

Time Biases: A Theory of Rational Planning and Personal Persistence

Forthcoming—OUP. DRAFT NOT FOR DISTRIBUTION.

MEGHAN SULLIVAN, UNIVERSITY OF NOTRE DAME

June 6, 2017

CONTENTS

List of Figures	v
Introduction	vii
1 The Received Wisdom	1
1.1 The Savings Struggle	1
1.2 The Received Wisdom	2
1.3 Discounting and Near Bias	3
1.4 Temporal Discounting and Risk Tolerance	6
1.5 Arguments for the Received Wisdom?	8
2 The Life Saving Argument	9
2.1 The Art of Measurement	9
2.2 The Life Saving Argument	11
2.3 Rationality Doesn't Require You to Care?	13
2.4 Bookstore Buddhism	14
2.5 Measuring a Whole Life	16
3 The Arbitrariness Argument	21
3.1 The Arbitrariness Argument	22
3.2 Arbitrariness and Bias	24
3.3 Should we be Permissive?	26
3.4 Problems for Non-Arbitrariness?	29
4 Personal Volatility	33
4.1 Single-Self and Multiple-Self Models	34
4.2 Two Routes to Personal Discounting	39
4.3 A Qualitative Basis for Egoistic Concern?	40
4.4 The Concern Argument	42
4.5 Simple Connectedness and Moral Valence	43

4.6	Egoistic Concern and Bias	48
4.7	The Persistence Argument	50
4.8	The Two-Part Case for Future Neutrality	54
5	Preferences about the Past	57
5.1	A Negotiating Tip for Accident Victims	57
5.2	Bias Toward the Future	58
5.3	Determining Discount Functions for the Past	59
5.4	Do Subjective Probabilities Matter?	63
5.5	Preferences and Control	64
6	The No Regrets Argument	69
6.1	No Pain, No Gain?	70
6.2	The Mixed Tradeoff Argument	71
6.3	Resisting Temptation	72
6.4	Regrets, Affirmations and Planning	74
6.5	The No Regrets Argument	77
6.6	The Dead Have No Regrets	79
6.7	The Teen Pregnancy Problem	80
6.8	Success and Neutrality	83
7	The Arbitrariness Argument (Again)	85
7.1	Einstein's Hollow Condolences	85
7.2	The Arbitrariness Argument (Again)	87
7.3	The Obvious Asymmetry	88
7.4	The Asymmetry of Emotion	90
7.5	The Asymmetry of Attachment	91
7.6	The Asymmetry of Control	92
7.7	The Asymmetry of Possibility	93
7.8	Non-Arbitrariness and Neutrality	96
8	Understanding Temporal Neutrality	99
8.1	A Perceptual Model for Past Discounting?	100
8.2	Self and Other Asymmetries	102
8.3	Time Biases as Evolved Emotional Heuristics	104
8.4	Two Mistakes About Temporal Neutrality	107
8.5	Planning in Neutral	108
9	Neutrality, Sunk Costs, and Commitment	111
9.1	The Sunk Cost Puzzle	112
9.2	Weak Forecasting and Weak Honoring	113
9.3	Sunk Costs and "Rational Irrationality"?	115
9.4	Sunk Costs and Rational Non-Reconsideration?	117
9.5	Temporal Neutrality and Rational Commitment	118

9.6	Sunk Costs and Structuralism?	120
9.7	Sunk Costs and Integrity?	122
9.8	Sticking to Your Plans	125
10	Neutrality and Extending Your Life	127
10.1	The Lazarus Problem	127
10.2	Afterlives	130
10.3	Caring about Difference?	131
10.4	Prioritizing Satisfaction	134
10.5	Temporal Neutrality and Life Extension	135
10.6	Caring about Phenomenology?	138
10.7	Better Retirements, Better Deaths	140
11	Neutrality and Meaning	141
11.1	The Temporal Argument for Nihilism	141
11.2	Meaning as a Form of Value	142
11.3	Future Bias about Meaning	145
11.4	From Future Bias to the Permanence Principle	147
11.5	From Permanence to Nihilism	148
11.6	Subjectivism about Meaning?	149
11.7	Heavenly Optimism?	150
11.8	Temporal Neutrality about Meaning	151
11.9	Meaning and Vulnerability	154
	Bibliography	163

TIME BIASES

You are time-biased if you have systematic preferences about when events happen. For instance, you might prefer to spend your tax return now rather than save it for a rainy day, believing full well that if you put it in savings, it would increase in value. This is a near-bias – caring less about the experiences money can buy when they are scheduled further in the future. You might prefer that a painful root canal scheduled for tomorrow was already over and done with. This is a future bias – caring less about an unpleasant experience because it's past. You might prefer that the best part of a vacation come at the end rather than the beginning. This is a structural bias – caring about the temporal sequence of your experiences. You might even have preferences about when events outside of the confines of your life happen. Perhaps you now prefer that environmental crises happen a long time after your death rather than in the next few generations, because immanent catastrophes would diminish the meaning of your projects. This is a kind of meaning bias – caring about when your life occurs in relation to other events.

Philosophers, economists, and social psychologists have tended to treat these time biases in a piecemeal fashion, offering different theories of the origin of such attitudes, different theories of how to measure their extent, and different verdicts on whether these various time biases are rational or irrational. But in this book, I will argue that our time biases are far more similar than we have acknowledged. And there is good reason to think that all time biases are irrational. Learning to recognize and overcome our irrational time preferences can help us become better planners. Indeed, if Plato is right, overcoming our time biases might even save our lives.

The book is divided into three parts. The first part (Chapters 1-4) will focus on whether it is rational to care less about events that would happen in your distant future. Most social psychologists and philosophers have simply

assumed that near bias is a pervasive but irrational attitude. But why is it wrong to care less about the distant future? I will isolate two philosophical arguments for why near bias is irrational and defend the arguments against objections. In the process, I will also offer a theory of rational egoistic concern.

The second part (Chapters 5-7) will consider whether it is rational to care less about some events just because they've already happened. This attitude, while common, has been largely ignored by social psychologists and philosophers. I will offer a theory of what it is to discount your past, and then argue that such a time bias is irrational for the same reasons it is irrational to be near-biased.

The final part (Chapters 8-11) will develop a theory of why we are time-biased even though we shouldn't be. I will explain what is (and, more importantly, isn't) entailed by temporal neutrality. And I will show how this theory can help us in determining when to stick to our past plans, in planning for the future in the face of radical change, and in understanding the meanings of our lives even if none of our plans make a permanent difference to the world.

If your curiosity is piqued and you'd like to get straight into the debates about time biases, you can skip ahead to the first chapter. But some readers might be interested in a preview of the kind of rationality at play in these debates. The rest of this chapter will describe the type of rationality that I am interested in, contrast it with some other common approaches, and then introduce the three main principles of rational planning that will reappear throughout the book.

WHAT YOU DO vs. WHAT YOU PREFER

This is a book about what you have most prudential reason to prefer and the extent to which temporal considerations should affect these preferences. Some philosophers – and many economists – think we ought to take our preferences as the fundamental inputs to a theory of rationality. On this view, there is no particularly deep question of why you have the preferences that you do. We just take it for granted that you have some preferences and then ask whether your set of preferences, taken as a whole, is logically consistent at a particular time. For instance, suppose Sparky the pyromaniac reports that right now he prefers blowing up cars to slowly burning them. But – Sparky immediately continues – right now he *also* prefers slowly burning the cars to blowing them up. The minimalists will deem Sparky irrational for having contradictory preferences. But they remain silent on whether there is anything irrational about wanting to torch vehicles in the first place.

We often have good reason to make sure our preferences conform to logi-

cal constraints. But we omit an important dimension of rationality if we take individual preferences to be immune from criticism. Rational beings can and should deliberate about what preferences they ought to have. And in deliberating about the reasons that back our particular preferences, we often come to realize that some of our preferences are irrational. For instance, I might start off the electoral year preferring that Bernie Sanders win the 2016 primary rather than Hilary Clinton, with little reflection on my reasons for this preference and not having formed any other electoral preferences. I find that I just like Bernie. This preference is logically consistent. But I might nevertheless find myself learning more about the various candidates' platforms and on the basis of new reasons come to prefer that Clinton win. There is an important sense in which my preferences are more rationally justified after this research and deliberation. And this improvement need not be explained as my preferences better cohering within some formal framework with any of my other preferences. It has to do with my uncovering the reasons for what I should want. Philosophical reflection can help us uncover reasons for and against our preferences in just this way, and your preferences are more rational to the extent that they are better supported by reasons.

Another major branch of philosophical work on rationality focuses on practical rationality – how we should determine our reasons for action. For instance, I might wonder if I have most reason to vote on November 3rd, or if I have most reason to stay home. Within this sphere, we may also draw a distinction between moral and prudential reasons. I may think I have the most moral reason to vote, since I think that universalizability is an important moral standard, and if everyone declined to vote, our democracy would fail. But I might also think that I have most prudential reason to stay home, since my individual vote will likely not make a difference to my future well-being, while staying home would allow me to sleep in an extra hour.

Aristotle thought that issues of prudence were only interesting insofar as our preferences relate to our actions: “Thinking itself moves nothing, but thinking that is for the sake of something and concerned with action does... Nobody now chooses to have sacked Troy in the past, and nobody deliberates either, about what has already come into being but rather what will be and admits of happening.”¹ In this, he is not alone; many contemporary philosophers assume that issues of practical rationality are only interesting insofar as they pertain to rational action.² But in exclusively focusing on reasons for action, this concept of practical rationality also overlooks an important dimension of rational deliberation. We often deliberate about what we prefer the world to be like, even if we have no control over whether the world will be that way. For instance, I might deliberate about which outcome I prefer for the US Presidential election, believing full well I have no control

¹Book 6, Chapter 2 of the *Nicomachean Ethics*.

²See for instance Chapter 1 of List and Pettit (2013) or Korsgaard (2009).

over the outcome. And often we deliberate about what we prefer while not really knowing how our actions might match up with these preferences. I may decide, for instance, that I prefer not to be kept on life support if I suffer a serious brain injury. But I might not know what course of action is best for realizing this preference. I may not even be able to bring myself to think now of any of the practical steps that realizing this preference would entail.

On the view of rationality investigated in this book, the preferences typically come first; the connection with action often comes later and sometimes not at all.³ Here are five more examples of preference-focused rationality that will re-appear in later chapters.

First, we can reason about our preferences for what the past could have been like. For instance, do you prefer your painful root canal to be over or to be still scheduled for a month from now? And what rational status do you give regrets – what should you prefer now about your past choices (and the preferences that guided those choices)? We have no control over the past, and this deliberation is not reasoning about what to do in the past. It is deliberation about what you prefer the world to have been like.

Second, we can deliberate about our personal post-mortem preferences. For instance, do you prefer that there be some form of spiritual afterlife? Or do you prefer that your natural death be the end of your existence? While we have control over whether we follow the precepts of some particular faith, we do not have any control over whether an afterlife is possible or what such a continued life would entail. Still we can deliberate about what we want, even if we have no control over it. And such deliberation might indirectly influence decisions that we can control – for instance whether to attend a Catholic Mass or an atheist Sunday Assembly each weekend.

Third, we can deliberate about how we want the world to go after our deaths. Do you prefer that climate change hasten the end of human civilization, so you can be part of the last chapter of the human story? Do you prefer that there be many future generations after ours, who take an interest in the kinds of projects you cared most about in your life? You might not be able to hasten or delay the apocalypse, but you can form preferences about what happens outside the confines of your life, and, as we'll see, such preferences can play an important role in the kinds of projects you pursue now.

Fourth, we can deliberate about our preferences over scenarios we aren't sure are physically or biologically possible. A patient suffering permanent locked-in syndrome can deliberate about whether he'd prefer to run marathons or climb mountains. A space enthusiast can deliberate about which galaxy

³Which is not to say I take any firm stand on the debate between causal and evidential decision theorists. The connections between such considered preferences and actions might be somewhat sophisticated, especially when facing Newcomb problems. For an overview of issues with translating "news value" accounts of rationality into rational decision-making, see Chapter 5 of Joyce (1999).

he'd prefer to visit, even in the absence of faster-than-light travel.

Finally, we can deliberate about our preferences over scenarios we aren't sure are metaphysically possible. For reasons related to their strange religion, the Pythagoreans strongly preferred that the square root of two be a rational number. Thomas Hobbes preferred that he be the scholar to discover how to square a circle. Both scenarios are metaphysically impossible, but not straightforwardly logically inconsistent. And both Hobbes and the Pythagoreans had reasons for their preferences – reasons a theory of rationality can weigh in on. We'll return to these issues in Chapters 5 and 7.

These kinds of deliberation are not only common, they are important to our agency and planning. Or so I shall argue here.

It would be useful to have a general name for the kind of rationality that concerns our reasons for wanting or not wanting different events to occur (or to have already occurred). For lack of a better term, I will call it *approbative rationality* – the rationality that governs what we should approve of happening.⁴ As with practical rationality, we can distinguish moral approbative rationality from prudential approbative rationality. Moral approbative rationality concerns your preferences given your moral reasons. Moral approbative rationality might require you to prefer a world where more wealth is redistributed to the actual world, because such a world would be overall happier. Moral approbative rationality might require you to prefer that World War I never occurred, even if you have no control over past wars. In contrast, prudential approbative rationality concerns reasoning about your preferences given your self-interest. Prudential approbative rationality might require you to prefer the actual world to a redistribution world, since you would likely lose money in a just redistribution scheme. Prudential approbative rationality might require you to prefer the actual world to one where World War I never happened, if the Great War was a precondition of your grandparents meeting and you eventually existing.

Approbative rationality isn't completely disjoint from practical rationality, because your reasons for action will tend to be based on your preferences, including ones that you've deliberated about. Furthermore, preferences that seem inert when considered on their own might be action-guiding when they are combined with your other beliefs and preferences. Indeed, many of the time-biased preferences we will discuss in the coming chapters have exactly this effect. What is distinctive about this approach is that we do not take preferences to be buried in us, waiting to be revealed by our choices. While

⁴There is an idea like approbative rationality in Richard Jeffrey's decision theory. He thinks we should model preferences over propositions in terms of the "news value" that learning such states of affairs obtain. Jeffrey (1983: 82-83). But unlike Jeffrey's system, approbative rationality (1) focuses on reasons you might have for taking certain "news items" as preferable to others and (2) permits you to have rational preferences for metaphysical impossibilities.

complex features of our psychology might obscure some of our preferences, many of them – including some that are most important to who we are – are the sorts of attitudes that we can and should support with reasons. How exactly our preferences connect with our reasons will be a recurring topic in the chapters to come.

THREE PRINCIPLES FOR RATIONAL PLANNING

This is a book about prudential approbative rationality, i.e. what preferences you should and shouldn't have insofar as you care about your self-interest. I will assume without much argument that there is a coherent distinction between self-interest, morality and other kinds of normativity. I also agree with the philosophers and economists who think that rational preferences must be logically consistent at a time. This is a principle we can call:

Consistency: At any given time, a prudentially rational agent doesn't prefer states of affairs that are logically inconsistent.

Consistency gives one account of what's irrational about Sparky's automotive destruction preferences – they straightforwardly contradict one another. In other cases, preferences are broadly logically inconsistent because of what particular terms mean. For instance, someone who prefers to both stay vegan and to eat cheeseburgers fails the consistency test because being vegan *just means* abstaining from animal products.⁵

To the Consistency criterion, I add two more, which will be defended in the chapters to come. First, I assume that being prudentially rational should, in normal circumstances, mean preferring that our lives go as well as possible into the future. We learn about prudential rationality (and write books about it!) because we want to improve our lives and the lives of those who follow our advice. This motivates a principle that I call:

Success: At any given time, a prudentially rational agent prefers her life going forward go as well as possible.

In Chapters 2, 6, and 8, we'll clarify this principle and survey some arguments for and against it.

I also think that being rational means forming preferences and choosing actions that are supported by reasons. Indeed, the focus on reasons is characteristic of approbative rationality and motivates a principle that I call:

⁵I will not, however, take consistency to rule out preferences over metaphysical impossibilities. Examples might include the squared circle or a rational root of 2. And I won't assume that metaphysical impossibilities all reduce to forms of logical impossibility.

Non-Arbitrariness: At any given time, a prudentially rational agent's preferences are insensitive to arbitrary differences.

In Chapters 3, 7, and 8, we'll clarify this principle and survey some arguments for and against it.

There are exotic cases where aiming at Success can seemingly conflict with Consistency and Non-Arbitrariness. For example, suppose an eccentric billionaire promises to bestow his fortune on you but only if you develop inconsistent preferences. Suppose further that he can somehow detect your preferences; maybe he is a brilliant psychologist or a telepath. In such a case, it seems you can't satisfy both Consistency and Success. Likewise we can imagine a problem case for Success and Non-Arbitrariness where the billionaire demands you form arbitrary preferences. We can even imagine a case where the billionaire promises a fortune only if you stop caring about your self-interest altogether, thereby making the Success criterion self-defeating.

How bad are such scenarios for an account of prudential rationality? We might avoid some of the conflicts by trying to reduce Consistency and Non-Arbitrariness to Success (or vice versa). Another option is to claim that they are independent but one of the rational standards always takes precedence over the others. A third option is to admit that we sometimes face rational dilemmas: planning problems where no matter what we end up partially failing.

I lean towards admitting that there are, occasionally, insuperable rational dilemmas – the eccentric billionaire problems being a paradigm. But I won't defend that in this book, and I won't argue for the supremacy of any of the three norms. Different arguments that I'll offer appeal specifically to versions of the Success or Non-Arbitrariness principles. The arguments can be taken separately if you find you accept one principle but reject the other.

Happily we rarely find ourselves faced with dilemmas involving billionaires fixated on spreading irrationality. Most problems of rational planning have solutions, and in many cases ensuring that your beliefs are consistent and non-arbitrary will help you to live better. Further, as a methodological point, in this book I'm committed to presenting real world planning problems whenever possible to make my argument. The theory on offer in this book should be tested against its ability to offer systematic and useful advice for how to form preferences in the face of common planning questions just as much as it solves the puzzles of the philosophers.

You'll notice that the three principles discussed above apply to agents *at a time*. There are two perspectives you might take when you think about your plans and your self-interest. One is the *present perspective* – situated as you are now, in the present, are your preferences rational now and is your life going well from the present vantage point? Another is the *timeless perspective* – considering your life as a whole, is it a life rationally lived and a life that goes

well? There is some philosophical disagreement about which perspective we ought to take when evaluating our preferences. And the perspectives can offer conflicting verdicts about whether we are doing well. The arguments in this book will assume you are planning from the present perspective. That is, they will be posed to answer the question of what you should prefer now rather than what would be valuable from some timeless assessment of your life. I favor the present perspective approach for two reasons. First, as I will describe in more detail in Chapters 4 and 7, I think there are real and irreducible distinctions between the past, present and future. We should not be so quick to abstract away from these differences when we consider our potential sources of reasons. Second, it is far more natural, I think, to be engaged in planning while also seeing our life as an event that we are in the middle of, remembering what happened, anticipating what will come next, and trying to point it in the right direction. So the temporal neutrality I will advocate does not require us to ascend to some timeless perch to be rational.

ACKNOWLEDGMENTS

This book was initially dreamed up while on a year-long research sabbatical in 2015-2016. I am extraordinarily grateful to the University of Notre Dame, the Immortality Project, the New Directions in Time Project, and the John Templeton Foundation for supporting that year of research. I am also intellectually indebted to Preston Greene, Tom Dougherty, David Brink, and Derek Parfit. All four have done extensive, interesting work on temporal neutrality that has significantly guided my thinking. Chapters 5, 6, 8 and 11 draw in certain sections from research that Preston and I conducted together. His insight runs throughout those arguments (and is frequently cited). Chapters 3 and 7 draw in part from joint work with Peter Finocchiaro (likewise cited). Johann Frick and Tom Hurka offered very helpful suggestions during my visit to Princeton, which substantially shaped the direction of Chapters 5 and 6. The entire manuscript benefitted extensively from feedback from my Spring 2016 graduate seminar on diachronic rationality: Elizabeth Anthony, Paul Blaschko, Rebeca Chan, Kate Finley, Liz Jackson, Ting Lau, Caleb Ontiveros, Callie Phillips, Luke Riel, and Louise Williams. Notre Dame has a wonderful community of smart, creative and relentless graduate students; I love sharing my work with them. Mike Rea and Tom Dougherty also gave extensive, extraordinarily helpful comments on the entire manuscript. Christine Grandy helped me make beautiful charts representing the various discount functions.

A version of Chapter 4 is forthcoming as an article in *Philosophical Issues*.⁶ And a version of Chapter 11 is forthcoming as an article in the *Georgetown*

⁶Sullivan (Forthcominga).

Journal of Law and Public Policy.⁷ I am grateful to both journals for their permission to reprint this material.

I am indebted to many departments and conferences for providing me venues to present portions of this book while it was in progress: the University of Iowa, NUS, NTU-Singapore, Syracuse, Hong Kong University, Lingnan University, the Centre for Time at Sydney, the ANU, the University of Manitoba, the University of Vermont, Rutgers, Duke, the University of Utah, the Notre Dame Law School, Princeton, UMSL, Brown, Columbia, the University of St. Thomas, the Georgetown Philosophy Department, the Notre Dame Center for Philosophy of Religion, the Immortality Project Capstone Conference, the Sydney Temporal Experience and Transformation Conference, the Iceland Future of the the Study of Time Conference, the 2015 Rutgers Religion and Value Workshop, the 2016 Eastern APA, the 2016 Orange Beach Epistemology Workshop, the Georgetown McDonough Business School Conference on Remote Effects, the 2017 Desert Philosophy Workshop, the 2017 APA Pacific Pre-Conference on Transformative Experience, the 2017 Toronto Time in the History of Metaphysics Conference, and the 2017 Milan Open Future Workshop.

Most importantly, I am grateful to my family – Mary, Liam, Patrick, Ambrosia and Connor – for teaching me humor, unconditional love, and that there is more to living than planning. This book is dedicated to them.

⁷Sullivan (Forthcomingb).