to benefit my past self over my present or future self.[212]

But it might not be the case that it's okay for someone else to benefit my past self if it's prudentially wrong for me to do so. If we really consider that Albert has no present and future reasons for completing his project, it seems weirdly patronising for Joan to insist on doing it for Albert's sake. If you yourself would decide to abandon a project you worked hard on, and someone else close to you would insist on continuing the project on your behalf, even if they had no reasons on their own to do so, and you explicitly expressed that you had no reason any longer to see the project completed, wouldn't that be disrespecting your change of mind when it comes to the project? So, someone else benefitting my past self, even if it's not prudent for myself to do so, might be wrong and patronising. In other words, why should someone else be temporally neutral on my behalf if I am future-biased?

Dorsey's second argument is more interesting. He argues that a strong future-bias held jointly with an attitude about cooperation with future-selves would be "normatively unsavoury". Consider

> Cooperation: to achieve a project-related good at time t, the success conditions of which occur at times later than

[212] Dorsey (2018), p.1914.

> *t*, requires cooperation between one's *t*-self and selves at times other than *t*.[213]

This principle states that different selves at different times need to commit to similar attitudes that lead to success of a project. The cooperative attitude, says Dorsey, fosters a relationship between different selves at different times that takes into account what the different selves decide, including recognising "the effort one's past self has put in, and cooperated for the sake of the success of the project."[214] While Dorsey remains somewhat vague about what this commitment precisely entails, we can stipulate that it at least means that my current self, when deciding and choosing, needs to take into account what my past selves had decided, and what my future self will decide. So even if Cillian Murphy's current self does not want to be a musician at all, he should take into account to some extend past efforts from his past selves to complete his project. His past self, after all, counted on his back-then future (and now present) self to be cooperative.

However, if Murphy is future-biased in the disconnecting way, we walk back on this commitment: his past self counted on his now current self's cooperation just as much as his current self is now counting on his future-self's cooperation. As every current self was a future self in the past, our past selves at some point counted on the cooperation of our future self. As our current selves, when planning projects, need the cooperation of our future selves, we shouldn't refuse this courtesy to our past selves. Or, as Dorsey

---

[213] Dorsey (2018), p. 1916.

[214] Dorsey (2018), p. 1918.

puts it, "you are refusing to grant your past self the same courtesy you're asking of your future self".[215]

I believe that there are several problems with Dorsey's argument, including that the Cooperation principle may be less obviously true than Dorsey thinks it is.[216] But what I would like to focus on is rather why this is supposed to be a problem for future-bias – why do I owe it to myself to grant my past self the same courtesy as my future self? Why is it "normatively unsavoury" for us to treat our past selves differently than our future selves? What Dorsey seems to be concerned about is hypocrisy in our behaviour when we take on projects as if our future selves were cooperative, but our current selves are not cooperating with our past selves. But to my eyes, there's nothing wrong or irrational about doing that – of course a future-biased person would not grant their past selves the same courtesy as their future selves, precisely because they are future-biased. Dorsey needs to provide a reason as to why this unsavouriness is bad, otherwise it's too easy for me to bite the bullet and accept that future-biased agents behave hypocritically.

One reason why we shouldn't embrace this unsavouriness might be that it leads to less projects being successfully undertaken –

---

[215] Dorsey (2018), p. 1921.

[216] Cooperation requires our current selves' attitudes to be dependent to some degree to what our past and future attitudes were or will be, making Cooperation a so-called diachronic norm. There is a lot of literature on diachronic norms about beliefs, and whether we ought to accept any of them given the substantial counterexamples advanced. (see Hedden 2015)

we count on our future selves to honour our past selves' commitment to ensure that a project we start gets carried over the finish line. And it might be the case that our strange behaviour of treating our future selves as if they won't be future-biased even if we are future-biased now will lead us to abandon more projects compared to us being temporally neutral. But that surely depends on what other principles and attitudes I follow to ensure my project commitments (e.g. I could adopt what Dorsey calls the "aggressive stance" and sanction future selves for not following projects), as well as how appealing my projects are over an extended time period. If I choose projects that stay appealing to my future selves, there is no need for my future selves to take into account my past selves' commitments. And if my projects do not appeal to my future selves, do I not owe them the right to reconsider their commitments?

So, Dorsey's two arguments fail. With this in mind, and the arguments in favour of a strong type of future-bias, we should treat future-bias as a comprehensive and lexically discounting preference pattern.

6.7 Conclusion

I have argued that we should understand future-bias in both a comprehensive way and with lexical discounting:

> Comprehensive Future-bias: An agent is future-biased iff she discounts her past parts of life against her present and future parts.
>
> *Disconnecting**: If an agent is future-biased, her reasons for preferring, choosing and acting now generated by past

> events, decisions and preferences will always be trumped by reasons generated on present and future events, decisions and preferences.

With this, we can see how it makes a difference whether Cillian Murphy is future-biased or temporally neutral: if Murphy is future-biased, his present and future reasons to act and choose will always trump his reasons from past preferences, and he should become an actor, not a musician. If Murphy is temporally neutral in a comprehensive way that includes all parts of life and not only pains and pleasure, Murphy will have to weigh his reasons from his past preference of becoming a Rockstar equally to his present and future reasons.

You might think that this comprehensive understanding of temporal neutrality and future-bias that isn't limited to pleasures and pains is much more radical and demanding than their minimalist versions. However, all authors in the debate defend a comprehensive version of temporal neutrality, which suggests that the appeal behind temporal neutrality runs deeper than mere balance of evaluation between past and future events. In other words, minimal temporal neutrality might be less demanding on agents, but isn't it arbitrary to limit temporal neutrality to just benefits and harms of events? Similarly, future-bias seems to have a deeper appeal than just pain being past, and should be

understood as comprehensive, as authors like Parfit at least implied in their works.[217]

So, in the end, we face a choice: if we understand temporal neutrality and future-bias as minimal, their rationality requirements might be easier for us to fulfil, but open to concerns of arbitrariness. However, if we understand both as comprehensive, they not only capture a broader intuitive appeal, but also make a difference in how we live our lives.

[217] See Parfit (1984) for his discussion on future-bias and temporal neutrality at the end of life.

7 Against Narrative Ethics

Boromir, Captain of Gondor, had a good start into his life. He was the firstborn son and heir, an outstanding warrior, and achieved glory early as a leader, taking back the lost city of Osgiliath from the enemy. Favoured by his father, Boromir was known for his strength, honour, his selflessness and love for his people. Later, however, he loses his hope and despairs after he fights battle after battle against the enemy, and in his desperate desire to protect his people, is unable to resist the influence of the ring of power. He betrays his vows, his friends, tries to steal the ring, realises that he has been corrupted, and dies fighting alone, thinking that he lost his honour.

Faramir, Captain of Gondor, had a bad start into his life. Always in the shadow of his elder brother, he favoured books and daydreaming over swords and horses, the "wizard's pupil" was disfavoured and neglected by his father. Known for his modesty and fair-mindedness, not for his love of warfare, he lost several military campaigns, losing Osgiliath to the enemy, almost leading to his death. However, he later manages to recognise the ring of power's corrupting influence and turns out to be the only human able to resist the ring. After the war, he recovers and is honoured for his wisdom and bravery, and finds peace.

Assume that both brothers had roughly the same number and quality of good days and bad days, pleasures and pains, achievements and failures in their lives: who lived a better life?

Boromir, whose life progressed from high to low, or Faramir, who started out low but ended high?

As a second question, let's focus on the ring of power: Is that what makes us care about the fact that Boromir became corrupted while Faramir resisted? Is it merely the consequences of their actions (breaking the fellowship, or allowing the ring to be destroyed, respectively) that make these events seem significant to us? Or does Boromir's fall matter to us beyond the pain and damage he causes, and Faramir's achievement more than a good outcome?

7.1 Introduction

One answer for both questions defended by many philosophers is that it's the narrative structure in both brothers' life that gives meaning to these events. It is because their life story adds a specific, particular quality, making a bad event tragic, or a good event heroic. Boromir's failure becomes intelligible to us because of his life story, and Faramir's resistance becomes an achievement because his life was structured in a way that told a story how he arrived there. Because of the narrative structure within both brothers' lives, we recognize Faramir's life as better than Boromir's, and recognize the moral depth of their achievements and failures.

Recently, philosophers such as Dorsey (2015), Rosati (2013), Glasgow (2013), Kauppinnen (2012), Raibley (2012) Portmore (2007), and less recently Velleman (1991) and MacIntyre (1981), have defended this popular explanation in moral theory: the narrative structure of a person's life matters as such. How a

person's life is structured, how different parts of a person's life relate and build to something meaningful matters, determining a life's value beyond momentary well-being.

Typically, the narrativity thesis in ethics states that there are meaningful, irreducible relations between life events that tell us something about the person's life story. Roughly, defenders of the narrativity thesis usually accept three claims[218]:

> Relationism: The value of a person's life does not only depend on the momentary value of events or parts of the life, but also on the value-affecting relations among its parts over time.
>
> Narrativity: The relevant value-affecting relations among its parts over time are narrative relations.
>
> Irreducibility: The narrative relations between parts of life over time are not reducible to other factors of a life's value.

From this, defenders of the narrativity thesis argue that stories matter for a life's value – so if I want to live a good life, I should try to live in a way that has meaningful relations between different parts of my life. Other things being equal, I have reasons to live my life as a story.

In this chapter, I will argue that this popular view is mistaken. The narrativity thesis is false because we should reject *narrativity*. After outlining the narrativity thesis in more detail,

---

[218] Rosati (2013), pp. 29-30.

I will argue that, if future-bias is rationally permissible, it provides an undercutting defeater to narrative relations between different parts of life, as my reasons for caring about those will disappear. If we are rationally permitted to disconnect from a past event, a relation between this past event and a future event will not generate an irreducible value to my life that provides me with a reason to make decisions about my life. In other words, if I can discount the normative significance of my past, I don't have to live my life as a story.

I further explain that future-bias offers an explanation as to why we think that the narrative shape of a life matters, as our preference for good things being ahead of us and bad things being past often leads us to think that our life ought to be shaped in a certain way. Hence, we can reject narrativity without losing its explanatory value when it comes to explaining why shapes of lives matter.

Finally, I will argue that future-bias as a preference pattern can provide us with a safeguarding rationale to avoid narrative fallacies – we often overestimate our understanding of the past, and together with our tendency for creating a story out of our life events, this can lead us to construct misleading narratives about ourselves. Future-bias can caution us against committing this fallacy.

In summary, we don't need any assumptions about a person's narrative self to explain a life's value, and we don't have to live our lives as a story. The narrative self doesn't matter, and the narrativity thesis is false.

7.2 The Narrativity Thesis

Let's first spell out the Narrativity Thesis. To do so, it might be helpful to outline first what defenders of it originally set out to deny:

> Additivity Thesis: The sum of momentary well-being sufficiently determines the value of a life.

According to this view, if we add up all the momentary good and bad experiences Boromir had, we would be able to determine how good or bad his life was as a whole. We'd know exactly why he suffered from betraying the fellowship, and we'd know exactly why being corrupted by the ring was bad for him. This view is typically held by Hedonists, but it is not exclusive to them – any theory of well-being that holds that there is nothing valuable beyond a person's well-being at particular moments.

Narrative ethics denies this – there is something valuable beyond momentary well-being, and the value of a life cannot be reduced to the sum of its valuable moments. This "something" that matters is the narrative ordering or structure of life. As described by Dorsey:[219]

> Shape of Life Hypothesis (SLH): The temporal sequence of good and bad times in a life can be a valuable feature of that life as a whole.

[219] Dorsey (2015), p. 305, also Velleman (1991), p. 50.

To motivate SLH, consider Dorsey's tales of O.J. Simpson and J.O. Nopmis, which are basically like Boromir and Faramir, but a bit more precise:[220]

> O.J. Simpson: O. J. Simpson was a celebrated college and professional football running back, film actor and producer, and sports commentator. In the midst of his success, Simpson was put on trial for murder. And though he was acquitted after a lengthy trial, many were convinced of his guilt, and his reputation had been ruined. Following his acquittal, he was held civilly liable for wrongful death in the same event and was later convicted of burglary, was sentenced to thirty-three years in prison, and is currently serving his sentence.
>
> J. O. Nospmis: J. O. Nospmis grew up midst gang-related violence and crime, was suspected at an early age of murder, and was sentenced at age twenty-five for armed robberies. Following her stint in prison, Nospmis was released and given an opportunity to coach basketball at a club for troubled youth. Her success at this endeavour, along with her rapport with players and amazing life turnaround brought her to the attention of schools, later universities. She retired after having coached her team to back-to-back NCAA Final Four appearances, and spent her remaining years as a popular and trusted broadcaster,

[220] Shortened from Dorsey (2015), pp. 304-305.

> offering insightful colour commentary on professional and college basketball.

Assuming that both lives have the same sum of momentary well-being, how does O.J Simpson compare to J.O Nospmis? If the Additivity Thesis is true, then both have lived a life equally good – there is no difference in the value of both lives, since there is no difference in momentary well-being. However, this sounds implausible – many people, so Dorsey, think that O.J Simpson is worse off even if he momentarily experienced the same as J.O Nospmis. Simpson's life is lacking something beyond Nospmis' life: an upwards trajectory, an upswing. The fact that Nospmis' life events are ordered in an upwards trajectory makes her life better than Simpson's, whose life contains a downwards spiral. And since both facts are beyond momentary well-being, the temporal sequence of life events matter – SLH is true, and the Additivity Thesis is false.

Why would the temporal sequence of life events matter? Stated by Velleman[221], and taken by Dorsey as most plausible, is the explanation that how events in life relate to each other affect how valuable they are. The relation between events or parts of life at least partly determine the value of these events for a person's life as a whole. We now arrive at the first component of the narrativity thesis mentioned above:

> Relationism: The value of a person's life does not only depend on the momentary value of events or parts of the

[221] Velleman (1991), p. 53.

life, but also on the value-affecting relations among its parts over time.

Note that this goes beyond what SLH says. It is possible to deny Relationism and hold onto SLH. For example, you might think that the shape of a life just matters intrinsically, or that we care about gains and losses – what makes Nospmis' life better than Simpson's is that she gains and he loses, explaining the value difference between both without invoking relations between different parts of life.[222]

However, *Relationism* states that what explains the value of the shape of a life is the relations between life events. Without looking at how the events in life relate to each other, we cannot determine what they mean for life as a whole.[223] To know how an event contributed to a life's value, we need to understand what the event means, or what role it plays in relation to other events in life.[224] If evaluated in isolation, Boromir being corrupted by the ring would not reveal the depth of the situation. Only by looking at Boromir's life story of being an honourable defender of his people do we understand the tragic quality of his failure to resist the ring. The way his life was structured leading to the event makes the corruption worse than it would have been if it were just a random, isolated event, as it isn't vices like wickedness or selfishness that leads to Boromir's failure, but precisely his virtues that make him succumb to the ring. Without the relational

---

[222] See Glasgow (2013).

[223] Velleman (1991), p. 53.

[224] MacIntyre (1981), pp. 211-212.

structure, we could not understand the actual value of the event. Hence, life events are only intelligible through them "fitting into the story", or in other words, through their relation to other events.

Now, why is it that the relation between life events determine the value, or even the intelligibility of an event in life? Following Velleman, Dorsey and MacIntyre, the relations between life events affect their value because they unify the agent's life to a narrative story.[225] We arrive now at the second component of the narrativity thesis – the relation between life events is a narrative one:

> Narrativity: The relevant value-affecting relations among a life's parts over time are narrative relations.

What "Narrative" exactly means is spelled out slightly differently by different authors in the debate. For example, authors disagree whether narrativity possesses final value or is simply contributory towards the value of a life.[226] The common core however is that a narrative relation between life events is a form of coherence between earlier and later activities that give rise to "meaningfulness".[227] Two life events stand in a narrative relation to each other if the latter coheres meaningfully with the former. "Meaningful coherence" is cashed out in terms of long-term goals

---

[225] Dorsey (2015), pp. 312-313, Velleman (1991), p. 53, MacIntyre (1981), p. 218.

[226] Velleman thinks the first, Kauppinnen the second.

[227] Velleman (1991), pp. 59-60, Dorsey (2015), p. 313, Kauppinnen (2012), p. 368, Rosati (2013), p. 34.

or projects – an agent is narratively unified if her life events are connected within long-term projects in a way that former activities of an agent positively informing the latter activities, so the agent appropriately feels fulfilment in her achievements if her long-term goals or projects are completed.

So, for Boromir's life to be narratively unified is for his former activities (being honourable and brave, fighting for his people, going to Rivendell etc.) to positively inform his later activities in terms of his overall long-term project (protecting the people of Gondor). Because of this long-term goal, the event of being corrupted by the ring gains a tragic meaning: They disrupt the coherence between his earlier activities and his later actions, so that he "loses meaning" when the events happen. His corruption by the ring is based on his overarching goal to protect his people, but do not fit his earlier activities, and therefore throw him off-balance, because breaking his oath and betraying his friends is not something an honourable captain of Gondor does.

Finally, we arrive at the final part of the narrativity thesis:

> Irreducibility: The narrative relations between parts of a life over time are not reducible to other factors of a life's value.

This means that there is something about narrative relations that is "over and above" the momentary factors in a life that contributes to a life's value. Velleman describes the contribution of narrative relations as a second-order good that cannot be explained in terms of momentary well-being because narrative relations contribute on a diachronic dimension, while momentary

well-being is synchronic.[228] His argument for irreducibility is that if the contribution of narrative relations could be explained in terms of momentary well-being, then learning from a misfortune in life would just be as learning the lesson from some other source.[229] Since the lesson from personal tragedy would only add value to the sum of momentary well-being, a book that teaches the same lesson would just be as good.

Imagine that Faramir, going through humiliation by his father and being overshadowed by his brother, as a result reflects on the true meaning of nobility and honour, comes to the conclusion that being a true knight is not something determined by use of force, helping him to resist the ring. Now compare this to Faramir reading a book with the same lesson or listening to Gandalf telling him that the ring is treacherous and to be resisted. According to Velleman, the second case cannot be equal to the first – in the second case, the book may help Faramir take the same decision and arrive at the same outcome of momentary well-being (the ring is destroyed and Sauron defeated), but in the first, the Faramir's action gains meaning. By learning from his own story, Faramir would gain a different value to his life than merely adding momentary well-being from another source. The information must come from his own past, his own narrative. Therefore, narrative relations contribute over and above momentary well-being, and are not reducible to them.

---

[228] Velleman (1991), p. 60.

[229] Velleman (1991), p. 54.

To shortly summarise so far, defenders of the narrativity thesis accept three claims[230]:

> Relationism: The value of a person's life does not only depend on the momentary value of events or parts of the life, but also on the value-affecting relations among its parts over time.
>
> Narrativity: The relevant value-affecting relations among parts of life over time are narrative relations.
>
> Irreducibility: The narrative relations between parts of life over time are not reducible to other factors of a life's value.

Through these three components, the narrativity thesis explains why the shape of a life matters – because there are irreducible, narrative relations between parts of a life that give meaning to life as a whole.

In what follows, I argue that the narrativity thesis, more specifically *Relationism* and *Narrativity*, should be rejected. I'd like to note that I do not aim to defend the additivity thesis – the value of a life may very well be determined by something beyond momentary value. Nor do I want to reject the shape of life hypothesis – on the contrary, I will try to offer an alternative explanation as to why we think the shape of a life matters. What I aim for is to show that we should reject the view that our past

[230] Rosati (2013), pp. 29-30.

activities meaningfully inform our future non-instrumentally, because we are permitted to disconnect from our past.

7.3 Caring About Our Past

Most people prefer bad things to be over. If you think back to a painful appointment at the dentist or to an awful social event, you are glad that it's over and done with. At the same time, you'd like good things to be ahead of you. You'd prefer your vacation or a good concert to be in your future rather than in your past, already gone. If you have this sort of preference, you are future-biased.

If a person is future-biased, she prefers bad things to be past and good things to be future. This preference seems to be quite natural – so natural that most would think it to be permissible to have this attitude. The permissibility of future-bias is classically motivated by Parfit's Past and Future Operations case, where an awakening patient with no recollection of what happened is wondering whether they would prefer to have had their very painful operation yesterday already, or a milder, less painful one tomorrow.[231]

The case shows that most of us would prefer bad things like pain to be past rather than future; even if the past pains are worse than the future pains. So, considering Parfit's operations-case, future-bias seems so natural that it ought to be permissible. Dorsey[232] states that it's so natural that we care more about

[231] Shortened from Parfit (1984), pp. 165-166.

[232] Dorsey (2016), p. 5.

events ahead of us, that we want bad things to be in the past that future-bias may be considered a brute feature of our practical rationality. So, anyone who would want to show that future-bias is rationally impermissible, and that we ought to be temporally neutral, would have to carry a quite heavy burden of proof, since the attitude's permissibility is so intuitively appealing.

As discussed in chapter 2, you could reply here that future-bias is, if rational at all, only permitted when it comes to hedonic goods, but not when it comes to non-hedonic goods. As Brink and Hurka have pointed out, we do not find it intuitively plausible for future-bias to be rational in cases of disgraces or achievements.[233] However, it isn't difficult to imagine a case where I would prefer a smaller, future achievement to a bigger, past achievement: Imagine that I'm a successful football player in the midst of my career – I wake up one night, hungover, forgetting my players records momentarily. When I look myself up, I can see two football player profiles, one that one the European championship five years back but won't win any more, and one that hasn't won any championships yet, but will win the Scottish premier league in five years. Which one would I prefer to be?

While it may be permissible for me to be temporally neutral here, it also isn't unreasonable to assume that a football player in the midst of their career who cares about playing football well, would prefer the smaller, future achievement over the bigger, past one, especially if we exclude instrumental benefits such as money, recognition, and focus on the achievement as such. A "doer", as

[233] See Brink (2010) and Hurka (1993).

Hurka puts it, wouldn't be unreasonable to be future-biased when it comes to achievements.

Additionally, I would like to add that temporal neutrality, the requirement opposing the rationality of any time-biases, is usually phrased in quite a comprehensive way by authors like Brink, Dougherty and Greene and Sullivan:

> Do not prefer an event over another only because of its temporal location and give equal concern to all parts of your life.[234]

Temporal neutrality is about all our parts of life, not only the hedonic parts: if you're temporally neutral about benefits and harms of events, why not be temporally neutral about other aspects of your life too? It would be arbitrary just to single out pain and pleasure to be temporally neutral about, so why not be temporally neutral about all parts of life? In a way, temporal neutrality about benefits and harms of events is part of a broader principle of comprehensive neutrality. Similarly, for future-bias: If we're future-biased about benefits and harms of events, why not appeal to a broader version of future-bias that encompasses preferences and choices too? This way, we have a more systematic and less arbitrary way of being future-biased.

In absence of a convincing counterargument, I will hence assume the following:

(1) Future-Bias is rationally permissible.

---

[234] See Brink (2010), Greene and Sullivan (2015), Dougherty (2010).

(2) Future-bias applies to all parts of past life, including events, preferences, and reasons.

Before I continue my argument, I need to outline a structural feature of future-bias. Another way of describing future-bias is as a discount function – the value of an event is decreased if it's past. Sullivan argues that there is reason to think that future-bias is not an exponential or hyperbolic discount function, where the value of an event would decrease the further it goes into the past, but *absolute*.

If future-bias is absolute, we assign no or almost no value to a past event – as soon as it is past, it's just gone. If future-bias weren't absolute, we'd have a function with three variables:

(1) The temporal distance between now and the past event.
(2) The value I'd assign to the event if it were present.
(3) A discount-function.

So, let's say that I have a discount rate such that every 7 days, the value of an event halves. My painful operation I just had is a 10/10 now on the pain-scale, but after a week, I'd rate it 5/10, after two weeks, 2,5/10 and so on. In this case, my future-bias could be represented as an exponential discount function. If future-bias would be a non-absolute discount function, my evaluation of the past event would be sensitive to the temporal distance to the past event. I'd map the value of the past event to how much time has passed. However as Parfit's Past and Future Operation's case shows, I do not. I just care about the past event being past. It would be absurd to hope for a future operation instead of a past one, regardless of how far away the past one is.

As soon as an operation is past, I will always prefer it to a future operation. Hence, future-bias is absolute.[235]

However, as we have discovered in the last chapter, there is several problems with that. Firstly, if we look at empirical data, it might not actually be the case that we discount absolutely, even if we don't discount hyperbolically or exponentially. As Fernandes helpfully summarises the empirical findings, people usually do give past events a significant amount of value, compared to future events, as well as only discounting past pain to a certain point: while 92-93% prefer past pain to equal future pain, and future pleasure to equal past pleasure, 53-54% switch their preference if the amount of past pain or pleasure is doubled. At the same time, people report to be indifferent between 5.6 hours of past pain and 2 hours of future pain. This seems to show that the amount of past pain is still significant.[236]

This suggests that when future-biased, agents don't just reduce the value of the past – they can still assign value to it, and sometimes in a way that does not resemble absolute discounting. Future-biased agents rather care differently about their past:

> *Disconnecting*: If an agent is future-biased, her past events, decisions and preferences cease to generate reasons for preferring, choosing and acting now.

---

[235] Sullivan (2018), pp. 49-50.

[236] As summarized by Fernandes (2019), p. 7. For the full studies, see Caruso et al. (2008), Caruso et al. (2018), Lee et all (2018).

If we are future-biased, our past events just cease to be reason-generating: our past pain does not provide us with any reasons for preferring or acting anything anymore, even if we still assign a pretty high value to it. The future pain, however, does give us reason to prefer it not to be the case, and reasons to try and avoid that pain – even if that pain might be lower than the past pain.

Disconnecting also works better than *Absolute Discounting* when it comes to non-hedonic goods and harms: if you look at a past big achievement in your life and compare it to a smaller achievement in the future, it might not necessarily make sense to speak of discounting your past achievement if you're asked which one you'd prefer. It might not be the case that the past achievement is worth almost zero to you. It might not even be the case that your past achievement is worth less to you now. The thing is just that you *care differently* about your past than about your future: your past achievement now doesn't generate a reason for preference like a future achievement would, even if in the past, your achievement did provide you with a reason for preference. Another example would be past friendships vs future friendships: it might not be that you discount your past, it simply ceases to be reason-generating for you. Hence, given that future-bias applies to non-hedonic cases, we are better served with *Disconnect* rather than *Absolute Discounting*.

And finally, Sullivan herself, one of the main proponents of temporal neutrality, also thinks so. In her book, she describes temporally neutral agents as being different from time-biased (read future-biased) agents: temporally neutral agents take their past events, preferences and choices into account in making

decisions, past preferences and choices can license present and future choices just as much as anticipated future preferences and choices can.[237] Hence, the difference between a temporally neutral agent and a future-biased agent is that the former does not disconnect from past events, preferences and choices, while future-biased agents do. As Sullivan treats temporal neutrality as a way of being connected to your past choices and preferences, future-bias should be treated correspondingly, as *Disconnecting*. Past preferences and choices are not taken into account for current and future considerations if I am future-biased.[238]

So, to summarize, I will assume the following going forward:

(1) *Future-bias consists in disconnecting from one's past*: If an agent is future-biased, her past events, decisions and preferences cease to generate reasons for preferring, choosing and acting now.

A past event can still hold instrumental value for a future-biased agent: you can learn from the past event, and you can still prefer a past event over another past event, but it won't be relevant to you in an actual preference situation with future events in the way that it in itself doesn't provide you with a reason in favour of any options.

In what follows, I will show that this leads to future-bias undercutting the narrativity component of the narrativity thesis,

[237] Sullivan (2018), p. 151.

[238] Sullivan (2018), p. 131.

and since future-bias is rationally permissible, we ought to reject the narrativity thesis as a result.

7.4 Against Narrativity

Recall Narrativity as a component of the narrativity thesis:

> Narrativity: The relevant value-affecting relations among parts of life over time are narrative relations.

My argument proceeds as follows:

(1) Future-bias applies to all parts of past life, including events, preferences, and reasons.

(2) *Future-bias consists in disconnecting from one's past*: If an agent is future-biased, her past events, decisions and preferences cease to generate reasons for preferring, choosing and acting now.

(3) For narrative relations between past and present parts of life to exist, our past needs to be reason-generating for present and future decision-making. (From Narrativity)

(4) If an agent is future-biased, she will disconnect from her past in a way that disrupts narrative relations between her past and present selves. (From (2) and (3))

I have established (1) and (2) above, and while we need to do some more footwork to establish (3), the premise should look plausible at first glance from the narrativity thesis and the necessary historical element in narrative relations between past and future parts of life. If we get this far, and given that future-bias is rationally permissible, we have established that future-bias

undercuts narrativity, as future-bias removes the basis of narrative relations to be value-affecting. So, why (3)?

If we accept Narrativity, then there is something beyond momentary value making my life good or bad: My life will be better the more parts of my life cohere meaningfully with other parts of my life. When pursuing actions and activities, I should not only care about my present and future well-being but link my actions up with my past activities so that my actions fit into my story. This way, I will increase the value of my life by giving it meaning – in acting as a narratively unified agent, I will make my life better overall, independently of my momentary well-being.

In other words, my past actions, activities, and preferences should give me reasons for acting now and in the future. If I wanted to become a pianist in the past, and took steps to achieve this goal, these past preferences and activities give me a reason now to go to the conservatorium, even if I know I prefer to study philosophy. So, as a narratively unified agent, my past gives me reasons to act now and in the future.

However, if future-bias is rationally permissible, I am allowed to disconnect from my past. In acting and choosing, I may render them not reason giving for future choices anymore, since future-bias consists disconnecting from my past parts of life. This means that past events don't influence my choices for my future activity beyond instrumental learning, once discounted. And as we established already, narrative relations are essentially historic, as

MacIntyre argues – one cannot make sense of narrative unity without looking at the past.[239]

Recall Boromir: if Boromir is allowed to disconnect from his past, then he may always prefer positive events to be in his future, and negative events to be in his past, regardless of the corresponding event. For example, his past activities as a defender of his people, retaking Osgiliath for Gondor, can be disconnected from his present self, as well as him protecting his little brother from their father. But if these events do not generate any reason for choice anymore, why should they matter in his life story? Why aim for coherence with actions that are not reason-giving anymore? The reason why he'd aim for meaningful coherence between past and future activities was the meaning in his past activities in the first place – if his past is discounted, the reason for him to construct a relation disappears. Hence, *Narrativity* is undercut by future-bias.

So, if I'm future-biased, my past won't positively/negatively inform my future choices in a non-instrumental way. My past can still inform my future-choice instrumentally, e.g. what Faramir learned from being a bookworm in the past can be good use for what he'll do in the future. But his past can be discounted in the sense that he doesn't have to make future choices dependent on what he did before. His activity as a young daydreamer not training with the sword doesn't matter as such - what he can learn about his past for his future choices, he could also learn from a book, a trusted friend, or some other sources. But with a

[239] MacIntyre (1981), pp. 221-222.

narrative relation between past and future, this would not be possible, it *must* be his past that informs his future. While I can still inform myself from my past, my reason to link my past activities to my future is undercut.

In summary, if an agent is future-biased, narrative relations in their life are undercut by the disconnect from their past – you cannot be future-biased while also trying to live your life as a story. To this point, you might say: if this is true, worse for defenders of future-bias – the narrativity thesis is more plausible and more important than future-bias, therefore we should reject future-bias's rationality to preserve narrative ethics. As I already mentioned, we went into this argument assuming that future-bias is rational, so we can appeal to an argument like

(1) Future-Bias is rationally permissible.

(2) If a rational agent is future-biased, she is rationally permitted to not act in a way that establishes narrative relations between her past and present parts of life.

(3) If an agent is rationally permitted to not act in a way that establishes narrative relations between her past and present parts of life, narrative relations between past and present parts of life do not bear intrinsic value.

(4) Narrative relations between past and present parts of life do not bear intrinsic value.

to reject *Narrativity*. Why (3)? Why does not acting to establish narrative relations being rational mean that these narrative relations are not value-bearers? We might appeal to the fact that

most people want their lives to be as good as possible, following a principle like

> Other things being equal, if I have the option between increasing my lifetime well-being and not increasing it, I should choose to increase it.

In other words, there should be some kind of connection between lifetime well-being and what we ought to do. If that is the case, and I am rationally allowed to not act in a way that establishes narrative unity in my life, that means (other things being equal) that narrative unity does not contribute to my lifetime well-being. Therefore, the narrativity thesis is false, and we don't have to live our life as a story.

7.5 Objection: Narrativity and Temporal Neutrality

To recap, let's assume that the narrativity thesis is correct: a life lived as a story is better than a life without. Then, other things being equal, you ought to live your life as a story. So, you should make decisions in a way that contribute to your life being a story, make choices that are meaningful to you. However, to make such choices, you need your past to contribute to your decision, in a non-instrumental, irreducible way. To make your career choice meaningful, you need your past events and activities to inform your choice a career adviser couldn't. That means that there should be a relation between your past and your future activities in a way that tells a story. But if you can disconnect from your past, make it irrelevant to your choices, how can you tell a story

about that? Shouldn't you, if the narrativity thesis is correct, not be future-biased?

If that is correct, being future-biased and living your life as a story can come into conflict, as your past preferences need to be reason giving for your future choices to have a narrative relation to them. And as future-bias is rationally permissible, we should reject the narrativity thesis. However, instead of that, proponents of the narrativity thesis could appeal to temporal neutrality, a requirement not to be time-biased and to give equal concern to all parts of life.

And indeed, Brink explains that temporal neutrality and a concern for narrative structure are compatible, and if there is value added by living your life as a story, even required by temporal neutrality due to concern for all parts of life.[240] However, that does not mean that the plausibility of temporal neutrality automatically lends support to the narrativity thesis. For example, Sullivan expresses scepticism, and outlines in her book that even given temporally neutral agents, commitments to narratively structured lives can lead to conflict with prudence:

> Golden Years: Frank is approaching retirement. He has spent most of his career working on a book, and if he threw himself into the project, he could finish it. But he gets little pleasure from the project anymore, and he thinks he would be happier if he abandoned the book and spent his golden years playing with his grandchildren. Suppose that

[240] Brink (2010), p. 4.

> a life where he finishes the book has more narrative value than one where he spends the rest of his life playing children's games.[241]

Should Frank finish his book, to live a life with better narrative value? Or should Frank abandon the project, leading to more happiness? In this case, being temporally neutral doesn't automatically favour the narrativist point of view: if Frank is temporally neutral, it could still be better for him overall to abandon his project.

What does this show us? That temporal neutrality per se does not support narrativity, it only supports it insofar it rules out future-bias as irrational. However, as I have explained, intuitively, future-bias seems rationally permissible, and to appeal to temporal neutrality just on the basis that it opposes the rationality of future-bias without any further connection to narrativity seems dialectically cheap.

7.6 Future-Bias and Shapes of Life

So far, I have argued that the key component of the narrativity thesis, namely that there are narrative relations between parts of life, should be rejected. This may be hard to accept for those who find the narrativity thesis intuitively appealing or have independent reasons for narrative unity of agents. However, in absence of an independent reason, I believe the appeal of the permissibility of future-bias to be sufficient to yield the conclusion. In what follows, I will add an explanation to show

---

[241] Sullivan (2018), p. 31. She expresses further scepticism on pp. 144-148.

why we feel drawn to the narrativity thesis, but that we don't need it to explain why we think that a shape of a life matters.

One aim of the narrativity thesis is, as outlined in the beginning, to explain SLH:[242]

> Shape of Life Hypothesis (SLH): The temporal sequence of good and bad times in a life can be a valuable feature of that life as a whole.

However, Future-bias offers an alternative explanation for why a life's shape matters, so that we can make do without the narrativity thesis and still hold on to SLH.

One of the alternative explanations for SLH Dorsey describes and then dismisses is the *"Later-is-Better"* view: We simply prefer goods to be later in life – the later a good occurs, the more it is worth, and vice versa for harms.[243] The reason why Simpson's life is worse than Nospmis' is that goods occur later and harms earlier in her life, while Simpson has harms later and goods at the beginning.

That's a time-preference which is roughly corresponding with future-bias. The earlier/later perspective is a non-perspectival description of the life that doesn't fit the narrative point of view – if we look at a life from the perspective of the person in the story, an agent always understands herself as being "within" the story, according to Rosati.[244] From that internal point of view,

---

[242] Dorsey (2015), p. 305, also Velleman (1991), p. 50.

[243] Dorsey (2015), p. 314.

[244] Rosati (2013)

the agent has a past, present and future. Of course, not all earlier-later relations translate into past-future relations. For example, there can still be two future events, expressed by earlier-later descriptions. In these cases, the "later-is-better" view seems implausible anyways, as Dorsey argues in his weekend cases:[245]

> The Lost Weekend. On Friday, I went over to a friend's house to watch a Friday Night Football game, had a great time, but drank rather too much. As a result, I was feeling very bad on Saturday, and recovered only slightly on Sunday.
>
> The Found Weekend. I drank rather too much on Thursday night. As a consequence, was feeling very bad Friday, recovered only slightly on Saturday, but was feeling fine on Sunday, when I went over to a friend's house to watch Sunday Night Football, and had a great time.

According to the "later-is-better" view, the found weekend must be better than the lost weekend, since goods occur later, and harms earlier on the former. Dorsey thinks that in cases like this, this doesn't seem to be the case. Although I do have the intuition that, other things being equal, the found weekend is actually better, let's grant that argument to Dorsey, and try to reform the "later-is-better" view. Let's replace it with *future-bias*. For cases where I am in between two events, "*later-is-better*" is future-bias. If it is Saturday, I will, if I am future-biased, prefer the found weekend to the lost weekend, simply because I'd rather have a

[245] Dorsey (2015), p. 316.

great time ahead of me rather than a hangover. However, before the weekend, as well as after, I would be indifferent between both – which I take to be the best explanation of the weekend cases.

The same goes for the Simpson/Nospmis case – future-bias explains what we think about them. At birth and at death, I would be indifferent between both lives. But in between, within the life of one of them, I'd clearly say that Nospmis has the better fate, because she has so much good ahead of her, and her misfortunes are past, while Simpson's fortune is over and gone already.

So, from the viewpoint of the agent, we can replace the "later-is-better" view with future-bias. Therefore, future-bias not only shows that we should reject *narrativity*, but also offers an alternative explanation for SLH. We think that the shape of a life matters, because we care about what is future, and what is past. This does not completely explain or capture narrativity's appeal, as there are going to be cases where narrativity provides a different judgement on whether the shape of a life matters than future-bias. For example, a case where a person suffers or achieves something in the mid-point of their life, which could be fully explained by meaningful narrative relations, could not be captured in the same way by pointing to a later-is-better view. But we can limit some of the loss of explanatory power resulting from rejecting narrativity somewhat and preserve some of SLH's appeal.

So, how do we explain the depth of Boromir's failure, and Faramir's achievement then? We can appeal to Boromir's good

days being in the past, and his despair for his people in the future, and we can outline Faramir's misery being behind him, and his achievements in front of him. This might not completely replace a narrative account, but at least partly explains what is going on in their story.

7.7 An Insurance against Narrative Fallacies

I have argued that, if future-bias is rationally permissible, then we ought to reject the narrativity thesis. I have also argued that future-bias offers an alternative explanation for why we think that the shape of a life sometimes matters. Some will think that, given the deep role narratives play in our lives, if my argument goes through, the worse for future-bias – the incompatibility just provides an argument against that time-preference. However, I do believe that a preference for the future is as embedded in our thinking as our tendency towards narrative storytelling, and therefore hope to have at least challenged the idea that our stories matter for the value of our lives.

But for those who still think that living a story is more important, I would like to close the chapter with some thoughts about storytelling. Consider the following case:

> You're on a job panel, and two candidates are shortlisted who are equally qualified and distinguished for the job. Candidate A's CV, however, tells a story: where they came from, how it fits with their direction in life. Candidate B's CV does not. Does the fact that you think that candidate A has a better life story make A a better choice than B?

I believe this to be a realistic example, and one that affects a fair amount of people. If we accept something like the narrativity thesis, we will be inclined to lean towards candidate A – even if B is just as good, or even better. Fractured lives look worse to us, and a life that tells a story looks better – even if the fractured life is just as good, or better, or if A's life isn't as coherent as it seems.

Daniel Kahneman describes something called a "narrative fallacy":[246] because we like to make sense of events, we tend to construct stories and narratives about our past, in our drive for coherency and explanation. When we hear about a set of events, we immediately try to set them into relation, to construct a story, even if the available information is very limited. Interestingly, so Kahneman, the less you know about something, the easier it is to build a coherent story about it. "The core of the illusion", he says, "is that we believe we understand the past [...] but in fact we understand the past less than we believe we do."

The narrative fallacy is treated as distinct from the so-called sunk-cost fallacy, where agents commit to honouring past investments by committing further resources to it, even if the outcome isn't desirable anymore. For example, if I buy a ticket to a Lana Del Rey concert, and between the purchase and the concert change my mind about Lana Del Rey and don't want to see her anymore, it would be a sunk-cost fallacy if I still go just because I bought the ticket.

---

[246] Kahneman (2011), pp. 201-202.

In the case of narrative fallacies, we look at a similar choice structure: a past decision or activity determining our present choice. However, with a narrative fallacy, my decision would not be based on my sunk investment, but on me trying to make sense of why what I did matches my identity, or my future choices.

Take an example: Suppose I construe myself as a very egalitarian person, who cares about oppressed and marginalised groups, and who tries to actively contribute towards more justice, in terms of gender, race and class. When I now learn more facts about my past, e.g. that I might have behaved inappropriately towards members of the opposite gender, that I was brought up in a certain social class, instilling me with biases, then it gets more difficult for me to construct a unifying narrative of me as an egalitarian. When I learn this, I can either form a complex narrative that takes into account my learning process – or I can shut out the parts of my past that do not fit my current narrative as a good feminist. For many reasons, the latter is easier to achieve, and this is what makes narrative fallacies not only misleading, but actively dangerous for our learning process.

The same can be said on a collective level: The United Kingdom has had a rather glorious past, with an empire where the sun never set. The British empire was a pioneer in exploration, trade and military, and was one of the greatest sea powers the world has ever seen. Nowadays, the United Kingdom takes this past to inform itself for its future activities – keeping and deepening ties with the Commonwealth, to contribute further towards a free trading world, using its knowledge from the past to continue to

drive the Commonwealth forward in terms of trade and cooperation.

This may sound very compelling and is a narrative that is not uncommon in the political landscape of the United Kingdom. However – regardless of what you think of this story in particular - the more facts we learn about the past of the British Empire, and the more we learn about what the Commonwealth nowadays actually is, how it functions, and how attitudes between members of the Commonwealth are, the more difficult it will be to tell a unifying narrative.[247] Whether you learn details of colonialism, wars for independence from the empire, or whether you learn about today's diminished role of the United Kingdom, the more facts you learn, the more difficult it is to actually tell the story – and many people will not opt for constructing a complex narrative, but for a simple one not containing difficult past events.

In summary, not only does narrativity not track the truth about our past – it seems that narrativity might actively contribute to a false understanding of who we were, since it's easier for us to accept a simpler, more meaningful story – even if the narrative is not accurate. And both on a personal and collective level, narrative fallacies risk embedding or creating self-identities that are at best misleading, and at worst oppressive.

Of course, most of the philosophical debate about narratives operates under the assumption that at least sometimes, we do

[247] See , for example, Hirsch (2018), Barnes (2020).

understand the past. However, it may turn out that under less idealised circumstances, we do not. Rejecting narrative ethics and discounting the past may be one way of avoiding the fallacy of overestimating our understanding of our past.

7.8 Conclusion

This chapter argued that the popular narrativity thesis in moral theory is mistaken – a life lived as a story is as such not better than a life without a story. I have shown that, if future-bias is rational and understood as including more than just hedonic goods and harms, this undercuts narrativity. At the same time, future-bias offers an alternative explanation to the shape of life hypothesis, capturing parts of the explanatory power the narrativity thesis possessed. Finally, I have suggested that future-bias is a good preference pattern to uphold, as it guards us against narrative fallacies – our tendency to overestimate our understanding of the past.

8 Conclusion

In this dissertation I have set out to defend the rationality of future-bias against arguments from temporal neutralists. I provided rebuttals to the following arguments:

1. Future-bias is impermissible because the intuitive support for its rationality is unstable.
2. Future-bias is impermissible because it is arbitrary to prefer goods to be in the future rather than past.
3. Future-bias is impermissible because it leads to pragmatic loss.
4. Future-bias is irrational because evolution, emotions, or other irrelevant influences have led us to be future-biased.

Beyond that, I have provided support for future-bias's rationality by arguing that

1. Our temporally embedded agency provides rational grounds for being future-biased.
2. The fact that we can sometimes influence the present and future, but not the past gives us reason to be future-biased.
3. Future-bias helps us to avoid committing to the sunk cost fallacy.

Taken together, I believe that these provide us with a systematic rationale not only for preferring bad things to be in the past, and good things to be ahead – these considerations support a broader principle explaining why we are future-biased: as I have argued in chapter 6 and 7, we are generally allowed to disconnect from our past and focus on the present and future.

I have shown that, if we accept this strong reading of future-bias as rationally permissible, this will undermine moral theories that are concerned with temporally extended selves, as I have demonstrated with whole life egalitarianism and narrative ethics.

Is this a good conclusion, even if plausible? In the beginning, I wrote about two different temporal perspectives in our moral and rational commitments: we think and act from a moment in time, and we think and act as temporally extended beings. For those who give priority to the latter perspective over the former, the conclusion that we are allowed to evaluatively disconnect from our past selves won't be a welcome one, and will invite resistance. How can it be a good thing to argue for fragmenting our moral and rational agency? How will we follow projects, or stay committed to our lifetime goals?

I believe, however, that my conclusion in this thesis is an encouraging one. Many, if not most of us live fragmented lives, where it isn't easy to make sense of our temporally extended agency, especially when it comes to our past. Many of us don't really understand what our past meant to us, what that means for our commitment and our direction in life. And even if we think that we know what the past means to us, this can turn out to be misleading due to our tendency to simplify how we evaluate our past out of the desire for coherence in our life stories.

So, the fact that we are rationally permitted to disconnect from our past should encourage us, as we can be full moral and rational agents even if we do not understand our past, or if we do not want our past to be part of ourselves.

And it will provide openings in our commitments: We will only stay committed to a project, an achievement, or a goal if we continuously want it, not because our past ties us to it. I believe this not to be a danger to our commitments, but a liberating encouragement to only follow those projects that we still want to follow.

Galen Strawson, in his quest against the narrative self, wrote:

> "But can Episodics be properly moral beings? [...] Diachronicity is not a necessary condition of a properly moral existence, nor of a proper sense of responsibility."[248]

Strawson was concerned mainly with narratively connected selves, arguing that we do not need to have (narratively) unified selves to be moral beings.

But I believe that the conclusion of this thesis should be read in a similar vein as an encouragement: We do not need to be (completely) temporally neutral to be a proper moral or rational agent. If you do not care about your past journeys, your failures, pains or achievements that are behind you – that's fine. You're still you, perfectly rational.

[248] Strawson (2004), p.450

8 Bibliography

Andersen, Sarah C. (2016): Present Me and Future Me. <https://sarahcandersen.com>

Arntzenius, F. & Elga, A. & Hawthorne, J. (2004). Bayesianism, Infinite Decisions and Binding. *Mind*.

Baier, A. (2004). Feelings that matter. In Robert C. Solomon (ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions*. Oxford University Press.

Beckstead, N. (2013). *On the overwhelming importance of shaping the far future* (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick)

Best, K., & Keppo, J. (2014). The credits that count: How credit growth and financial aid affect college tuition and fees. *Education Economics*, *22*(6), 589-613.

Bigelow, J. (1996). Presentism and properties. *Philosophical perspectives*, *10*, 35-52.

Bradley, B. (2009). *Well-being and Death*. Oxford University Press.

Bratman, M. (2012). Time, Rationality, and Self-Governance. *Philosophical Issues, 22.*

Brink, D. (1999). Eudaimonism, Love and Friendship, and Political Community. *Social Philosophy & Policy*, *16*, 252–289.

Brink, D. (2010). Prospects for Temporal Neutrality. In C. Craig (Ed.), *The Oxford Handbook of Philosophy of* Time. Oxford University Press.

Broome, J. (1999). Normative requirements. *Ratio*, *12*(4), 398-419.

Broome, J. (2005). Does rationality give us reasons? *Philosophical Issues*, *15*, 321-337.

Broome, J. (2018). Against Denialism. *The Monist*.

Bruckner, D. (2013). Present desire satisfaction and past well-being. *Australasian Journal of Philosophy*, *91*(1), 15-29.

Brueckner, A. and Fischer, M. (1986). Why is Death Bad? *Philosophical Studies*, *50*(2), 213-221.

Caney, S. (2020). Climate Justice. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 edn.). = <https://plato.stanford.edu/archives/sum2020/entries/justice-climate/>.

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.

Caruso, E. M. (2010). When the future feels worse than the past: A temporal inconsistency in moral judgment. *Journal of Experimental Psychology: General*, *139*(4), 610.

Chang, R. (2004). Can Desires Provide Reasons for Action?. In R. J. Wallace, P. Pettit, S. Scheffler & M. Smith (Eds.), *Reason and Value: Themes From the Moral Philosophy of Joseph Raz*(pp. 56-90). Oxford University Press.

Christensen, D. (1991). Clever Bookies and Coherent Beliefs. *The Philosophical Review*.

Cohen, G. A. (2000). If you're an egalitarian, how come you're so rich. *The Journal of ethics*, *4*(1-2), 1-26.

Cooke, R., Barkham, M., Audin, K., Bradley, M., & Davy, J. (2004). Student debt and its relation to student mental health. *Journal of Further and Higher Education*, *28*(1), 53-66.

Daly, C., & Liggins, D. (2010). In defence of error theory. *Philosophical Studies*, *149*(2), 209-230.

Deasy, D. (2015). The moving spotlight theory. *Philosophical Studies*, *172*(8), 2073-2089.

Deng, N. (2015). How A-theoretic deprivationists should respond to Lucretius. *Journal of the American Philosophical Association* 1 (3):417-432.

Dougherty, T. (2011). On Whether to Prefer Pain to Pass. *Ethics*.

Dougherty, T. (2015). Future-bias and Practical Reason. *Philosopher's Imprint*.

Dorsey, D. (2017). Future-Bias: A (Qualified) Defense. *Pacific Philosophical Quarterly*, *98*, 351-373.

Dorsey, D. (2018). Prudence and past selves. *Philosophical Studies*, *175*(8), 1901-1925.

Dorsey, D. (2015). The Significance of a Life's Shape. *Ethics*.

Dorsey, D. (2016). Future-Bias: A Qualified Defence. *Pacific Philosophical Quarterly*.

Dorsey, D. (2019). A Near-Term Bias Reconsidered. *Philosophy and Phenomenological Research*, *99*(2), 461-477

Fernandes, A. (2019). Does the temporal asymmetry of value support a tensed metaphysics? *Synthese*, 1-18.

Finlay, S. and Schroeder, M. (2017). Reasons for Action: Internal vs. External. In E. N. Zalta (Ed.), , *The Stanford Encyclopedia of Philosophy* (Fall 2017 edn.).= <https://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external/>.

Finnis, J. (2011). *Natural law and natural rights*. Oxford University Press.

Finocchiaro, P., & Sullivan, M. (2016). Yet Another "Epicurean" Argument. *Philosophical Perspectives*, *30*(1), 135-159.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, *40*(2), 351-401.

Glasgow, J. (2013). The Shape of a Life and the Value of Loss and Gain. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, *130*(5), 769.

Greene, P., & Sullivan, M. (2015). Against time bias. *Ethics*, *125*(4), 947-970.

Griffin, J. (2000). *Well-being and morality: Essays in honour of James Griffin*. Oxford University Press.

Griffith, P. E. (1997). What emotions really are. *The Problem of Psychological Categories, Chicago-London*.

Hájek, A. (2005). The Cable Guy Paradox. *Analysis*.

Hare, C. (2008). A Puzzle about Other-Directed Time-Bias. *Australian Journal of Philosophy*, *86*(2), 269-277.

Hare, C. (2009). *On myself, and other, less important subjects*. Princeton University Press.

Harman, E. (2009). I'll be Glad I Did It. *Philosophical Perspectives*, 23.

Heathwood, C. (2008). Fitting attitudes and welfare. *Oxford studies in metaethics*, *3*, 47-73.

Hedden, B. (2015). Time-Slice Rationality. *Mind*.

Hedden, B. (2015). Options and Diachronic Tragedy. *Philosophy and Phenomenological Research.*

Hedden, B. (2015). *Reasons without persons: Rationality, identity, and time*. Oxford University Press.

Hurka, T. (1987). The Well-Rounded Life. *Journal of Philosophy 84*, (12), 727-746.

Hurka, T. (1996). *Perfectionism*. Oxford University Press.

Kant, I. (1889). *Immanuel Kants Kritik der reinen Vernunft*. Mayer & Müller.

Kauppinen, A. (2012). Meaningfulness and Time. *Philosophy and Phenomenological Research*

Kauppinen, A. (2018). Agency, experience, and future bias. *Thought: A Journal of Philosophy*, *7*(4), 237-245.

Kagan, S. (2011). Do I Make A Difference?. *Philosophy and Public Affairs*.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan Publishers.

Kerr, A. D. (2014). *Affective Rationality* (Doctoral dissertation, The Ohio State University).

King, O. C. (2018). Pulling apart well-being at a time and the goodness of a life. Ergo.

King, O. C. (2019). The good of today depends not on the good of tomorrow: a constraint on theories of well-being. *Philosophical Studies*, 1-16

Korsgaard, C. (1989).Personal Identity and the Unity of Agency: A Kantian Response to Parfit. *Philosophy and Public Affairs*.

Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press.

Kolodny, N. (2003). Love as valuing a relationship. *The Philosophical Review*, *112*(2), 135-189.

Kolodny, N. (2005). Why be rational?. *Mind*, *114*(455), 509-563.

Kolodny, N. (2007). How Does Coherence Matter? *Proceedings of the Aristotelian Society, 107*(1pt3), 229-263).

Latham, A., Miller, K., Norton, J., Tarsney, C. (2020). Future-bias in action: Does the past matter more when you can affect it?. *Synthese*.

Levin, S. (2017). "Millionaire tells millennials: if you want a house, stop buying avocado toast". *The Guardian*. <https://www.theguardian.com/lifeandstyle/2017/may/15/australian-millionaire-millennials-avocado-toast-house>.

Lowenstein, G. E. J., & Elster, J. (1992). *Choice over Time*. New York: Russel Sage Foundation.

Lowry, R. and Peterson, M. (2011). Pure Time Preference. *Pacific Philosophical Quarterly,* 92(4), 490– 508.

Maclaurin, J., & Dyke, H. (2002). 'Thank goodness that's over': the evolutionary story. *Ratio*, *15*(3), 276-292.

MacIntyre, A. (1981). *After Virtue*. University of Notre Dame Press.

McKerlie, D. (1989). Equality and time. *Ethics*, *99*(3), 475-491.

Moller, D. (2002). Parfit on pains, pleasures, and the time of their occurrence. *Canadian journal of philosophy*, *32*(1), 67-82.

Noah, Paul (2017): Donald Trump's Inauguration in: The New

Yorker, Jan 2017. <http://paulnoth.com>

Nussbaum, M. (2004). Emotions as Judgments of Value and Importance. In R. C. Solomon (Ed.), *Series in affective science. Thinking about feeling: Contemporary philosophers on emotions* (p. 183–199). Oxford University Press.

Paul, L. A. (2010). Temporal experience. *The Journal of Philosophy*, *107*(7), 333-359.

Parfit, D. (1984) *Reasons and Persons*. Oxford University Press.

Parfit, D. (1986). Comments. *Ethics*, *96*.

Perović, K. (2019). Three Varieties of Growing Block Theory. *Erkenntnis*, 1-23.

Prosser, S. (2016). *Experiencing time*. Oxford University Press.

Pummer, T. (2013). Intuitions about large number cases. *Analysis*, *73*(1), 37-46.

Raibley, J. (2012).Welfare over Time and the Case for Holism. *Philosophical Papers*.

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Raz, J. (2000). The Central Conflict: Morality and Self-Interest. In Griffin, James et al. (Eds.), *Well-being and morality: Essays in honour of James Griffin*. Oxford University Press.

Rosati, C. (2013) The Story of a Life. *Social Philosophy and Policy*.

Rosenbaum, S. (1989). The symmetry argument: Lucretius against the fear of death. *Philosophy and Phenomenological Research*, *50*(2), 353-373.

Rosenberg, J. F. (2005). *Accessing Kant: a relaxed introduction to the Critique of Pure Reason*. Oxford University Press.

Schechtmann, M. (2007)Stories, Lives and Basic Survival: A Refinement and Defence of the Narrative View. *Royal Institute of Philosophy Supplements*.

Scheffler, S. (1994). *The rejection of consequentialism: A philosophical investigation of the considerations underlying rival moral conceptions*. Oxford University Press.

Sidgwick, H. (2019). *The methods of ethics*. Good Press.

Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy and Public Affairs*, *1*(3), 229-243.

Solomon, R. C. (1988). On emotions as judgments. *American Philosophical Quarterly*, *25*(2), 183-191.

Srinivasan, A. (2018). The aptness of anger. *Journal of Political Philosophy*, *26*(2), 123-144.

Strawson, G. (2004). Against narrativity. *Ratio*, *17*(4), 428-452

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, 109-166.

Street, S. (2015). Does anything really matter or did we just evolve to think so? In A. Byrne, J. Cohen, G. Rosen, and S. Shiffrin (Eds.), *The Norton Introduction to Philosophy* (pp. 685-693). New York: Norton.

Street, S. (2009). In Defence of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters. *Philosophical Issues, 19*.

Sullivan, M. (2012). The minimal A-theory. *Philosophical Studies*, *158*(2), 149-174.

Sullivan, M. (2018*). Time-Biases: A Theory of Rational Planning and Personal Persistence*. Oxford University Press.

Taylor, C. (1991). Sources of the Self: The Making of the Modern Identity. *Ethics*.

Unger, P. (1996). *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press.

Vavova, K. (2014). Debunking evolutionary debunking. *Oxford studies in metaethics*, *9*, 76-101.

Vavova, K. (2015). Evolutionary debunking of moral realism. *Philosophy Compass*, *10*(2), 104-116.

Vavova, K. (2018). Irrelevant influences. *Philosophy and Phenomenological Research*, *96*(1), 134-152.

Velleman, D. (1991). Well-being and time. *Pacific Philosophical Quarterley*, *72*, 48-77.

Woollard, F. (2015). *Doing and Allowing Harm*. Oxford University Press.