



MONASH University

AXIOLOGIES AT ODDS
RECONCILING INCOMPATIBLE THEORIES OF
THE GOOD

SAMUEL MAWSON
B.A. (HONS)

*Submitted for the degree of Master of Arts
Department of Philosophy
Monash University, 12 April 2017*

Copyright Notice

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

© Samuel Mawson (2017)

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution.

To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.



Samuel Mawson

12 April 2017

Abstract

Time discounting and Prioritarianism are both defensible components to include in an axiological theory. However, I argue that combining these features can lead to undesirable results. Prioritarianism is the view that a fixed improvement in someone's wellbeing matters more, morally speaking, the worse off that person is. Time discounting represents the view that an event's moral significance depends on when that event occurs. Typically, this involves ascribing less importance to events in the distant past and future than to events that occur nearer to the present. I demonstrate that if an axiology incorporates both time discounting and Prioritarianism, it can produce inconsistent evaluations over time. The kind of inconsistency generated by this combination is particularly problematic because it can lead an axiology to endorse a series of decisions that are ultimately disadvantageous by the axiology's own evaluative standards. I argue that, not just Prioritarianism, but also any other good that supervenes on the distribution of wellbeing over time, risks generating this kind of inconsistency if it is paired with time discounting. These results are indicative of a broader class of problems that can arise when one includes multiple goods that supervene on the distribution of wellbeing as part of an axiology. Next, I demonstrate a second inconsistency result that can occur when this class of goods is paired with discounting and I propose a list of criteria that axiologies must adhere to in order to avoid these inconsistencies and other undesirable results. These criteria prohibit a number of combinations of goods that supervene on wellbeing. Finally I investigate ways of defending non-standard representations of these goods that would allow them to be combined without violating the criteria I have proposed.

Acknowledgements

I am extremely grateful to Toby Handfield (my supervisor) for his encouragement, his patience with my frequent meanderings into new areas of literature, and most of all for the countless helpful critiques and suggestions that he provided at every stage of this thesis. I would also like to thank Justin Oakley for the thoughtful comments he provided on a number of partial drafts of this thesis.

I have benefitted from seminar questions from Monima Chadha, Lloyd Humberstone, John Thrasher, and Robert Simpson. I would like to thank Robert Simpson in particular for the support and encouragement he has given me throughout my time at Monash University, and for helpful discussions during the early stages of my thesis.

Early versions of Chapter 1 were presented at the Australasian Association of Philosophy Conference in 2016 and the Australasian Postgraduate Philosophy Conference in 2015. I was grateful to receive stimulating comments and questions from the attendees at these talks, especially Suzy Killmister and Garrett Cullity.

My thanks also to Hilary Greaves, with whom I discussed my early ideas for this thesis. Although none of those ideas made it into the final manuscript, my work has greatly benefited from reading her insightful analyses of numerous topics in axiology.

I would like to thank my parents for their unfailing support throughout my academic career and for enduring several long discussions about population ethics while I was developing my ideas for this thesis.

Finally, I offer my thanks to Kelsey Kingsbury for her patience and her constant support and encouragement throughout this process.

Glossary of Mathematical Terms

A	<i>an additive representation of a pattern good</i>
b_i	<i>birth date of individual i</i>
d_i	<i>death date of individual i</i>
D	<i>discount factor that adjusts the contribution a good makes to overall goodness</i>
g	<i>contribution that time t in person i's life makes to overall goodness</i>
G	<i>overall goodness of a state of affairs</i>
i	<i>a person</i>
I	<i>the input used in a pattern good function</i>
N	<i>the set of all natural numbers</i>
P	<i>Prioritarian Function (a strictly increasing concave function)</i>
R	<i>a replacement representation of a pattern good</i>
t	<i>time</i>
T_n	<i>denotes a specific time. T_1 refers to time = 1, and so on.</i>
w_i	<i>total lifetime wellbeing of person i</i>
w_t	<i>wellbeing of all people who exist at time t</i>
$w(i, t)$	<i>wellbeing of person i at time t</i>

Contents

Chapter 1

I.	Introduction	1
II.	Axiology	2
III.	Prioritarianism	8
IV.	Discounting	11
V.	Combining Discounting and Prioritarianism	21
VI.	The Money Pump	26
VII.	Why the Money Pump Occurs	32
VIII.	How Alternative Combinations of Discounting and Prioritarianism Relate to the Money Pump	33

Chapter 2

I.	Introduction	36
II.	Ways of Representing a Pattern Good Mathematically	38
III.	Types of Goods	40
IV.	What Types of Goodness are Associated with the Common Pattern Goods	43
V.	Constraints on How an Axiology Can Represent Two or More Pattern Goods	46
VI.	Conclusion	54

Chapter 3

I.	Introduction	55
II.	Pattern Goods Rule out Separability	55
III.	Dispersion	58
IV.	What Counts as Wellbeing	62
V.	Indexing	64
VI.	Combining Discounting with Other Pattern Goods	68
VII.	The Two Main Types of Arguments for Discounting	70

VIII.	Standard Discounting Expresses Two Kinds of Pure Time Bias	70
IX.	The Democratic Representation Argument	75
X.	The Special Relations Argument	78
XI.	Over-demandingness Objection	79
XII.	Optimal Investment	81
XIII.	Infinite Times Paradox	82
XIV.	Conclusion	83
	Reference List	85

CHAPTER 1

I. Introduction

This chapter is about trying to reconcile two attractive ideas in moral theory. One idea is Prioritarianism, which holds that the interests of people who are badly off should be prioritised over the interests of people who are comparatively well off. The other idea is relative time discounting, which holds that events that occur at times near to the present moment matter more, morally speaking than do events that are temporally remote. These views both have appealing qualities and many people defend their use in practical morality. However, I will argue that it is inadvisable to employ a moral theory that affirms both of these views simultaneously. It may be that, at most, only one of these views can be a component of a satisfactory theory.

Some theorists hold that events in the distant future and past should contribute comparatively less to overall goodness than events near to the present. In other words the present is special. Hence this view is known as relative time discounting. It holds that an event's moral significance is relative to how close it is to the present moment. Discounting appeals to many theorists because it approximates the evaluations inferred from people's real decision making. Discounting also has some appeal because without it, some typical classes of moral theory recommend investing implausibly large amounts into promoting the wellbeing of people who live in the distant future as opposed to people who are alive today.

Many people think concern for the worst off should play a role in the evaluation of states of affairs. This value is commonly represented by either Prioritarianism or by concern for equality. In this essay I will concentrate on Prioritarianism, which holds that a fixed improvement in a person's wellbeing matters more, morally speaking, the worse off that person is¹. In general Prioritarianism delivers similar results to concern for equality, but it has a number of distinguishing features that can be

¹ This summary of the Prioritarian view is borrowed from Toby Ord (2015).

advantageous. Furthermore, there is considerably more consensus about how to formally represent Prioritarianism in the kinds of theories I will be discussing, so the argument using Prioritarianism is a bit more straightforward. Given this, I will use Prioritarianism as the focus of my argument. However the results I demonstrate for Prioritarianism also apply to a number of common formulations of concern for equality.

I will demonstrate that if a theory incorporates both relative time discounting and Prioritarianism it delivers problematic inconsistencies in evaluation; in certain contexts, an axiology that represents these two views will endorse a series of decisions that are ultimately disadvantageous by its own lights. For instance, it might recommend selecting a particular course of action, and then subsequently recommend seeking to reverse that course of action at a cost, producing an outcome worse than if the initial action had not been taken in the first place.

Section II will specify the kinds of moral theory I will limit my discussion to and outline the assumptions that I will use for the rest of this thesis. Section III will give a more detailed exposition of Prioritarianism. Section IV will give an exposition of discounting and briefly discuss some of the arguments against discounting future times. Section V will discuss the ways in which a formal theory can be designed to represent both discounting and Prioritarianism. Section VI will discuss how, on what is otherwise the most appealing way of combining discounting and Prioritarianism, a dangerous inconsistency results. Section VII will discuss which of alternative ways of representing these two views are capable of avoiding the inconsistency. Section VIII concludes the discussion by summarizing the relative impacts of the alternative approaches.

II. Axiology

In this thesis I will be discussing theories that fall within the field of axiology. Axiological theories are solely concerned with determining the goodness of states of affairs. This means they omit further considerations that bear on how one should act such as duties or virtue ethical concerns. Hence the distinction between axiological and other theories of morality is sometimes characterised as the distinction between theories of the good and theories of the right. Theories of the good are concerned

with how good states of affairs are in and of themselves; theories of the right are concerned with determining what constitutes right action for a given person in a given circumstance. It is almost universally agreed that the goodness of states of affairs has at least some bearing on right action. Some theories, such as classical utilitarianism, make the goodness of outcomes the sole consideration relevant to right action; other theories introduce further constraints that bear on the rightness of action. I will set aside questions regarding which other considerations (if any) should bear on right action and focus solely on theories of the good.

Axiological theories often include two key aspects: firstly, an account of the goods that they take to be morally significant. A ‘good’ is just something desirable. For instance these could be mundane things such as money or medicine. But I will use the term to refer specifically to moral goods – these are the features of a world that contribute towards its overall goodness. One common example of a moral good is wellbeing. Secondly, a systematic method for determining overall goodness on the basis of the goods that are present in a state of affairs; in other words, a way of aggregating the moral significance of individual goods to determine their cumulative significance.

Many theorists working in axiology hold that people’s wellbeing is the most important factor in determining the overall goodness of states of affairs. For instance, John Broome (2004) argues that considerations unrelated to people’s wellbeing, such as beauty that no one appreciates, are either irrelevant to axiology, or if they are relevant, it at least does no harm to assess the contribution that they make to goodness separately from the contribution made by wellbeing. Since this is a common view, and the results I will be discussing depend only on values related to wellbeing, I will focus on wellbeing and different approaches to aggregating its moral significance². However I am open to the possibility that there may be other morally relevant goods. Examples of candidate goods that I set aside include: the

² Furthermore, the results I discuss will occur in any axiology that aggregates wellbeing in the way I discuss, this will be the case whether or not the axiology also incorporates other goods.

wellbeing of animals, the ‘health’ of ecosystems, and beauty that no one appreciates³.

I will assume that wellbeing can be represented on a cardinal scale and that interpersonal comparisons of wellbeing are possible. If wellbeing is represented on an *ordinal scale*, it is possible to make judgments of the form: outcome A is better for this person than outcome B; if wellbeing is instead represented on a *cardinal scale*, it is possible not only to make these sorts of judgments, but also to make judgments of the form: outcome A is twice as good for this person as outcome B. A cardinal scale not only provides a ranking of different levels of wellbeing, but also a measure of their relative magnitude. By interpersonal comparisons being possible I mean that it is possible to make definitive statements about whether a given state of affairs is better for one person than it is for another. In other words, each person’s cardinal representation of wellbeing shares a common scale. Making these assumptions is essential for the kinds of theories I will be discussing. These are fairly standard assumptions to make in axiology, but they are not trivial. Historically, a substantial amount of technical work and argumentation has gone into defending their use⁴. However due to limitations of scope I must gloss over them here.

I will further assume that a person’s momentary wellbeing (that is, his or her level of wellbeing during a given time interval) is the basic unit of goodness (as opposed to overall lifetime wellbeing). Therefore, all the wellbeing present in a state of affairs can be represented as a series of cardinal wellbeing scores, each corresponding to the level of wellbeing enjoyed by a particular person during a particular interval of time⁵. I will refer to this arrangement as ‘the distribution of wellbeing over times

³ If one wishes, animals can be incorporated into this framework relatively easily by considering them as persons and attributing wellbeing levels to them.

⁴ For a summary of the arguments and technical work that have gone into justifying these assumptions see Hilary Greaves (2014).

⁵ Using people’s momentary wellbeing as the basic unit of goodness presupposes that momentary wellbeing is sufficient, in aggregate, to exhaustively determine the overall goodness contributed by people’s wellbeing. This assumption is controversial though. Some

and persons’, or ‘the distribution of wellbeing’, for short. For a given person, the levels of wellbeing he or she experiences over a series of time intervals can be represented in a diagram like the one in figure 1 below. I will use this kind of pictorial representation throughout to help illustrate my arguments.

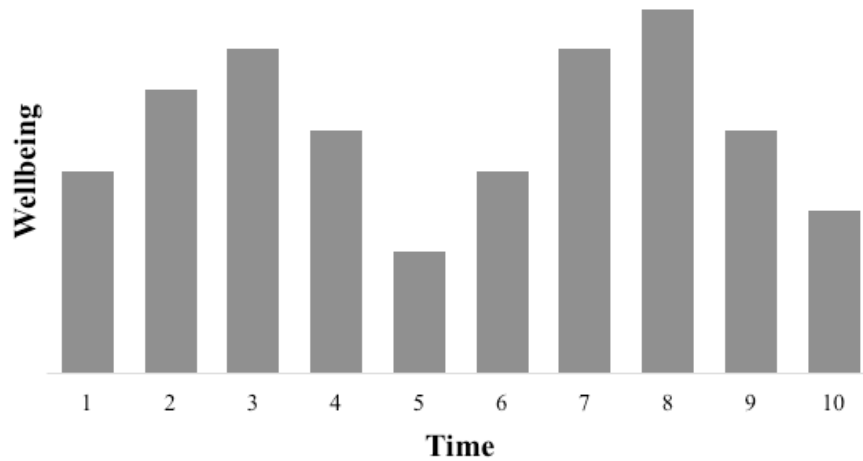


Figure 1. Representation of wellbeing over time

The theories I will be discussing hold that the overall goodness of a state of affairs can be represented by a function that takes as its arguments the values in each locus on the distribution of wellbeing and outputs a single numerical value; the greater this numerical value, the greater the goodness of the state of affairs. Like wellbeing, the outputs of this function are cardinal values, so they provide not only a ranking of states of affairs in terms of goodness, but also a representation of the magnitude of the differences in goodness between states of affairs. The most well-known example

theorists, including David Velleman (2000) have argued that a person’s overall lifetime wellbeing depends on more than just momentary wellbeing, and that its amount and therefore its contribution to goodness can not be ascertained by examining momentary wellbeing alone. Velleman argues that other factors such as the narrative relations between parts in a person’s life are also relevant to determining lifetime wellbeing. I am sympathetic to Velleman’s arguments. However my aim here is not to defend the claim that momentary wellbeing is sufficient to determine overall goodness. Instead I wish to examine what sorts of ethical positions can be faithfully represented by an axiology that uses momentary wellbeing as the basic unit of goodness. Hence I shall maintain this assumption.

of such a function is the one endorsed by classical utilitarianism. Classical utilitarianism states that overall goodness is determined by the sum total of people's wellbeing. Greater total wellbeing translates to greater overall goodness. This can be represented by an expression like the one shown below, where w_i represents the total lifetime wellbeing enjoyed by the i^{th} person in a state of affairs (person 1, person 2, etc.). And n is the total number of people who ever live in the state of affairs that is being evaluated. So this function steps through each person who ever exists in the state of affairs and sums their wellbeing.

$$\text{Overall goodness} = w_1 + w_2 + w_3 + \dots + w_n$$

This representation is straight forward, but it takes lifetime wellbeing values, rather than values for wellbeing at a time, as its arguments. A more appropriate function for my purposes here is shown in the expression below (1):

(1)

$$G = \sum_i \left(\sum_{t \in N(i)} w(i, t) \right)$$

Where $w(i, t)$ represents the momentary wellbeing of person i at time t . $\sum_{t \in N(i)} w(i, t)$ takes the sum of the momentary wellbeing levels for all values of t that exist in the set of times in the life of person i . $\sum_i()$ takes the sum of the values returned by the function inside the parentheses for every possible value of i (every person that exists in the state of affairs). I use G here to stand for overall goodness. Expression (1) defines G as the sum of the sums of all of the momentary wellbeing across each person's life.

This sort of formal mathematical approach is somewhat removed from commonsense morality but it does have a number of advantages. It can provide a clear guide for action since it unambiguously ranks the goodness of different states of affairs; the systematic approach ensures that the same evaluation criteria will be applied consistently; and because it ranks the overall goodness of states of affairs on a cardinal scale, it enables the use of expected utility theory, which is widely considered to be the best available formal theory of decision making for situations in

which the consequences of one's possible actions are uncertain. Expected utility theory involves multiplying the goodness of an outcome by the probability that that outcome will follow from a given course of action. This is done for all outcomes that might follow from that course of action, and then the results are summed to determine the expected goodness of that course of action.

By assuming that wellbeing is the only good under consideration this approach might seem to be limiting, but there are a number of values which are not synonymous with wellbeing but which supervene on wellbeing that can be accounted for by this framework. The straightforward summation function described above represents the classical utilitarian idea that all wellbeing at all times for all people counts equally. In other words, a unit of wellbeing contributes the same amount to overall goodness regardless of the time at which it occurs and the person whose wellbeing it is. In the rest of this thesis I will be discussing aggregation functions that diverge from this model so that the value of a given unit of wellbeing is dependent on a range of factors such as the amount of wellbeing enjoyed by other people. By using functions that make the contribution of wellbeing depend on other factors it is possible to express a variety of compelling ethical views. It is relatively straightforward to use this framework to express a single view that deviates from the classical utilitarian model. However, as I will discuss in the course of this chapter, if one attempts to construct an axiology that expresses multiple such views simultaneously, it can become difficult to represent the individual views faithfully and there is serious risk of the axiology recommending foreseeably self-defeating courses of action⁶.

⁶ I note here that in this thesis I will omit discussion of variable populations. When applied to fixed populations, axiological theories generate fairly sensible results. However, if axiologies are used to evaluate choices between outcomes that contain different numbers of people, they consistently produce extremely unintuitive results (Derek Parfit (1987) established a number of pioneering results in this field). In fact, it has been proven by Gustaf Arrhenius (2000) that *no* axiology can avoid violating at least one extremely plausible desiderata when applied to scenarios with variable population. The issues surrounding variable population are extremely interesting, but to discuss them here would take me too far afield, so I will limit myself to cases in which all outcomes under consideration contain the same number of people.

III. Prioritarianism

I begin my discussion with Prioritarianism, which makes use of a particular kind of aggregation function in order to express special concern for the worst off. It represents the belief that an increase in wellbeing for someone who is already well off is less morally significant (i.e. has less value) than an equivalent increase in wellbeing for someone who is comparatively badly off⁷. The distinctive feature of Prioritarianism is that it holds that improvements in a person's wellbeing can be said to provide a diminishing marginal contribution to overall goodness. If an agent ascribes diminishing marginal value to some good, she judges that the quantity of that good is not proportional to its contribution to overall goodness, but instead, the more of the good there is to begin with, the less overall goodness will be increased by an increase in the quantity of the good. For example, the importance people ascribe to having a particular amount of money often behaves like this; generally speaking, the more money one has, the less important it is to have a given amount more⁸.

The standard Prioritarian aggregation function differs from classical utilitarianism in that it applies a weighting function to each person's lifetime wellbeing scores before summing them. Prioritarianism can be represented by an expression like the one below.

(2)

$$G = \sum_i P \left(\sum_{t \in N(i)} w(i, t) \right)$$

Expression (2) is the same as the classical utilitarian formulation I gave earlier (1), except that it applies the function P to each person's lifetime wellbeing (the sum of all of their momentary wellbeing levels), and then sums the outputs of this function P for each person to determine overall goodness. The function P is chosen to have a

⁷ As Broome (2015) notes, this view has its origins in economic theory (for example see Atkinson and Stiglitz (1980)), but it has since been adopted and argued for by philosophers, most notably Derek Parfit (1995; 1997; 2012).

⁸ Barring Threshold cases as discussed in Frankfurt (1987).

slope that is always positive but becoming ever flatter, so that a higher wellbeing score always contributes more to overall goodness, but improvements to someone's wellbeing matter less the higher that person's wellbeing is to begin with⁹. When the relationship between the quantity of a person's lifetime wellbeing and the contribution that person's wellbeing makes to overall goodness is plotted, the graph is concave: it curves downwards as shown in the figure 2 below.

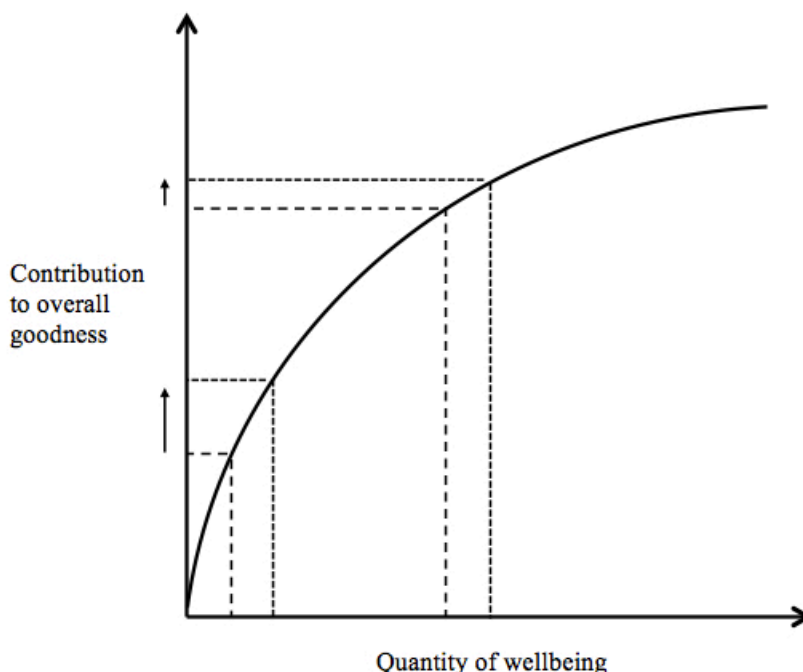


Figure 2. Concave function used to express Prioritarianism

As indicated by the sets of dotted lines intersecting the curve, a same-size increase in quantity of wellbeing will deliver a lesser increase in overall goodness when the starting quantity is greater. This should encourage actions that improve the wellbeing of the worst off more than actions which improve the wellbeing of those who are already well off.

⁹ This shape can be achieved a number of ways. One possible approach would be to make $P(x_i) = \log(x_i)$ where x_i is person i 's lifetime wellbeing score.

In general, Prioritarianism delivers similar results to concern for equality, but it has a number of distinguishing features that can be advantageous. One of the advantages of Prioritarianism is that there is comparatively little uncertainty over how it should be formally represented in axiology. For example, in his book *Inequality* (1993), Larry Temkin demonstrates that there are numerous plausible formulations of concern for equality, but they differ in their evaluations of which types of distributions are more equal and it is often not intuitively obvious which formulation is giving the *right* evaluation. By contrast, there is general agreement regarding how Prioritarianism should be represented in an axiology. Another difference between Prioritarianism and concern for equality is that some versions of the latter are famously vulnerable to what Derek Parfit (1997) calls ‘The Leveling Down Objection’; briefly, the objection is that some versions of egalitarianism are committed to saying that if someone who is relatively well off is harmed so that their wellbeing level is more similar to that of others, then there is at least one way in which this is a good thing, because it leads to a state of affairs with greater equality. Many people regard it as absurd to suggest that an outcome is in some ways made better if at least one person is harmed and no one is made any better off. An advantage of Prioritarianism is that it expresses concern for the worst off without generating this implication. The reason Prioritarianism avoids this result is that unlike egalitarianism, it is not at all concerned with people’s relative levels of wellbeing. It bases its concern for the worst off on people’s absolute levels of wellbeing; the higher a person’s wellbeing, the less good it is to give them extra, regardless of how they fare relative to others. This means according to Prioritarianism, the contribution a person’s wellbeing makes to overall goodness can be determined independently of any other person’s wellbeing, whereas this is not possible on most egalitarian accounts.

In light of these advantages, and because my results are more straightforward to demonstrate using Prioritarianism, I will use Prioritarianism as the focus of my argument. However the self-defeating recommendations that result from combining Prioritarianism and discounting will also be generated if common formulations of concern for equality are combined with discounting.

Because Prioritarianism uses lifetime wellbeing as input it has the formal characteristic of making times interdependent for the purposes of evaluation. By that I mean it makes it impossible to determine the moral significance of the wellbeing a person enjoys at one moment in time without referring to the distribution of wellbeing across other times in that person's life. According to Prioritarianism, the amount that the level of wellbeing in year t of a person's life will contribute to overall goodness is dependent on the wellbeing levels at other times in that person's life. So if that person's wellbeing is very low throughout the rest of their life, the wellbeing in year t will contribute more to overall goodness than it would if that person's wellbeing were very high throughout the rest of their life. As I will go on to explain, it is this feature of Prioritarianism that can cause it to produce inconsistent evaluations when combined with time discounting.

IV. Discounting

Time discounting is another modification on the classical utilitarian framework, one that makes the value of wellbeing depend upon a non-wellbeing factor: the time at which that wellbeing occurs. Time discounting is a way of expressing the view that events diminish in significance the more temporally remote they are from the present moment. On this basis, the overall goodness of a state of affairs is more strongly influenced by events that occur near to the present moment than events that occur relatively further into the future or past.

Two main kinds of argument are used to defend discounting: arguments that future wellbeing is intrinsically less valuable than wellbeing nearer to the present and arguments that make no particular claim as to the intrinsic value of wellbeing at different times, but which hold that discounting must be incorporated in an axiology to avoid absurd or undesirable consequences. Arguments in the former category are often labeled as advocating 'pure time discounting' or 'pure time preference'. For convenience, I will refer to the latter category of arguments as 'pragmatic time discounting'.

One example of an argument for pure discounting is the claim that it is permissible to value the wellbeing of people differently depending on one's degree of

connectedness to them. So just as it might be thought to be acceptable to have greater concern for people with whom we have a special relationship (family, friends, perhaps fellow citizens), so too it is permissible to have a greater degree of concern for those temporally near to us than those in the distant future ¹⁰.

A common argument for pragmatic time discounting is that axiologies that do not incorporate some form of discounting often generate absurd recommendations because they hold that goodness can be maximized by the present generation sacrificing almost all of its resources for the sake of future generations. The reason for this is that by investing resources now they will typically produce greater returns in the future. If wellbeing increases with consumption of resources and future people's wellbeing is treated as being equally important as present people's wellbeing, then it will generally be more beneficial to invest the present generation's resources so that they can provide a greater improvement in the wellbeing of future generations. Kenneth Arrow (1999) calculates that under fairly plausible circumstances, an axiology without discounting would require us to save over two thirds of our income for the benefit of future people; many regard this conclusion as unacceptable.

There are many other arguments that have been advanced to defend the use of discounting, but I will postpone my discussion of them until Chapter 3. For now it is sufficient to note that although there are a range of arguments used to defend discounting, they typically all employ a similar mathematical representation to express this view. My concern at this stage is with the consequences of combining that mathematical representation with the mathematical representation of Prioritarianism.

Many philosophers are highly skeptical of discounting. It is often regarded as axiomatic that wellbeing should count equally regardless of when it occurs ¹¹. As

¹⁰ This view is defended by Kenneth Arrow (1999).

¹¹ Many philosophers, both historical and contemporary, have argued that time impartiality is self evidently the correct view to take. See, for example (Sidgwick, 1890; Ramsey, 1928; Pigou, 1932; Broome, 2008; Dasgupta, 2008; Dietz, Hepburn, & Stern, 2008).

Hilary Greaves (*n.d.* p. 13) summarizes “but of *course* (runs the thought) the value of *utility* is independent of such locational factors: there is no possible justification for holding that the value of (say) curing someone’s headache, holding fixed her psychology, circumstances and deservingness, depends upon which year it is” ¹².

With this in mind it is worth noting that I make no claim that discounting *ought* to be used in axiology. My interest is in demonstrating the problems that can arise if discounting is expressed together in an axiology with certain other views such as Prioritarianism. Even if one is firmly against discounting, these results may be of interest. Firstly, because discounting is in fact used in most practical applications of axiology for policy making, so it is important to understand its implications and limitations ¹³; and secondly, because difficulties in combining discounting with other views could potentially constitute a further argument against discounting.

Time discounting is typically represented by using a function that weights the value of occurrences at each time such that occurrences nearer to the present moment are weighted more heavily in the determination of overall goodness. The standard assumption in the literature on discounting is that a function for determining overall goodness should take the ‘discounted-utilitarian’ form shown in expression 1 below. Greaves summaries the standard discounting process as follows:

The overall value of a state of affairs is computed by (1) calculating the amount of well-being present at each time t , by summing momentary well-being levels $w(i,t)$ across all individuals i who are alive at time t ; (2) ‘discounting’ the wellbeing at each time t by a factor $[D(t)]$ that represents how important well-being at t is, relative to well-being at other times; (3) summing the resulting discounted quantities across time (*n.d.* p. 5).

¹² I will discuss a number of other objections to discounting in Chapter 3.

¹³ Discounting is routinely used in policy assessment for areas such as climate change mitigation and healthcare distribution. See for example Dan Brock (2004) on the use of discounting in public health policy and Greaves’ (*n.d.*) review of the role of discounting in evaluating possible climate outcomes.

(3)

$$G = \sum_t D(t) \sum_{i \in N(t)} w(i, t)$$

Note that this is similar to the function (1) expressing classical utilitarianism except for two differences. One, rather than aggregating each person's lifetime wellbeing first, this approach aggregates the wellbeing across all people at each time first. The utilitarian and Prioritarian expressions incorporate $\sum_{t \in N(i)} w(i, t)$, which takes the sum of the momentary wellbeing levels for all values of t in person i 's life; by contrast, the discounting expression uses $\sum_{i \in N(t)} w(i, t)$, which sums the momentary wellbeing levels of each person i that exists at time t . The second difference is that rather than simply summing the results of this first aggregation, it multiplies the total wellbeing across all people at a given time by a discount factor, which is determined uniquely for each time by the discount rate function. Then after the total wellbeing levels for each time have been multiplied by their respective discount rates, the resulting values are summed to determine overall goodness¹⁴.

The function $D(t)$ outputs a specific value for each time t . This output is the *discount factor* for wellbeing at time t . It always takes a value between 0 and 1. This number is multiplied by the wellbeing level at the time in question to give the effective value of the wellbeing at that time. The discount factor represents the threshold for the amount we are willing to sacrifice now to get an extra unit of wellbeing at time t . Suppose we are only willing to sacrifice half a unit of wellbeing now to get an extra unit of wellbeing 10 years from now, then the discount factor for $t = 10$ would be set

¹⁴ I should note that the expression given in (3) is actually a slight variant on the standard approach to discounting. Economists generally work in continuous time, thus the equation would contain an integral $\int dt$ rather than a sum \sum_t . Using the integral approach would give the equation $V = \int dt \cdot D(t) \cdot \sum_{i \in N(t)} w(i, t)$. The simplified approach I am using requires splitting time into discrete periods and assigning well-being scores and discount factors representative of these time intervals. To use the discrete form as I am is common practice (Greaves n.d.), though it is slightly less elegant. I will continue to use the discrete form as it will make the exposition of my argument more straightforward and no substantive issues turn on this choice.

to 0.5. A low discount factor results in the relevant event being substantially devalued, whereas a discount factor close to one makes little difference to the event's original significance. Generally this factor declines with time – the later the time, the lower the amount we are willing to sacrifice for one extra unit of wellbeing at that time.

The function D often takes the following form: $D = \exp(-rt)$, where r is the *discount rate*¹⁵. The discount rate determines the speed at which the discount factor will decline as the value of t moves further away from zero. Because of the exponential relationship between the discount rate and the discount factor, small changes to the discount rate can lead to very significant changes in the amount that future wellbeing is discounted. If the value for the discount rate is set to be relatively high, then the axiology will use a discount factor close to zero for times in the far future and therefore will tend to recommend projects that favour the short term. By contrast, if an axiology uses a low discount rate that will mean that the values of future events are multiplied by a discount factor closer to one, so the axiology will be more likely to favour projects that make sacrifices in the short term for the sake of long term benefits.

When using time discounting, it is an open question whether *past* times should be valued in the same way as future times. All prominent arguments for time discounting hold that events are less valuable the further they occur into the future. A further question that needs to be answered is how should we value occurrences at different times in the past? They might be considered equally significant as an event a similar amount of time into the future, or less significant (perhaps having no significance at all), or more significant.

These different views on the relative importance of past events can be represented by selecting a discount rate for past events independently of the discount rate that will be used for future events. Typically, the present moment is set to $t = 0$, future times use positive values for t , and past times use negative values for t . Inserting

¹⁵ This could also be written as $D = e^{-(rt)}$. e is a mathematical construct for modeling exponential phenomena

these values into the function $D(t)$, this means that if the same discount rate is used for both past and future times, future times will decrease in effective value the further they are into the future, whereas past times will increase in effective value the further they are into the past, thus making events more and more significant the earlier in time they occur. In Broome's (2004) terminology, *relative discounting* and *neutral discounting* are two different ways of representing the relative value of times. Neutral discounting applies the same discount rate to all times, thereby making past events more important than future or present events. By contrast, relative discounting applies different discount rates to past wellbeing and future wellbeing, allowing a range of possibilities for what the relative value of past and future events will be. Below is a diagram depicting the ways that neutral and relative discounting could influence the contribution of a person's wellbeing to overall goodness. Note that the specific amounts that the contribution to wellbeing is altered will depend on the discount rates being used, but these general patterns are distinctive to relative and neutral discounting; neutral discounting will always make past times more significant than present or future times, whereas relative discounting has the option to make past wellbeing equally as valuable as future wellbeing.

The difference between the ways that relative discounting and neutral discounting alter the contribution of wellbeing is illustrated in figure 3 below. The person whose wellbeing is depicted in figure 3 has a uniform level of wellbeing across all times shown (prior to discounting). The person's actual level of wellbeing can be seen at time = 0 (the present moment), when discounting has no effect. At past times (negative values for time), relative discounting diminishes the contribution of the person's wellbeing, whereas neutral discounting amplifies its contribution.

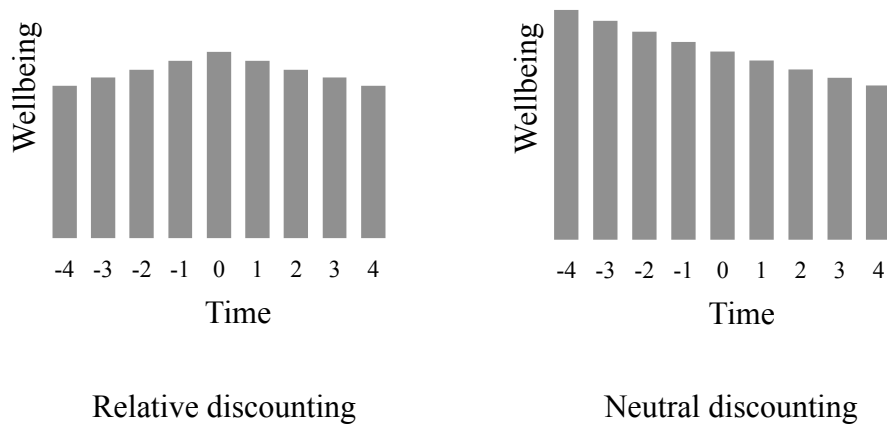


Figure 3. Hypothetical contribution that a person's wellbeing at different times makes to overall goodness after applying neutral discounting or relative discounting.

In order to represent a view that past times matter as much as or less than future times, it is standard to use a different discount rate for past times. For instance, if past times are evaluated using a *negative* discount rate that is otherwise equivalent to the (positive) future discount rate, then an event t years into the past will be discounted as much as an event t years into the future. By making the discount rate for all past times negative and very large, it would be possible to make past times count for almost nothing ¹⁶.

¹⁶ It may seem that the choice of $t = 0$ as the present time and past times as negative values of t is somewhat arbitrary, and thus possesses an undue amount of influence here. There are alternate ways of representing both times and the discount rate, but as far as I am aware none of them offer significant advantages over this approach. For example, one could do away with negative values of t by setting $t = 0$ as the beginning of time, and make all times (values of t) after that positive. This would have the effect of making past events more significant than present and future events, just as neutral discounting does. Alternatively, negative values of t could be treated as positive values by altering the function D to be $D = \exp(-|r|t)$. The vertical lines on either side of the r indicates the absolute value, i.e. the (positive) distance the number is from zero. This approach is effectively the same as using negative r for past times and positive r for future times.

Neutral discounting delivers some highly unintuitive results. One of these is that it contradicts commonly held preferences regarding the timing of good and bad experiences relative to the present moment. In general, people prefer that bad experiences be in the past rather than in the future and that good experiences be in the future rather than in the past. Suppose that at some point in time you are to undergo a certain painful experience such as a root canal. Would you prefer to be walking out of your dentist's office, with the painful experience over and done with, or to be walking in, with the bulk of your pain still to come? I think most people would prefer that pain be in their past, even though both scenarios involve experiencing the same amount of pain. They just differ in where that pain is located relative to the present moment ¹⁷. Neutral discounting gives the opposite response to this question. It entails that past pain is more significant than future or present pain, and therefore that it is better for some painful event to be in one's future than in one's past.

Another disadvantage of neutral discounting is that it can recommend that we work to benefit people who lived in the distant past, rather than people who are alive today. It may be that people can be made better or worse off by events that occur after their death. For example, Aristotle argued that after a person had died she could still be benefited or harmed by the fortunes of her descendants ¹⁸. Broome (2004) points out that, if this is so, neutral discounting could recommend devoting significant resources towards benefiting people who lived in the distant past, since their wellbeing would count for so much more than that of present or future people. For example, suppose Shakespeare benefits when people honour his legacy, an axiology that used neutral discounting might rank honoring Shakespeare's legacy as more morally important than making large improvements in the welfare of people who are alive today.

¹⁷ These preferences, and arguments for and against their rationality are discussed at length in Derek Parfit (1987) and Caspar Hare (2007; 2008; 2013). This particular example is borrowed from Hare (2013).

¹⁸ *Nicomachean Ethics*, book 1, chapter 10, in Crisp (2014).

This problem is somewhat similar to the over-demandingness objection that is raised in defense of discounting¹⁹. As discussed above, without discounting, a typical axiology would judge the optimal scenario to be one in which the current generation gave an enormous share of their resources to benefit people who live in the future. Discounting is able to counter this effect by making benefits to future people count for less in the evaluation of outcomes. Neutral discounting generates this problem in reverse though, rather than recommending the present generation give excessively to benefit future people, it might recommend giving excessively to benefit past people. However, in this case the problem is more severe than the standard over-demandingness objection, because even if limits were placed on how much a generation or an individual could be required to aid others, it would still be somewhat absurd to give greater priority to benefiting people who existed a given amount of years into the past than to people who will exist that amount of years into the future. What could be special about past people that would make it more morally important to benefit them than future people? Intuitively it seems clear that benefitting past people should either count for the same or count for less than benefitting future people, either way past people should not count for more.

Relative discounting is able to avoid the problem of placing excessive weight on past events. One approach to relative discounting is to use a discount factor of zero for all past times²⁰. I will call this view *complete future bias*, since it holds that only present and future events are relevant in moral deliberations. Once an event moves into the past it would have zero bearing on the determination of the overall goodness of a state of affairs. This approach succeeds in avoiding the Shakespeare problem, however, as I will discuss in Section VIII, there are serious problems that arise for

¹⁹ Greaves (*n.d.* p. 15).

²⁰ Setting the discount factor as zero for past times would require more than a change to the discount rate. However it could be achieved with modifications to the function R such that:

$$D(t) = \begin{cases} t \geq 0 = \exp(-rt) \\ t < 0 = 0 \end{cases}$$

This expression states that, if the value of t is 0 or greater, $D(t) = \exp(-rt)$, but if the value for t is less than zero, $D(t) = 0$.

an axiology that expresses complete future bias in conjunction with Prioritarianism or concern for equality.

A less extreme way of interpreting relative discounting is to use discount factors for past times that are symmetrical with the discount factors used for future times, such that events become less significant the further they are from the present, regardless of whether they occur earlier or later than the present moment. This approach is compatible with the common preferences regarding the timing of good and bad experiences relative to one's present experience and it avoids the problem of prioritizing past people over present people. Thus there are good reasons to prefer relative discounting to neutral discounting. In practice, when discounting is discussed it is often assumed to be relative discounting, and most advocates of discounting are in favour of relative discounting²¹.

Unfortunately, relative discounting is not without its problems, notably, relative time discounting leads to inconsistent evaluations over time. When events move from the present into the past, their value will be altered relative to future events. Preston Greene and Meghan Sullivan (2015) have demonstrated that this can lead an agent to regret previous decisions. For example, this sort of discounting might tell me that the prospect of eating a cookie tomorrow is more valuable than having eaten a cookie two or more days ago. So, no matter when I eat a cookie, afterwards, I will eventually regret the decision and wish that I had saved the cookie for a later day.

If I can choose when in time a good occurs, no matter what time I choose, after that good has become past, relative time discounting will cause me to judge that the good would be more valuable if it were in my immediate future. This inconsistency is bothersome, but since we can't change the past, it cannot lead an agent to reverse prior decisions so it does not pose any significant problem to practical decision-making. In her discussion of relative discounting, Greaves emphasizes that the inconsistency generated by relative discounting does not lead to practical difficulties. She writes:

²¹ Greaves (*n.d.*).

It is worth noting well, however, that the only sort of inconsistency that can result from the discounting structure in question is the phenomenon of foreseeable regret that we have seen in this example. This is arguably not as bad as foreseeable backtracking, in which an agent decides at time i to pursue a certain course of action at time $i + 2$, while recognizing that at time $i + 1$ she will no longer think that this course of action is best, and so will attempt to reverse her decision (*n.d.* p. 14).

By itself, relative discounting does not recommend sequences of actions in which one will seek to reverse a prior decision, however, as I will argue in Section X, some ways of jointly representing relative discounting and Prioritarianism *do* recommend such problematic reversals. By recommending this kind of reversal, an axiology can engineer an outcome that it would ultimately regard as disadvantageous by its own lights, regardless of from what temporal perspective the outcome was evaluated. Before developing this claim, I need to discuss some of the ways in which an axiology can be designed to express both priority for the worst off and temporal discounting.

V. Combining Discounting and Prioritarianism

In her forthcoming review article, Greaves (*n.d.*) describes three variations on the standard discounting function that incorporate Prioritarianism. For convenience, I will call these Time-Slice Prioritarianism, Uniformly-Discounted Prioritarianism, and Post-Discounting Prioritarianism. In what follows I will argue that all of these approaches face problems that make them unsatisfactory as representations of the two views in question. In Chapter 3 I will introduce a fourth approach that incorporates both of these views and which is potentially less vulnerable to objections.

Time-Slice Prioritarianism

This approach is the same as the default discounting function described above, except a concave function P is applied to each person's wellbeing at each time t before these values are summed across times and discounted (4).

(4)

$$G = \sum_t D(t) \sum_{i \in N(t)} P(w(i, t))$$

This function P is of the same kind as the one used in the Prioritarian expression (2) I described in section III. Overall value is computed by (1) calculating the amount of Prioritarian adjusted wellbeing present at each time t by applying the concave function to the momentary wellbeing level $w(i, t)$ of each individual i who is alive at time t , and then summing the outputs of the concave function; (2) ‘discounting’ the wellbeing at each time t by a factor $D(t)$ that represents how important wellbeing at t is, relative to wellbeing at other times; (3) summing the resulting discounted quantities across time.

This approach notably differs from standard Prioritarianism in that it involves applying the concave Prioritarian function to the momentary wellbeing at each unit of time within a person’s life and then summing the results, as opposed to standard Prioritarianism, which first sums the wellbeing across a person’s life and then applies a concave function to the total. In other words, it ascribes diminishing marginal value to momentary wellbeing rather than to lifetime wellbeing.

Time-Slice Prioritarianism seems to me to be an unsatisfactory replacement to standard Prioritarianism for two reasons: it packages concern for the worst off with additional moral judgments that are not necessarily agreeable, and it often fails to adequately represent concern for the worst off.

In addition to expressing concern for the worst off, Time-Slice Prioritarianism also favours outcomes in which the distribution of wellbeing is more rather than less equal between the times within a person’s life. Because the Prioritarian function is applied to each time in a person’s life independently, for any given lifetime quantity of wellbeing, that wellbeing will contribute more to overall goodness if it is spread evenly across the person’s life than if there is variation in the wellbeing level over time. Concern for having an equal distribution of wellbeing across the times within each person’s life is a value distinct from concern for the worst off and it requires an

independent justification. As such, this view is at risk of being less defensible than standard Prioritarianism.

Even if concern for equality between the times in a person's life is justified, this view faces the more serious problem of sometimes failing to adequately express concern for the worst off. Consider the distributions of wellbeing in the following two worlds. In World 1, Person A and Person B alternate being 1 unit of wellbeing worse off than the other. By contrast, in World 2, Person B is 1 unit worse off than Person A at every time (different times are indicated by T1, T2 etc.).

World 1

	T1	T2	T3	T4
Person A	1	2	1	2
Person B	2	1	2	1

World 2

	T1	T2	T3	T4
Person A	2	2	2	2
Person B	1	1	1	1

Using only snapshot valuations one cannot tell these two worlds apart, they both seem identical with respect to equality. At each moment in time, one person's level of wellbeing is 1 lower than the other person's. Time-Slice Prioritarianism is insensitive to the cumulative difference made by the fact that in World 2 it is always Person B who has one fewer units of wellbeing whereas in World 1, Person A and B alternate being worse off or better off. In World 1, Person A and B have the same total wellbeing across all times, whereas in World 2, Person A has 4 more total wellbeing than Person B. Despite this discrepancy in total wellbeing, Time-Slice

Prioritarianism would rank these two worlds as being equally good. This is clearly a failure to display concern for the worst off.

Concern for the worst off ought to include not just concern for the worst of at any given moment, but also concern for the worst off over time. Indeed, of these two elements, the latter seems more important than the former. Many egalitarian philosophers (broadly construed) have reached similar conclusions. It is commonly held that a person's overall wellbeing across the whole of their life is the appropriate unit for determining how outcomes fare with respect to equality ²².

Uniformly-Discounted Prioritarianism

Uniformly-Discounted Prioritarianism is described by the following expression (5):

(5)

$$G = \sum_i D(b_i) P \left(\sum_{b_i < t < d_i} w(i, t) \right)$$

Here b_i and d_i are the birth and death dates of individual i respectively. This approach uses the standard Prioritarian method of applying a concave function to each person's total lifetime wellbeing score. The outputs of these Prioritarian functions are then discounted using the time of the person's birth to determine the relevant discount factor (since the function D operates on b_i rather than t). The results of multiplying the outputs of the P functions by their respective discount factors are then summed to determine overall goodness. It is worth noting that in this equation the discount factor is determined using the person's birth date, but it could be based on the time of the person's death, the middle of her life, or any time during her life for that matter. What characterizes this approach is that it uses a single time

²² See, for example, Rawls (1971) and Dworkin (1981).

in a person's life as input to the function D to determine the discount factor that will be applied to all times in that person's life.

I find that Uniformly Discounted Prioritarianism is also not very appealing. On this approach, all of the wellbeing across a person's life is subject to the same discount factor. Broome (1992, p. 57) criticizes discounting a person's lifetime wellbeing at a single rate on grounds of arbitrariness. He points out that there seems to be: "no reason why the wellbeing of a person should be tied down to the date when she is born rather than, say, the date when she dies". A further problem with this approach is that it would favour unequal distributions among people living contemporaneously, since, at any given time, the oldest living people would have a higher weight assigned to their wellbeing than the youngest living people due to their having an earlier birth date used to determine the discount factor. Arguably, this inequality is acceptable, since although young people would be disadvantaged as a result, they would later be advantaged when they in turn became the older generation. Nonetheless, it seems problematic that an axiology should consistently favour an unequal distribution among people living at any given moment, especially given that Prioritarianism is generally intended to favour outcomes with a high degree of equality.

Post-Discounting Prioritarianism

This approach applies pure discounting directly to momentary wellbeing at each time before that the resulting scores for each time are summed to compute a discounted version of each person's total lifetime wellbeing (6). A concave function is then applied to these discounted lifetime wellbeing scores and the results are summed to determine overall goodness.

(6)

$$G = \sum_i P \left(\sum_{b_i < t < d_i} D(t) \cdot w(i, t) \right)$$

This approach is importantly different from the previous two in that it applies time discounting to wellbeing scores *before* they are input into the Prioritarian function.

As I will go on to explain, this feature causes Post-Discounting Prioritarianism to produce highly undesirable results.

Earlier I discussed the regret phenomenon that arises from all forms of relative discounting. Greaves (*n.d*) noted that although this brings about inconsistencies in how choices are evaluated over time, the inconsistency is not particularly problematic since the evaluation of the relative goodness of an occurrence only changes once that occurrence is in the past, so the change in evaluation can not prompt a change in behavior. If Post-Discounting Prioritarianism uses relative discounting, then it is vulnerable to a more serious kind of inconsistency. It can result in scenarios where the axiology changes its recommendation of an optimal future course of action over time, even when the predicted future outcomes do not change at all. This means that at one time the axiology might recommend allocating future wellbeing in a certain way, and then at a later time (but still prior to when that future wellbeing would occur) it would recommend reversing decision and allocating that future wellbeing in a different way. By instructing an agent to reverse their decisions about undertaking a course of action, such an axiology can produce suboptimal outcomes. In particular, this kind of preference reversal can cause the axiology to recommend a series of actions that will result in an outcome that it regards to be worse than what would happen if none of the actions were taken. In the decision theory jargon, this phenomenon is often called a “money pump”. In what follows, I will demonstrate how this effect occurs²³.

VI. The Money Pump

Before proceeding I must acknowledge a debt. My argument in this section follows a similar pattern to the one made by Tom Dougherty (2011) wherein he demonstrates that an agent who is both risk averse and cares more about future events than past events is liable to be money pumped. Both arguments explore the combination of a concave value function and a factor that modifies the values of events according to their proximity to the present moment. My argument

²³ The concept of a money pump first appeared in Davidson, McKinsey, and Suppes (1955),

demonstrates that a similar pump can result from different views than the ones Dougherty discusses.

I will now give a general overview of how these two views combine to produce a what is termed a money pump. Suppose I am making decisions on the basis of an axiology that incorporates Prioritarianism and relative time discounting. It discounts past times and future times so that they are each discounted the same amount for each year they are removed from the present moment. Suppose also that there are two people who have the same total wellbeing and live during the same times, but whose wellbeing over time is skewed in different directions. Person A has more wellbeing near the beginning of her life, whereas Person B has more wellbeing near the end of his life. These wellbeing distributions are depicted in figure 4 and figure 5 below, respectively.

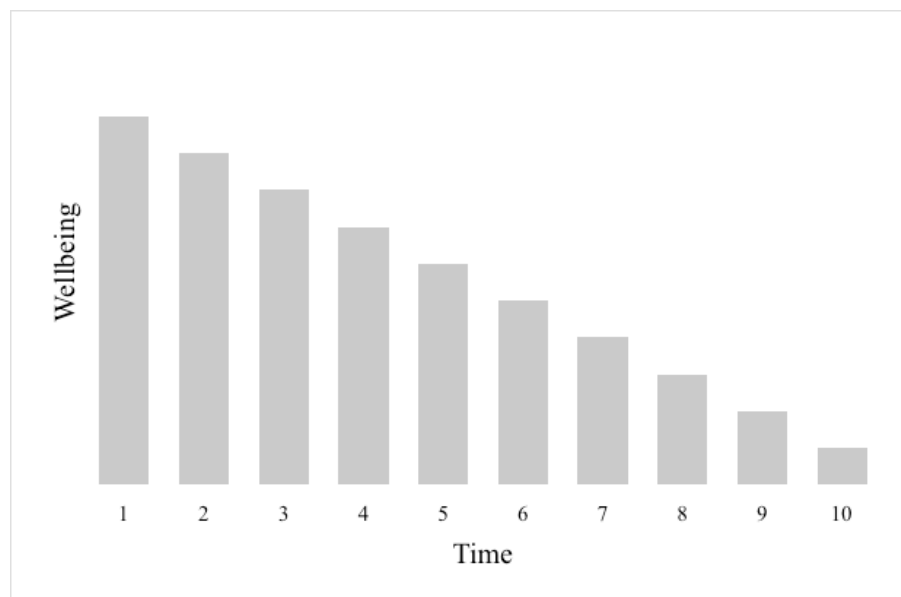


Figure 4. Person A's undiscounted wellbeing

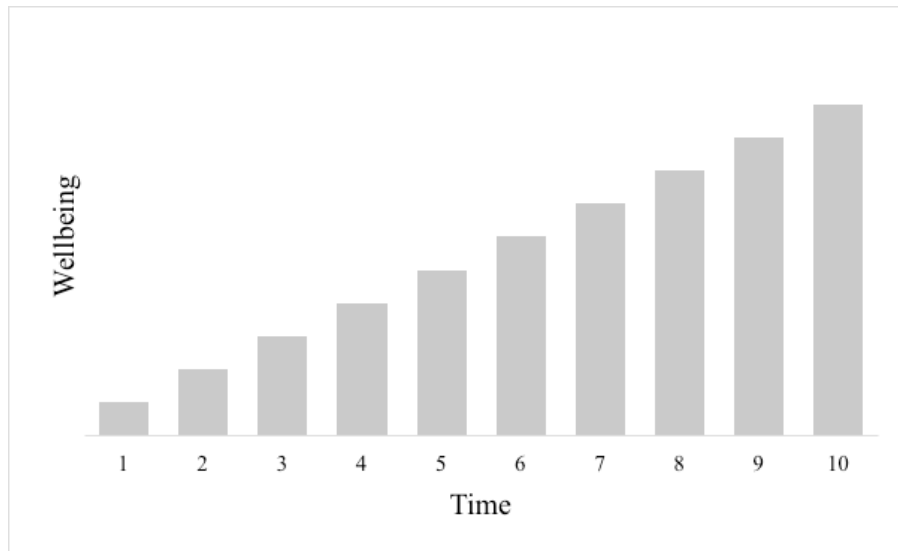


Figure 5. Person B's undiscounted wellbeing

Suppose discounting at 10% per year from the perspective of time T1. Figure 7 shows the effect of this discounting on Person A. The darker region represents how much effective wellbeing is left after discounting. The lighter region shows how much of the original wellbeing score is effectively removed by discounting. Likewise, Figure 8 shows the effective discounting from the perspective of T1 on Person B.

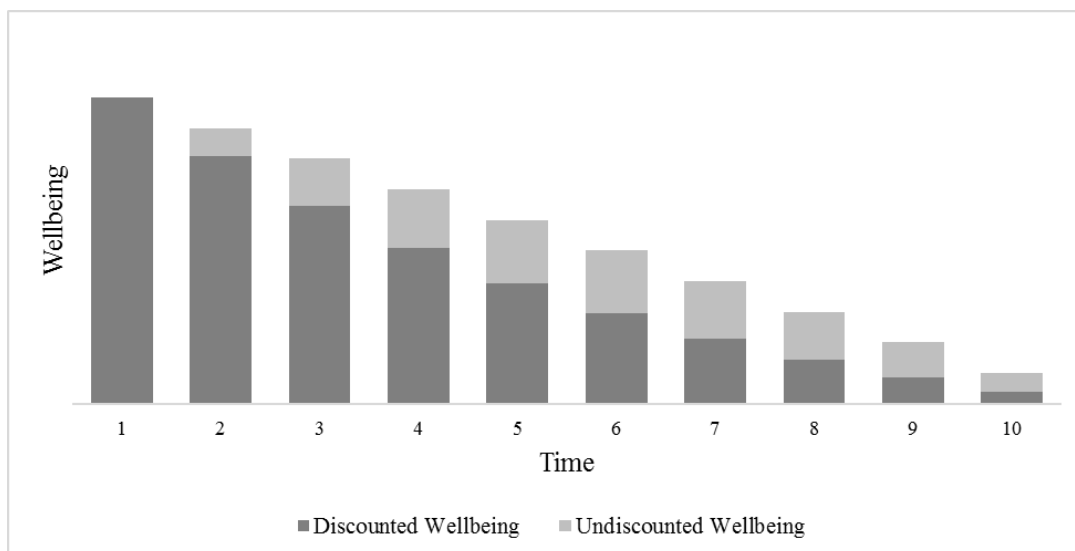


Figure 6. Person A's discounted wellbeing from the perspective of T1

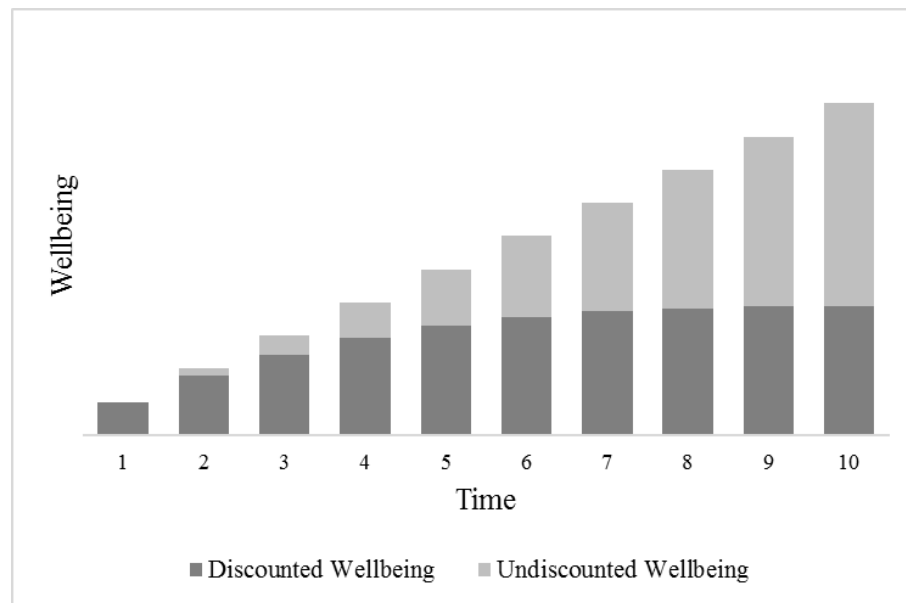


Figure 7. Person B's discounted wellbeing from the perspective of T1

From the perspective of earlier times, person B, whose wellbeing is skewed primarily towards the end of his life will be more heavily affected by relative discounting since the bulk of his total wellbeing is remote in time. As a result, his total discounted wellbeing will be lower than that of person A.

At later times the converse is true. The person whose wellbeing is skewed towards the beginning of his or her life will be more heavily affected by discounting. Figures 9 and 10 depict the impact that discounting from the perspective of T9 has on the contribution made by these lives.

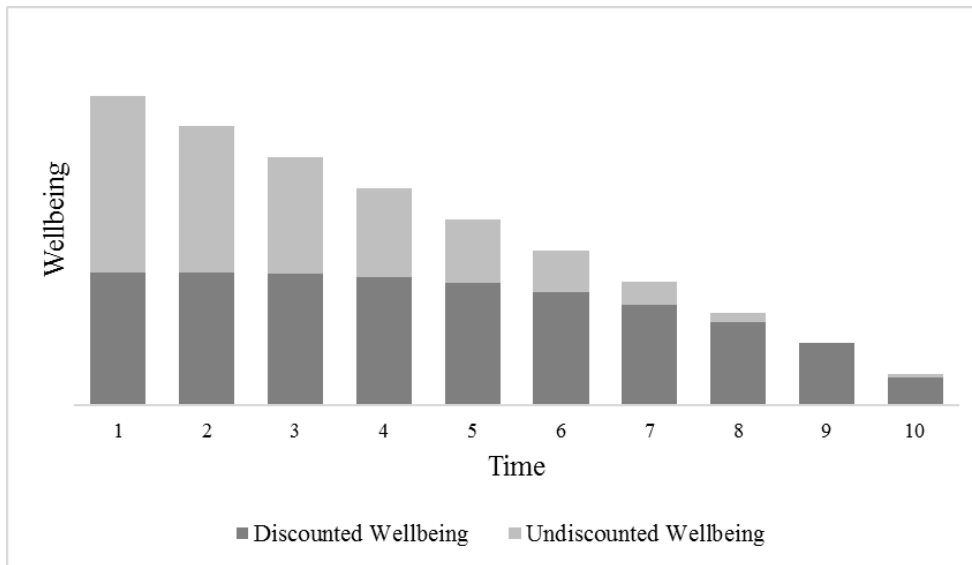


Figure 8. Person A's discounted wellbeing from the perspective of T9

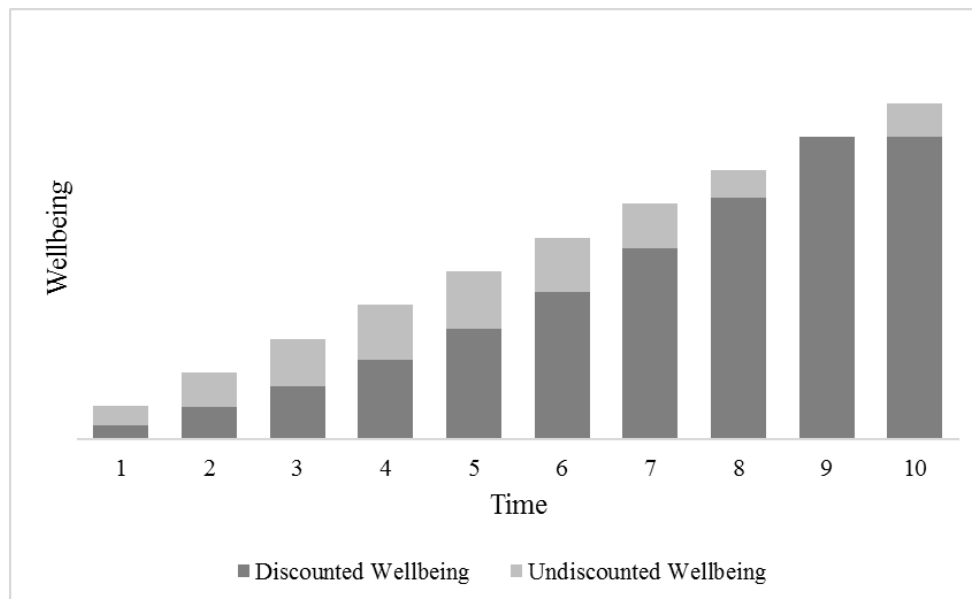


Figure 9. Person B's discounted wellbeing from the perspective of T9

As can be seen, from the perspective of T9, Person A is more substantially affected by discounting than person B. So when comparing two people with wellbeing skewed across time in different directions, an axiology that uses relative time discounting will switch its evaluation of which life contributes more to goodness depending on whether it evaluates from an earlier or later perspective in time. From the perspective of the earlier time, Person A's lifetime wellbeing contributes more to overall goodness than Person B's does. Conversely, from the perspective of the later time, Person B's wellbeing contributes more to overall goodness.

Prioritarianism holds that it is always better to give a same size benefit to a worse off person than a better off person, other things being equal. The Prioritarian function is applied after discounting, so it will evaluate who is a better recipient of benefits based on discounted wellbeing scores. From the perspective of earlier times, it will determine that Person B is worse off and therefore a better candidate for receiving a benefit. But from the perspective of later times, Prioritarianism will produce the opposite result, it will determine that Person A is worse off, and that they are therefore a better candidate for receiving some benefit.

Suppose at T1 I have the option to enact a plan that will give a small benefit to one person at a time near the end of his or her life, but at the same time it will also result in a slight cost to the other person. Say at T10 one person will be slightly better off and one will be slightly worse off by the same amount. From the perspective of an earlier time, B is worse off so according to Prioritarianism it may be worth it to give him a small benefit at a small cost to A. Adhering to my axiology, which includes discounting and Prioritarianism, I decide to enact this plan.

Now suppose that at T9 I have the option to try to reverse my decision. I can remove the extra wellbeing I would have given to B and restore the wellbeing lost by A. From the perspective of T9 time I would prefer to make this change, since A now appears to be the worse off person. Prioritarianism dictates that this wellbeing would count for more if it were enjoyed by A than if it were enjoyed by B. However, suppose that I cannot make this change without some transaction costs. If I try to reverse my original decision, B will end up slightly worse off than they were before I allocated them a benefit, and A will not get back quite as much wellbeing as I originally took from them. Even if this is the case, my axiology might still instruct me to make the change, since according to Prioritarianism the benefit to the worse off person would provide more goodness than would be lost by harming the better off person. This difference in goodness might be sufficient to outweigh the loss in goodness resulting from the decline in total wellbeing. Supposing this is the case, my axiology will instruct me to reverse my earlier decision thus making A slightly better off, and B slightly worse off, but leaving them both slightly worse off than they would have been if I had not acted in the first place.

In this way, the combination of discounting and Prioritarianism can endorse a series of actions that result in an outcome that is worse than the status quo from all perspectives of evaluation. From the perspective of the early time, the first decision to allocate slight benefits and costs is better than the status quo, and from the perspective of later time, the inefficient reversal is better than accepting the results of the first decision. But according to the evaluations of the axiology, from both perspectives in time, it would be better if neither action had been taken than if both actions had been taken. The results of the two actions leave each person slightly worse off and make the outcome in no way better than the original status quo. This is a money pump.

VII. Why the Money Pump Occurs

The problem comes from the fact that Prioritarianism makes the evaluation of wellbeing across different times in a person's life interdependent. It holds that the moral significance of the wellbeing at a given time in a person's life depends on the quantity of wellbeing across all other times in that person's life. This means that discounted past wellbeing can bear on the decision of whom to allocate a benefit to in the future.

Relative time discounting causes a shift in the relative value of goods as they move from the future to the past. If we limited our evaluation to only the wellbeing levels that occur at future times, relative discounting would never cause a reversal in the evaluation of which person's wellbeing contributes more to overall goodness. However when past wellbeing levels are taken into account, the evaluation of which person's wellbeing contributes more *can* change depending on the temporal perspective from which the evaluation is made. In particular, if two peoples' wellbeing levels are each skewed in different temporal directions (so that one person is better off earlier and the other is better off later), one person's effective wellbeing will be more significantly reduced at earlier times, and the other's effective wellbeing will be more significantly reduced at later times.

By itself, the change in evaluation of wellbeing that has become past would not make a difference to the evaluation of future benefits or harms, but aggregation

processes that make times interdependent can cause this change in the relative value of past wellbeing to affect the moral significance of future wellbeing. In the case of Prioritarianism, a person will be deemed a better candidate for a future benefit if they have less lifetime wellbeing. The relative change in the value of discounted wellbeing that becomes past can lead to a change in the evaluation of who has less lifetime wellbeing and therefore a change in the evaluation of how future benefits and harms are best allocated. In this way, the combination of relative time discounting and aggregation processes that make the evaluation of times interdependent can generate a money pump.

VIII. How Alternative Combinations of Discounting and Prioritarianism Relate to the Money Pump

Only-Future Prioritarianism

Only-Future Prioritarianism uses the standard Prioritarian function, except that rather than apply this function to each person's overall lifetime wellbeing score, it applies it to the total wellbeing they each enjoy between the present moment and the end of their lives. Past wellbeing is ignored²⁴. By ignoring past wellbeing it exhibits an extreme form of future bias.

I think that Only-Future Prioritarianism fails to completely exemplify concern for the worst-off. If a person has fared poorly in the past, it seems that this should provide some reason to prioritize his or her interests in the present or future. Only Future Prioritarianism denies this. However, even if Only-Future Prioritarianism can be defended as satisfactory expression of concern for the worst off, it is not a viable alternative for the purposes of avoiding the money pump. This view does not generate a money pump in the example I described above, but it turns out to be capable of generating a money pump via a slightly different route. By ignoring past times, this view exhibits the same pattern as axiology that expresses both standard discounting and relative time discounting; at early times it can give priority a person whose wellbeing is concentrated towards the end of his life, and at later times it can

²⁴ This variation on Prioritarianism has some defenders in the public health care literature. Ottersen (2013) outlines the case for and against its use in that context.

reverse this preference and give priority to a person whose wellbeing is concentrated towards the beginning of her life.

Consider two lives with the following distribution of wellbeing over time:

	T1	T2	T3
Person A	4	2	2
Person B	2	3	2

From the perspective of T1, Only-Future Prioritarianism takes each person's wellbeing across all three times into account so it recommends allocating a benefit to person B rather than person A. But from the perspective of T2, this view only takes the wellbeing from T2 onwards into account, so it would now recommend giving a benefit to person A rather than person B. If, at T1 one has the option to enact a process that would redistribute some wellbeing at T3 from A to B, Only-Future Prioritarianism would recommend doing so. Suppose then that at T2 one has the option to inefficiently reverse this procedure, leaving both A and B slightly worse off at T3 than they would have been to begin with. Only Future Prioritarianism will recommend undertaking the reversal. By recommending this sequence of actions it will leave both people worse off than if no action had been taken. This is a money pump. Thus, Only-Future Prioritarianism is not a viable alternative to standard Prioritarianism if one is seeking to avoid money pumps.

Time-Slice Prioritarianism

Time-Slice Prioritarianism (described in section V) has the advantage that it is able to avoid money pumps. This view involves applying the concave Prioritarian function to each unit of time within a person's life and then summing the results, as opposed to standard Prioritarianism, which first sums the wellbeing across a person's life and then applies a concave function to the total. Because Time-Slice Prioritarianism applies the concave function separately to each time in a person's life, rather than to the total wellbeing across that person's life, it does not have the property of making times interdependent for the purpose of evaluation, so it will not generate a money pump.

Uniformly-Discounted Prioritarianism

Uniformly-Discounted Prioritarianism will also not generate a money pump because the discounting function is applied to the output of the Prioritarian function (as opposed to the other way around). The money pump occurs only if a function that makes times interdependent is applied to the output of the discounting function.

There are three features needed to generate the money pump, one is that the effective quantity of a good at a certain time can be altered relative to goods at other times when that good moves from being in the future to being past. The second necessary feature is that the effective quantity of past goods can influence the values of future goods. Relative discounting entails the first property, and concave functions that take goods across multiple times as their input (like standard Prioritarianism) entail the second property. Finally, not only do these two qualities need to be present to generate a pump, the functions that have these properties need to be applied in a certain order.

The money pump occurs when the discounting function is applied first, such that altered values for events that have become past are then used as input to the function which makes times interdependent; this has the flow on effect of making future events change their value as a result of the change in the value of past events. If the functions are applied the other way around, then any interaction between times is already resolved and accounted for before adjusting the effective quantities of goods based on the time at which they occur, so the changing values of past goods cannot have any impact on the contribution of future goods. Hence Uniformly-Discounted Prioritarianism, which applies the Prioritarian function first and then the discounting function, is not susceptible to the money pump.

Although this approach is able to avoid the pump, I nonetheless think it is an unsatisfactory way to represent concern for the worse off (as I discussed in section V). In Chapter 3, I will investigate how it might be possible to combine discounting and Prioritarianism in a way that does not suffer the same drawbacks as Uniformly-Discounted Prioritarianism and Time-Slice Prioritarianism, but which is still able to avoid the money pump due to applying the concave function prior to the discounting function.

CHAPTER 2

I. Introduction

Time Discounting and Prioritarianism both hold that the amount that a given instance of wellbeing contributes to overall goodness can depend on the way that other instances of wellbeing are distributed across times and or people. In other words, these views hold that moral significance of wellbeing depends not only on the quantity of wellbeing present, but also on patterns in the distribution of wellbeing. Because of this, Broome (2004) refers to the values expressed by these kinds of views as *pattern goods*. On Broome's account, pattern goods do not show up directly in the distribution of wellbeing. Instead, the contribution they make to overall goodness supervenes on the distribution of wellbeing; it depends on the way that wellbeing is distributed.

Another example of a pattern good is the contribution to goodness of the 'shape' of a person's wellbeing distribution over time; some people believe that it matters how wellbeing is distributed across a person's life. They think that it is better to have a life that begins badly and ends well than it is to have a life that starts well and ends badly. This value is sometimes referred to as concern for the 'shape of a life'²⁵.

Example of two lives with different 'shapes'

	T1	T2	T3
Person A	6	4	2
Person B	2	4	6

²⁵ Authors including David Velleman (2000), Michael Slote (1984), Larry Temkin (2011), and Joshua Glasgow (2013) have all defended versions of this view. See Dorsey (2015) for a recent article that summarizes much of the literature on this subject.

The above table shows two lives each of which contain the same total wellbeing but which have different distributions of wellbeing over time. Person A has more wellbeing near the beginning of her life and her wellbeing gradually declines over time. By contrast, Person B begins her life with relatively low wellbeing and her wellbeing increases over time. An axiology that took into account only the total quantity of wellbeing would rank these two lives as equally good, but an axiology that incorporated concern for the shape of a life might consider B's life to be better because the wellbeing level in B's life trends upwards over time whereas it trends downwards in A's life. In addition to Prioritarianism, discounting, and shape of a life, concern for equality can also be considered a pattern good ²⁶.

Pattern goods offer appealing ways of expanding upon the classical utilitarian model in order to express a greater range of ethical positions, so it is worth exploring the extent to which they can be successfully represented in quantitative axiologies like the ones I have been discussing. It is relatively straightforward to express a single pattern good using axiology ²⁷. However, the possibility of the money pump that I described in Chapter 1 demonstrates that incorporating multiple pattern goods in an axiology is potentially more fraught. Of the expressions I discussed that combined Prioritarianism and discounting, all of them either failed to faithfully represent one of the pattern goods in question, or advocated self-defeating courses of action in the form of a money pump. In this chapter I discuss some of the ways that pattern goods can be given mathematical representations and propose a series of criteria that must be met when representing multiple pattern goods. I claim that adhering to these criteria is necessary if an axiology is to avoid generating money pumps and other problems of this sort.

Section II of this chapter will contain a brief discussion of two different ways that a pattern good can be represented mathematically. These two styles of representation

²⁶ Other pattern goods that I will not discuss here include the value of longevity and the survival of the species – see Hirose (2015) for an investigation of how the value of longevity might best be mathematically represented.

²⁷ Provided one can identify a function that produces evaluations in line with the concerns underlying the pattern good

differ in whether or not the output of the pattern good function will subsume the value of the pattern good's inputs or will remain independent from them. As I will describe in Section V, the choice of which of these approaches to use can have implications for whether and how multiple pattern goods can be successfully represented in the same axiology. Section III discusses some of the kinds of goods that could potentially appear as inputs or outputs to the functions used to represent pattern goods. Section IV briefly canvasses what the most appropriate inputs and outputs would be for common pattern goods. In Section V, I will propose a series of criteria that must be met if multiple pattern goods are to be represented in an axiology without incurring problems such as the money pump described in Chapter 1.

II. Ways of Representing a Pattern Good Mathematically

Before proceeding, I must briefly explain some terminology that I will be using. Throughout the rest of this thesis I refer to what I call the 'inputs' and 'outputs' of pattern good functions. Briefly put, the inputs are the goods or quantities that the value of the pattern good supervenes on. Usually pattern goods take some form of wellbeing as their input. For example, standard representations of Prioritarianism and time discounting both make a unique contribution to overall good that depends on the way that wellbeing is distributed. In some cases, using a good as input involves taking that good as an argument in a function. For example, I have defined standard Prioritarianism as being the sum of the functions: $P(\sum_{t \in N(i)} w(i, t))$ applied to each person. This function could be rewritten as $P(w_i)$, where w_i represents person i 's lifetime wellbeing. So the Prioritarian function takes a person's lifetime wellbeing as its input.

In other cases, such as with discounting, the standard representation does not quite fit this model. The standard discounting expression I gave multiplies total wellbeing a time by the function $D(t) = \exp(-rt)$ (refer to expression (3)). In this expression, the input (total wellbeing at a given time) is multiplied by a function rather than being used as the argument in a function. However, this expression could be rewritten so that total wellbeing at a time was used as the argument to a function representing

discounting. For example, $D(t) \sum_{i \in N(t)} w(i, t)$ could be rewritten as $F(w_t) = (\exp(-rt)) * w_t$, where w_t is total wellbeing at time t . In any case, when I discuss the inputs and outputs of pattern good functions, I will be less concerned by the specific mathematical representations of them and more concerned with whether given inputs and outputs are appropriate to represent the ethical position that underlies the use of the pattern good. The input to a pattern good is the kind of good it supervenes on, and the output is the kind of contribution it makes to overall goodness. In what follows I will give a more thorough account of the kinds of outputs that pattern good functions can have.

In this section I will discuss two alternate ways of representing pattern goods. The first approach to representing pattern goods is something I will call *replacement representation*. Replacement representations of pattern goods subsume the quantities they take as input, and output a new quantity that is simultaneously representative of their input and the contribution made by the pattern good; in other words, the output for the pattern good represents the adjusted contribution that the input quantity makes to overall goodness after the pattern good is taken into consideration. The standard Prioritarian function is an example of this: it takes a person's lifetime wellbeing as its input, and it outputs the contribution that person's life makes to overall goodness. This output represents the person's wellbeing after taking the pattern good into consideration.

The second approach involves calculating the contribution of the pattern good as distinct from the contribution of any quantity that was used as input to the pattern good function. I will refer to this way of representing pattern goods as *additive representation*. Rather than replacing the input with a new quantity that represents both the input and the pattern good in question, additive representations generate a distinct quantity that represents nothing but the contribution of each pattern good. From here, overall goodness can be determined by summing the contribution of the pattern good and the quantity that was used as input to the pattern good function (along with any additional quantities that are relevant to overall goodness).

Any replacement representation of a pattern good can technically be converted into an additive representation and vice versa. The following steps can be taken to convert a replacement representation to an additive representation: (1) First, calculate the total quantity of the pattern good's input that is present in the state of affairs under consideration; (2) next, calculate the modified contribution of that input using the replacement style function as normal; (3) then, subtract the total quantity of the input from the modified contribution of the input according to the pattern good. The resulting quantity is the additive representation of the contribution of the pattern good. It can be added to the input quantity to give the same overall goodness scores as would have been achieved by using the replacement representation of that pattern good. This process can be described the following expression:

(7)

$$R(I) - \sum I = A(I)$$

Where I is the pattern good function's input, $R(I)$ is the replacement form of the pattern good function, and $A(I)$ is the additive form of the pattern good function. It also possible to convert an additive representation of any pattern good into a replacement representation; one straightforward way to do this is to sum the total quantity of the pattern good's input that is present in a state of affairs with the additive contribution of the pattern good. Since the resulting quantity accounts for both the input and the contribution of the pattern good it can be said to represent the adjusted contribution of the input to overall goodness after taking the pattern good into account.

III. Types of Goods

In this section I will discuss some of the distinctions between different types of goods that might be used as inputs or outputs for pattern good functions. The standard input for most pattern good functions is wellbeing, either momentary or lifetime. However, there are a range of possible inputs and outputs that pattern good

functions might use. As I will argue in Section V, when multiple pattern goods are represented in an axiology, it is often necessary to have those goods interact in certain ways. For example, by having the *output* of one pattern good function be used as *input* for another pattern good function. Because of this, it is quite important to determine what the appropriate inputs and outputs are for any pattern goods that are to be included in the axiology.

The different kinds of goods that I will discuss are distinguished by three main features. These are: (1) the type of thing that is represented by this good e.g. wellbeing, equality, general good, etc.; (2) the indexical properties of this good, i.e. is this good a property of times and/or lives (in the sense that an instance of wellbeing is enjoyed by a particular person at a particular time); and (3) from what perspective is this good being evaluated? Is it perspective neutral, or is it from the perspective of a particular moment in time? If a good has the property of being dependent on the temporal perspective from which it is assessed, its value may change based on whether one is evaluating it from one time or another.

Note that here I am describing categories of goods, not specific instances of goods that happen to fall within that category. For example, as I am using it, the term ‘time indexed general good’ does not mean general good at one specific time, rather it means the category of goods that are indexed to times – specific instances of this type of good could be indexed to any time.

Time and Person Indexed Wellbeing

This category of goods describes wellbeing that is enjoyed by a particular person at a particular time within her life. Therefore, if a pattern good function takes time and person indexed wellbeing as its input, it will take one or more wellbeing scores corresponding to specific people at specific times as its input.

Person Indexed Wellbeing

Also referred to as lifetime wellbeing, this category refers to wellbeing scores attributable to the whole of a given person’s life as opposed to any specific time in the person’s life. This might be equal to the total time indexed wellbeing belonging

to the person, or (according to some views such as shape of a life) it might be a function (other than summation) of the person's momentary wellbeing scores.

Time Indexed Wellbeing

This category pertains to wellbeing that is a property of a particular time but not of any particular person. This can be used to describe the total wellbeing across all people at a given moment in time. This is the standard input for the time discounting function.

The next four categories of goods describe the specific contribution made by a pattern good. These are the kinds of goods that would result from using the additive style representation, whereas general good (discussed further below) indexed to the relevant dimensions would be the appropriate output for a replacement style approach. It is worth noting that I have described these categories as indexed 'pattern goods', but I am merely using this as a stand in to describe a family of different kinds of goods, each one corresponding to a particular pattern good. For example, 'person indexed pattern good' encompasses many kinds of goods such as person indexed Prioritarian contribution, person indexed shape of a life contribution, and so on.

Time Indexed Pattern Good Contribution

A specific instance of this type of good would be overall goodness contributed by a particular pattern good at a particular time.

Person Indexed Pattern Good Contribution

As above, but indexed to people rather than times.

Person and Time Indexed Pattern Good Contribution

As above, but indexed to both people and times.

The next three categories describe varieties of general goods. General goods are not attributable to any one source such as wellbeing or a specific pattern good. Instead they are an amalgam of multiple such goods. Typically, the outputs of replacement pattern good functions will take this form.

Person Indexed General Good

Unlike person indexed wellbeing, this category can accommodate views that hold that a person's contribution to overall goodness is not equivalent to her overall wellbeing. For example, Prioritarianism holds that a person's contribution to overall goodness is a function of her overall wellbeing. Hence, the standard representation of Prioritarianism takes person indexed wellbeing as its input and outputs person indexed general good.

Time Indexed General Good

This is similar to the previous category, but rather than being indexed to people it is indexed to times. It describes the goodness due to multiple types of goods that is contributed by a given interval of time.

Person and Time Indexed General Good

An instance of this type of good would be the amalgam of nonspecific kinds of goodness contributed by a particular person at a particular time.

All of the above forms of goods can be either goods from the perspective of a particular time, such as the present moment, or perspective neutral goods. If a good is dependent on the temporal perspective of the evaluator, then the 'quantity' or effective value of that good changes as time passes. This category is important because relative time discounting inputs time indexed perspective neutral wellbeing, and outputs time indexed general goodness *from the perspective of the present moment*.

Overall Goodness of a State of Affairs

This category refers to how good a state of affairs is overall. This incorporates all other forms of goods present in the state of affairs.

IV. What Types of Goodness are Associated with the Common Pattern Goods?

In what follows I will briefly note the types of goods that are used as inputs and outputs for the standard ways of representing some common pattern goods. This will be relevant because, as I will argue, whether or not multiple pattern goods can be

represented simultaneously in an axiology often depends on what types of goods those pattern goods take as their outputs and inputs.

Prioritarianism

In its standard formulation, Prioritarianism takes peoples' lifetime wellbeing levels (Person Indexed Wellbeing) as its input and outputs overall good contributed by a person (Person Indexed General Good)²⁸.

Concern for Equality

Generally, functions intended to represent concern for equality will take either person indexed wellbeing or person and time indexed wellbeing as their input. There are other approaches that focus on equality of capabilities or some other type of good rather than equality of wellbeing, but within this context where I have assumed that wellbeing is the basic unit of goodness, person indexed wellbeing and person and time indexed wellbeing are the most appropriate candidates. An egalitarian function that takes person indexed wellbeing as its input will evaluate how much total lifetime wellbeing people have relative to each other. Egalitarian functions that take person and time indexed wellbeing as input are concerned not only with the relative amounts of lifetime wellbeing that people enjoy, but also disparities in people's wellbeing levels from moment to moment.

There is some disagreement regarding what the appropriate output for equality representations is. Broome (2004) argues that, insofar as equality is good, it is good because it improves people's lives, therefore he argues that equality really constitutes an aspect of wellbeing and as such the output of an equality function should be either person indexed wellbeing or person and time indexed wellbeing depending on whether or not the function is sensitive to how inequality affects people's wellbeing at specific times. Other theorists such as Temkin (2011) hold that equality has value over and above its impact on people's wellbeing. Representations that express this view would have to output general good or contribution from equality depending on whether they used replacement or additive representations.

²⁸ Broome (2004, pp. 65-66)

These outputs might also be indexed to times and or persons depending on the desired effect.

Shape of a Life

The input for functions representing the shape of a life pattern good must be time and person indexed wellbeing, since the contribution from shape of a life is unique to each person and it depends on the way wellbeing is distributed across different times in a person's life.

The output for this pattern good seems relatively flexible with regard to whether it is in the form of wellbeing, pattern good, or general good. What is clear however is that the output should be indexed to persons. *Prima facie* it also seems that the output should not be indexed to times, since it does not represent the goodness that resides at any particular time or set of times, but rather the goodness that arises out of the relationship between the occurrences across a number of times.

Discounting

Traditionally the input for discounting functions is Time Indexed Wellbeing, so they take as their input a vector, each object in which represents the total wellbeing across all people at a given interval of time. Discounting is usually represented by a replacement function, so its output subsumes the goods it takes as input. Hence the output of the standard relative discounting function is time indexed general good from the perspective of the present moment. As time passes and the perspective of evaluation shifts, so too will the evaluations of the contributions made by events at each time. Discounting can also be performed so that it applies to a single person's momentary wellbeing rather than the total wellbeing across all people at a time. Rather than taking time indexed wellbeing as its input and outputting time indexed general good, it can take time and person indexed wellbeing as its input and output time and person indexed general good. Post-Discounting Prioritarianism (discussed in Chapter 1) uses the later kind of discounting function. It first calculates the discounted wellbeing enjoyed by a person (person and time indexed general good) at each time in her life, and then sums these and applies the Prioritarian function to the total. If this second type of discounting is applied to each person who exists at a

given time and the results are summed, it will give the same value as is achieved by using the standard discounting function.

V. Constraints on How an Axiology Can Represent Two or More Pattern Goods

In this section I propose and defend a set of four constraints on the ways that two or more pattern goods can be simultaneously represented if they are to avoid problematic results like the money pump described in Chapter 1.

Constraint One

In order for one pattern good to be properly conceived of as influencing the contribution of another, the *output* of the function intended to represent the first pattern good must be the same kind of good as is deemed an appropriate *input* for the function that represents the second pattern good.

If two or more pattern goods are expressed by an axiology, those pattern goods can either interact with each other or be independent, depending on how they are mathematically represented. In order for one pattern good to influence another, the output from the function that represents the first pattern good must be used as input for the function that represents the second pattern good. The result of this will be that whatever contribution the first pattern good makes to overall goodness will have some bearing on the contribution that the second pattern good makes. If one is to use the output of one pattern good as the input for another pattern good in this way, it is important that the kind of good represented by the output of the first pattern good is the same kind of good that the second pattern good specifies as an appropriate input. For example, suppose the output of a relative time discounting function is used as input for a shape of a life function; traditionally, shape of a life would use time and person indexed wellbeing as its input, however discounting traditionally outputs time indexed general good (from the perspective of the present moment). If the shape of a life function uses discounted wellbeing scores (in the form of person and time indexed general good) as its input, it will treat lives with a constant level of wellbeing throughout as though their wellbeing sloped downwards over time. Shape of a life, as it is traditionally construed, is a representation of the value of having

certain kinds of distributions of *wellbeing* over time, not as a representation of the value in having certain distributions of *general good as viewed from the perspective of the present moment*. To use the incorrect input for this pattern good function would change its meaning so that the formal representation no longer accurately reflects concern for the shape of a life.

It is worth noting here that Post-Discounting Prioritarianism violates this constraint since it uses the sum of a person's discounted momentary wellbeing scores (person indexed general good from the perspective of the present moment) as input for the Prioritarian function, which would normally take lifetime wellbeing (person indexed wellbeing) as its input. So not only does Post-Discounting Prioritarianism produce a money pump, it will also fail to adequately reflect Prioritarian concerns due to its use of an incorrect input.

Constraint Two

An axiology cannot include two or more *replacement representations* of pattern goods that take the same kind of good as their input.

The reason for this is that if multiple replacement representations use the same input, they will each output a new quantity that represents both the contribution of their respective pattern goods and the contribution of the quantity that was used as input. Therefore, if multiple replacement representations take a single quantity as input, they will output multiple quantities that represent the contribution of that initial input. Thus, an axiology with two such representations would effectively double count the contribution of the input quantity to overall goodness. For example, suppose an axiology expresses two pattern goods as replacement functions that take wellbeing as their input. This means that each pattern good function would generate a new quantity representing the modified contribution of wellbeing after applying that pattern good. In order to have overall goodness express the contributions of both pattern goods, each of the new quantities would need to be summed together. The problem with this approach is that it exaggerates the contribution of wellbeing to overall goodness. Since the new quantities generated by the two functions represent both the pattern good and value of wellbeing, an

axiology using this approach would essentially count the value of wellbeing twice, once for each pattern good used.

This problem can generally be avoided by recasting any excess replacement functions as additive functions. Because additive pattern good functions output a quantity that is only representative of the pattern good's unique contribution, they cannot generate this double counting problem.

Constraint Three

If time discounting is to be represented in an axiology, the output of the discounting function must either influence or be influenced by every other pattern good expressed in the axiology.

The reason for this is that if discounting does not interact with the other pattern goods in this way, the contribution of the other pattern goods will overwhelm the contribution of discounted wellbeing at times in the further future but not times in the nearer future; this inconsistency in evaluations over time generates a money pump.

Consider an axiology that expresses both time discounting and concern for equality. Suppose that it uses the standard replacement representation for time discounting and an additive representation for concern for equality. Assume that neither of these pattern goods takes the output of the other pattern good as input²⁹. Whether this axiology calculates the contribution of wellbeing or discounting first makes no difference, since neither contribution will bear on the contribution of the other. Once

²⁹ This argument does not depend on the axiology using these particular representations; it could also use additive representations for both pattern goods, or use additive representation for discounting and replacement representation for equality. Since the pattern good functions in this example do not interact, it will make no difference to the evaluations given by the axiology which of these representation styles is chosen. Note that the axiology cannot represent both pattern goods using the replacement approach since doing so would violate Constraint Two (because they both take some form of wellbeing as their input).

the contributions of these two pattern goods have been computed, overall goodness can be expressed as the sum of discounted wellbeing and the contribution from equality.

Suppose this axiology is used to evaluate two states of affairs that I will call state 1 and state 2. Both have the exact same distributions of wellbeing for the next 99 years into the future, but at the 100th year, state 1 has substantially less wellbeing than state 2, but wellbeing is distributed slightly more equally in state 1 than it is in state 2. These wellbeing distributions are pictured below. The height of the bars indicates the quantity of wellbeing enjoyed by a person at time t . The width of the bars indicates how many people exist at that level of wellbeing (both states contain the same number of people).

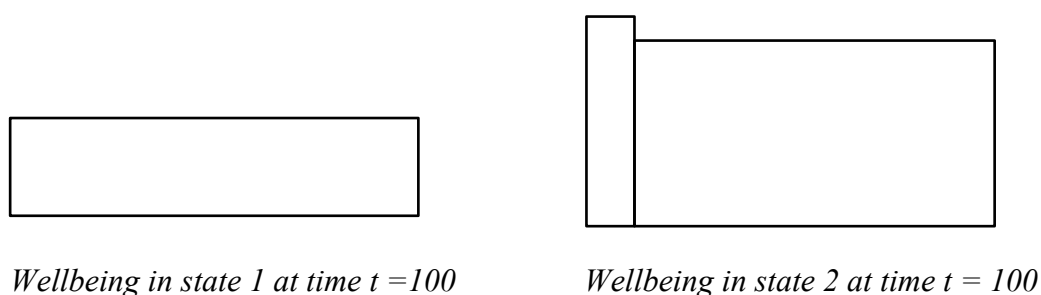


Figure 10. Distributions of wellbeing at $t = 100$ in states 1 and 2

At $t = 100$ in state 2, every person is significantly better off than their counterparts in state 1, but state 2 has a small amount of inequality which is not present in state 1.

Because discounting diminishes the contribution of wellbeing at time t by an amount proportional to the quantity of wellbeing at time t , the state that has a greater quantity of wellbeing will receive a greater reduction (in absolute terms) in the contribution of its wellbeing. This means that discounting reduces the absolute difference between the contributions made by the wellbeing in each state. For example, suppose that at $t = 100$, there are ten people in state 1 and each of them enjoys a momentary wellbeing level of 50, likewise there are ten people in state 2, but one of them enjoys a wellbeing level of 100, and the other nine enjoy a wellbeing level of 90. This means that at $t = 100$, the total wellbeing level in state 1

is 500 (50×10) and the total wellbeing level in state 2 is 910 ($100 + 90 \times 9$), so the difference between the total wellbeing levels in these states is 410. If the standard discounting function is applied using a discount rate of 0.05³⁰, the discounted contribution from wellbeing at $t = 100$ in state 1 will be $500 \times e^{-0.05 \times 100} = 3.35$, whereas the discounted contribution from wellbeing in state 2 will be $910 \times e^{-0.05 \times 100} = 6.097$. Therefore, after discounting, the difference (in absolute terms) between the contribution of wellbeing in states 1 and 2 is $6.097 - 3.35 = 2.747$. If the axiology evaluates the same moment after 10 years have passed, so that the moment occurs at $t = 90$, the discounted contribution of wellbeing in state 1 will be $500 \times e^{-0.05 \times 90} = 5.554$, whereas the discounted contribution from wellbeing in state 2 will be $910 \times e^{-0.05 \times 90} = 10.109$. This yields a difference of $10.109 - 5.554 = 4.555$ in discounted contribution of wellbeing between the two states. The higher the value of t (i.e. the further into the future the moment in question is) the more substantially discounting will diminish the absolute difference between the contributions of the wellbeing in the two states. This means that when the moment in question is 100 years into the future, the axiology will register almost no difference between the contribution from wellbeing in the two states, and as time passes and the moment becomes closer to the present, the contribution from wellbeing will differ more and more between the two states until eventually the moment in question occupies the present ($t = 0$), so the wellbeing levels will be multiplied by discount factor of 1 (as $e^{-0.05 \times 0} = 1$), and the difference between the contributions of wellbeing in the two states will be 410.

Since time discounting has no bearing on the representation of concern for equality that is used in this axiology, the inequality in state 2 will register as being just as significant at $t = 100$ as if it were to occur at $t = 90$ or $t = 0$. This means that, whatever the particular values that the equality function outputs for states 1 and 2, the difference between the contributions from equality in the two states will remain fixed, regardless of from what time the states are evaluated. So, as time passes and

³⁰ The discount rate is the value for r in the function $R(t) = e^{-r t}$, as described in Section IV of Chapter 1. Economists who advocate discounting typically select values for r that lie close to 0.05. Greaves (*n.d.*).

the moment depicted in the diagram above moves closer to the present, the difference between the contributions to goodness made by the total wellbeing in the two states will increase, whereas the difference between the equality contributions in the two states will remain the same. This property of the axiology makes it vulnerable to a money pump.

To illustrate this money pump, I require some hypothetical outputs given by a concern for equality function. Specifically, I require a value for the difference between the equality contributions in state one and state 2. A discussion of the specific functions that can be used to represent concern for equality would take me too far afield; however it will suffice to use the variable x to stand in for the difference between the levels of equality in state 1 and state 2. It is not necessary to have a specific value of x to illustrate that a pump will occur.

In the scenario described above, there is a single time in the future at which states 1 and two differ; the wellbeing distributions in the two states at that moment are shown in figure 10. To speed up the exposition, I will hence forth refer to this moment as the ‘difference event’. I will use the variable y to describe the difference between the discounted wellbeing levels in state 1 and state 2 during the difference event. Depending on the moment in time from which the difference event is evaluated, the value for y will differ. At times when x is greater than y , state 1 will be preferred over state 2, since the extra equality value it provides is greater than the amount of extra discounted wellbeing that state 2 provides. I must stipulate one further condition for this money pump to occur, which is that, when the difference event occupies the present moment, the value for x is less than the value for y . In other words, when the difference event is not at all impacted by discounting, the difference between the total wellbeing levels in the two states must be greater than the difference in the equality values between the two states. Given these conditions, a money pump will occur.

A hypothetical agent is evaluating states one and two from some moment in time prior to the difference event. Suppose that if no action is taken at the present moment, state 2 will occur. It is guaranteed that, at some point in time such that the difference event is in the future, the value for x will be greater than the value for y .

Precisely how far the difference event must be into the future for this to occur depends on the specific equality function and discounting function used in the axiology, but given that the value for x is insensitive to time and the value for y diminishes the further the difference event is into the future, there will be some point in time at which $x > y$. At this point in time, if the difference between x and y is sufficiently large, the axiology will judge it worthwhile to make some small sacrifice (perhaps slightly reducing the wellbeing at the present time) in order to bring about state 1 rather than state 2. Suppose that an agent using the axiology follows this advice and makes a sacrifice to set events in motion that will bring about state 1.

After a certain amount of time has passed, the axiology will reverse its evaluation and judge state 2 to be the superior option. When this occurs will depend on the specific value of x , but as I stipulated, x is less than the value that y takes when the difference event occupies the present time, so there will be some point at which the value of y overtakes the value of x . When the value for x is less than y , the extra equality value offered by state 1 will be less than the extra discounted contribution of wellbeing offered by state 2. Suppose that shortly after this shift occurs, the agent has the option to make a further sacrifice to reverse her earlier action so that state 2 rather than state 1 will be brought about. Since she now regards state 2 to be the superior option, the agent elects to make this sacrifice to ensure that state 2 will occur.

By taking this series of actions, the agent has arrived at the status quo that existed before she took either action, however she has also sacrificed a small amount of wellbeing by taking these actions. By adhering to the evaluations given by the axiology, the agent has engineered a state of affairs that is in one way worse and in no way better than the status quo that would have resulted if she had not acted. This final circumstance is worse than the status quo from whatever perspective in time it is evaluated. Hence, this is a money pump.

If an axiology includes functions representing time discounting and a second pattern good that do not influence one another, then that axiology can potentially be money pumped, provided that the following conditions are met: (1) the states under

evaluation contain at least one moment at which total wellbeing is greater in one state and the contribution from the second pattern good is greater in the other state; and (2) the difference between the values of second pattern good's contribution in the two states at that that moment is less than the difference between the undiscounted wellbeing totals at that moment.

Note that I only claim an axiology that meets these conditions can *potentially* be money pumped; whether or not it is actually money pumped will depend on whether the other features of the states are at close enough evaluative parity for the shift in relative values of the second pattern good and discounted wellbeing at a given moment to be decisive. In other words, the pump will not occur if other features of the states (such as the distribution of wellbeing at other times) ensure that one state will consistently be preferred over the other. These conditions are numerous, but given a long enough period of time, a situation fitting the requirements for this money pump is highly likely to manifest. For this reason, I argue that it is necessary to impose Constraint Three on axiologies that express time discounting (either relative or non relative) along with one or more other pattern goods.

Constraint Four

If *relative* time discounting is included in an axiology, in addition to interacting with every other pattern good (as per Constraint Three), it must be applied after any pattern goods that make times interdependent; therefore the outputs of any pattern goods that make times interdependent must collectively be used as input for the discounting function.

As I demonstrated in the first chapter, if an axiology applies discounting prior to a pattern good that makes times interdependent for the purposes of evaluation, then that axiology can be money pumped; hence the need for this constraint. This constraint is relevant to all of the non-discounting pattern goods I have discussed, since they each take goods ranging across multiple times as their input. In other words the contribution they make depends on the relationship between quantities of goods at different intervals of time. In other words, they make times interdependent. This constraint is potentially difficult to adhere to because using the output of a

discounting function as input for another pattern good will typically be disallowed by Constraint 1. By default, time discounting takes time indexed wellbeing or time and person indexed wellbeing as its input. Of the pattern goods I have discussed, only concern for equality produces a form of time indexed wellbeing as its output (and, as I mentioned in Section IV, whether or not it does this is debated). Unless a way can be found to make the output of a pattern good function that makes times interdependent (such as Prioritarianism) compatible with the input of relative time discounting, those two pattern goods should not be represented in the same axiology.

VI. Conclusion

From these constraints it can be seen that if relative time discounting and Prioritarianism are to be successfully combined in an axiology, one or the other of these pattern goods must influence the other's contribution (as per Constraint Three). Furthermore, it must be Prioritarianism that influences discounting's contribution and not the other way around, since Prioritarianism makes times interdependent (Constraint Four), and in order to legitimately influence discounting's contribution, the output of the Prioritarian function must represent the same kind of good as the discounting function requires for its input (Constraint One). Given the default outputs and inputs I stipulated, it seems that the output of the Prioritarian function (person indexed general good) does not match the input of the discounting function (time indexed wellbeing). The upshot of this is that if Prioritarianism and relative time discounting are to be represented in the same axiology, one or both of them must be expressed in a way that deviates from the standard formula. I argued in Chapter 1 that some previously proposed ways of doing this (i.e. Time-Slice Prioritarianism and Uniformly-Discounted Prioritarianism) were unsatisfactory. In Chapter 3, I will investigate whether it is possible to find (and to justify) representations of these pattern goods that are compatible with each other and which do not distort the original intent of the views they represent.

CHAPTER 3

I. Introduction

As I discussed in Chapter 2, the constraints on how multiple pattern goods can be simultaneously represented in axiology rule out the combination of the standard Prioritarianism and relative time discounting functions. These constraints would also rule out combining relative time discounting with concern for the shape of a life and some versions of concern for equality (such as the approach endorsed by Temkin (1993), which entails that the equality function cannot output wellbeing), since the standard functions for these pattern goods make times interdependent and they do not produce outputs compatible with discounting. In this chapter I will discuss the prospects for using nonstandard representations of pattern goods to enable combinations that would otherwise be prohibited by these constraints. Throughout, I will use the task of matching relative time discounting and Prioritarianism as a case study. I will discuss a number of formal maneuvers that can be used to generate alternative representations of pattern goods so that their outputs could be used as input for a discounting function; in particular, I will discuss approaches that allow the output of a pattern good function to be indexed to times, even when this is not its default representation. I will also discuss how different arguments in favour of discounting suggest different types of goods as being the appropriate input for the discounting function. Depending on which of these arguments (if any) are deemed compelling, the appropriate input for discounting may differ, and consequently it may be more or less difficult to justifiably combine relative time discounting with other pattern goods.

II. Pattern Goods Rule out Separability

One of the consequences of including pattern goods in an axiology is that, depending on the pattern good, the axiology must forgo certain assumptions of separability. Separability is a formal property that pertains to the ways aggregation functions are able to operate on the dimensions of times and lives. Roughly

speaking, separability of a dimension means that each location on that dimension can be evaluated independently of the other locations. For example, if times are separable, then the amount that goods indexed to a given time contribute towards overall goodness does not depend on what goods occur at any other times.

Before I explain separability in more detail it will help to briefly discuss ways that the process of aggregation might be broken into chunks. Broome (2004) discusses two broad approaches to aggregation: the people route and the snapshot route. The snapshot route begins by determining the overall value at each time. It does this aggregating the wellbeing levels that each person experiences during that interval of time. It then proceeds to combine the values for each time to generate a single value for the whole population across all time. By contrast, the people route first determines the overall value of each person's life. It does this by aggregating the wellbeing levels at each time within a person's life to generate a quantity representing how the total wellbeing across that person's life contributes to overall goodness. The values for each person's life are then aggregated together to determine the overall goodness of the state of affairs.

Each of the two routes to aggregation requires that value be separable with respect to at least one dimension. The people route requires the assumption that value is separable with respect to people's lives, whereas the snapshot route requires the assumption that value is separable with respect to intervals of time. In what follows I will provide a more technical definition of these two kinds of separability and explain how they relate to the snapshot and people routes to aggregation.

An aggregation function that evaluates a state of affairs based on the distribution of wellbeing across time can be represented with this general form:

$$(8) \quad f(w_1^1, w_2^1, \dots, w_t^1, w_1^2, w_2^2, \dots, w_t^2, w_1^3, w_2^3, \dots, w_t^3, \dots)$$

The variable w indicates a level of wellbeing enjoyed by a person at a time. The subscript numbers indicate times and the superscript numbers indicate persons. This function takes as its arguments every instance of wellbeing enjoyed by a person at a

time. Unlike the other expressions I have discussed in this thesis, it does not break the aggregation process down into steps first applying a series of functions to distinct regions in the distribution and then applying a function (such as summation) to the outputs of these functions. Instead, it operates on every part of the distribution simultaneously. If and only if value is separable with respect to times, the above function can be expressed in the form:

$$(9) \quad f' (f_1(w_1^1, w_1^2, w_1^3 \dots), f_2(w_2^1, w_2^2, w_2^3 \dots), f_t(w_t^1, w_t^2, w_t^3 \dots))$$

The values of $f_1(\cdot)$, $f_2(\cdot)$, and so on are real numbers, and the function $f'(\cdot)$ is strictly increasing in each of its arguments. (A function is said to be *increasing* in one of its arguments if and only if the value of the function increases whenever an argument increases while the other arguments remain constant). The expression (9) shown above applies a function to each interval of time separately before applying a final function that takes the outputs of the other functions as its arguments. $f_1(\cdot)$, $f_2(\cdot)$, and so on can be interpreted as functions representing snapshot valuations at each time. Separability of times assumes that we can value the distribution of wellbeing at a time independently of people's wellbeing's at other times. The snapshot valuations must be enough, taken together, to determine the overall value of the distribution. If times are not separable, function (8) cannot be expressed in this form and the snapshot route to aggregation cannot be taken.

Correspondingly, to assume that lives are separable is to assume that the value function (8) shown above can be expressed in this form:

$$(10) \quad f^*(f^1(w_1^1, w_2^1, \dots, w_t^1), f^2(w_1^2, w_2^2, \dots, w_t^2), f^3(w_1^3, w_2^3, \dots, w_t^3), \dots)$$

Once again, the values of $f^1(\cdot)$, $f^2(\cdot)$, and so on are real numbers, and the function $f^*(\cdot)$ is increasing in each of its arguments. $f^1(\cdot)$, $f^2(\cdot)$, and so on can be interpreted as value functions representing valuations of a person's entire life. Separability of people assumes that we can value the wellbeing in a person's life independently of

the distribution of wellbeing in other people's lives ³¹.

By definition, pattern goods rule out one or more forms of separability because they hold that the contribution of goods at a given index *does* depend on the quantities of goods at other indexes. Furthermore, if a pattern good rules out separability across a dimension (like times or lives), then by default, the output of that pattern good will not be indexed to any loci in that dimension. This is because the pattern good's contribution is derived from the joint influence of a series of loci as opposed to a discrete locus. This poses a problem for combining pattern goods with discounting, since discounting must take inputs that are indexed to times and any pattern good that rules out separability of times will (by default) not have an output that is indexed to times. In the following sections I will discuss ways that these pattern goods can be represented so that their output *is* indexed to times. This is easy to achieve formally, but because it diverges from what might be called the natural way of representing these pattern goods, it requires some justification.

III. Dispersion

In the following sections, I will discuss ways of indexing the output of pattern good functions to times when the output does not naturally take this form. This is relatively simple to achieve in formal terms. For example, the output of any pattern good could be arbitrarily assigned to a particular moment in time. However, it is not necessarily the case that such a move can be justified as a claim about the underlying values which the pattern good function is intended to represent. Indexing the output to times involves converting the output of the pattern good into a different form of good. For example, if the output of the pattern good takes the form of Person Indexed General Good by default, once that output has been indexed to one or more times, it would then take the form of Person and Time Indexed General Good. Such a move must be justified. It must be the case that the ethical concerns that the pattern good represents can be pinned down to a particular time or times. Furthermore, a defense will need to be mounted regarding the choice of which times

³¹ This formal definition of separability is sourced from Broome (2004).

to assign the output to. As I will discuss in section VI, it is not always obvious what the appropriate time to assign a pattern good's output to would be.

The first approach I will discuss is called *dispersion*; this is a technical maneuver discussed by Broome (2004) as a means of maintaining assumptions of separability in conjunction with pattern goods. If this strategy can be justified, it will mean that the axiology can represent the pattern good in question so that its output is indexed to times. This would be one step closer to making the output of the pattern good an appropriate input for a time discounting function.

Roughly speaking, dispersion involves taking a value like Prioritarianism that would normally be represented by a function that outputs general good or pattern good contribution which is not indexed to times, and finding a way to represent the output of this pattern good as Time and Person Indexed Wellbeing. Dispersion takes the output of a pattern good and distributes it across loci in the distribution of wellbeing such that summing the quantities at those loci will yield the same result as the output of the pattern good³². Once a pattern good has been dispersed in this way, the pattern good's contribution is accounted for without any further need for functions that make the contribution of one locus dependent on another. Thus, any time interdependence incurred by the pattern good function has already been accounted for. Therefore, it cannot introduce problems by making times interdependent post-discounting. Dispersion also makes it possible to index the output of the Prioritarian function to times so that it could potentially be used as input for a discounting function.

For each person, the output of the Prioritarian function could be dispersed across loci in the distribution corresponding to that person's life. This would involve carrying out a series of operations to replace the distribution of wellbeing with a

³² This is assuming that the pattern good has a replacement representation. If the pattern good has an additive representation, then summing the post dispersion wellbeing loci will yield a result equivalent to the sum of the pattern good's contribution and the total wellbeing present. Either way, summing the wellbeing loci post dispersion will account for both the contribution of wellbeing and the contribution of the pattern good.

new distribution that is reflective of the way that Prioritarianism alters the contribution that wellbeing makes to overall goodness. This new distribution would be such that using a simple summation function on it would yield the same result as would have been achieved by applying the Prioritarian function to the original distribution of wellbeing. The summation function is separable with respect to both times and people. It would therefore be possible to apply an aggregation function that consisted of a summation function modified to incorporate discounting (such as expression (3) discussed in Section IV of Chapter 1) to the modified distribution.

(3)

$$G = \sum_t R(t) \sum_{i \in N(t)} w(i, t)$$

Doing so would yield a result that was expressive of both Prioritarianism and discounting, and since the expression (3) is separable with respect to times, it does not feature any time interdependence, and so it is not at risk of generating a money pump. Note that expression (3) is formally identical to the standard discounting function; the only difference is that the momentary wellbeing levels it is operating on have been modified to reflect the contribution of Prioritarianism.

In order to index the contributions of a pattern good to times, dispersion treats the output of the pattern good as time indexed wellbeing, which allows times to be separable, even after taking into account a pattern good that would normally make times interdependent. By treating the pattern good as wellbeing, it also entails indexing the pattern good to lives, which means that it can also allow for aggregation functions that are separable with respect to lives. This makes dispersion quite powerful. However, as Broome (2004, p. 112) argues, its use also requires substantial justification. As Broome notes, dispersion can technically be applied to any pattern good. He writes that: “a theorist could always disperse [the contribution made by a pattern good] in a formal way by arbitrarily dividing it into parts and allocating each part to a person. She could then say that each person’s allocated part was a component of that person’s wellbeing.” However Broome argues that using dispersion in this way would be a mistake. He insists that to use dispersion is actually to make the substantive claim that the value in question really is a

component of people's wellbeing. As such he thinks it is only appropriate to use dispersion if one can mount a successful argument that the output of the pattern good in question truly does comprise an aspect of wellbeing.

Broome demonstrates this view by showing how he approaches the value of equality. Equality is often construed as a value over and above the quantity of wellbeing that is concerned with people's relative levels of wellbeing. According to this approach, the amount of 'equality value' a society has might be expressed as a function of the comparative levels of wellbeing of each of the members. However Broome claims that rather than expressing this value as a function of the group's levels of wellbeing, we ought to disperse it so that it comprises a component of each individual's wellbeing. Broome argues that this move is justified because, insofar as equality is valuable, it is valuable because it is good *for* people. He claims that: "The badness of inequality is not a separate negative value beyond people's own wellbeing. It is part of people's wellbeing itself" (2004, p. 111)³³. If this is so, then the goodness of equality ought to show up in people's wellbeing levels. Broome defends the dispersion of equality by arguing that although it might seem appropriate to account for equality by using a pattern good function, it is more properly construed as a component of people's wellbeing.

One of Broome's key motivations for this claim seems to be that unless an instance of dispersion fits his criteria, it will be redundant. He argues that:

Unless an instance of dispersion is genuine [...], it will be no help with aggregation. If we know the aggregate value of a distribution, we can always disperse the value arbitrarily to people. But to do that, we would have to know the aggregate value first. Since we are trying to find aggregate value, this arbitrary manoeuvre would be pointless (2004, p. 112).

Since an artificially dispersed pattern good will make the same contribution to overall goodness as the undispersed version, Broome claims that dispersing the output of pattern goods that do not represent wellbeing provides no benefit. As has

³³ See also Broome (1991, Ch. 9) for an extended version of this argument.

been established, dispersion does have benefit in cases where its use can prevent a money pump, so by itself, this argument of Broome's does not make a compelling case for restricting dispersion.

A stronger argument is made implicitly by Broome when he stresses the importance of representing distinct kinds of goods as they really are. When reflecting on his earlier work, *Weighing Goods*, he argues that general good and wellbeing are distinct categories and must be treated as such.

The mistake I made is to forget that wellbeing is *personal* value or goodness, which in section 4.2 of this book I contrasted with *general* goodness. When a person lives her life, her wellbeing is the goodness of the life for her. We have to separate this from the general goodness of the life, or how good it is that she lives this life. It may be generally better to save a present person rather than a future person, even though being saved is no better for the present person than it is for the future person (2004, p. 93).

Dispersing a pattern good that cannot properly be considered an aspect of wellbeing is inadvisable for two reasons. Firstly, it misrepresents the dispersed pattern good by treating its output as a kind of good that is not in keeping with the view that the pattern good was intended to represent. However, by itself this misrepresentation is not especially troubling since it makes no difference to the evaluations produced by the axiology. A more significant problem arises if the dispersed pattern good is combined with a second pattern good that takes wellbeing as its input. To do so would be a violation of Constraint 1 (See Chapter 2 Section V). Using something other than wellbeing as input to a pattern good function that is intended to take wellbeing as its input could significantly distort the contribution that the pattern good makes to overall goodness. This is true even if the non-wellbeing input has been arbitrarily treated as wellbeing.

IV. What Counts as Wellbeing

If we adhere to Broome's criteria then, in order for a pattern good to be eligible for dispersion, its output must be such that it could plausibly be considered an aspect of

wellbeing. There does not seem to me to be any way to reinterpret Prioritarian concerns as an aspect of wellbeing rather than general good without substantially changing their meaning. The mathematical processes used to represent Prioritarianism are antithetical to such a reinterpretation. Prioritarianism is premised on there being a disconnect between wellbeing for an individual and overall goodness. The standard formulation holds that increases in a person's wellbeing provide a diminishing marginal contribution to overall goodness. To disperse Prioritarianism seems to imply paradoxically that an individual's wellbeing provides a diminishing marginal contribution to her own wellbeing. Given this, the classic Prioritarian formulation of diminishing marginal value cannot be considered a component of wellbeing. As such, Prioritarianism cannot meet Broome's criteria for dispersion, so if we accept Broome's criteria as reasonable, dispersion cannot be used to avoid the money pump described in Chapter 1. However, in the next section I will propose an alternative route to indexing Prioritarianism's contribution to times, one which does not require that Prioritarianism represent an aspect of wellbeing.

With the exception of the example of Prioritarianism, I will not discuss whether other pattern goods can or can't be represented as wellbeing. A full discussion of the general criteria for determining whether a good counts as wellbeing is beyond the scope of this thesis. However, for overviews of canonical positions on this issue see Parfit (1987) Appendix I and Shelly Kagan (1992). Bernard Gert (1990) has also proposed an interesting set of criteria for whether something counts as wellbeing that is based around a simple list of conditions. The fact that Prioritarianism cannot be represented as wellbeing is enough to show that, by itself, dispersion cannot solve all compatibility problems between pattern goods. However, I claim that there are alternative pathways to indexing the output of pattern goods to times that do not come with the requirement that the good in question be wellbeing (or any particular sort of good). This means it is not crucial to identify whether a pattern good can be justifiably represented as wellbeing, since my method allows any pattern good to be indexed to times whether or not it constitutes wellbeing.

V. Indexing

In this section I will argue that the outputs of pattern good functions can be indexed to times, or lives, or both, without needing to treat those outputs as aspects of wellbeing. Instead, this is achieved by taking whatever type of good the pattern good outputs by default, and converting it to the most similar form of good that is also indexed to the desired dimension (times or lives). So, for example, if a pattern good function outputs person indexed general good by default, and one wanted to index this output to times, this approach would convert this output to person and time indexed general good. The advantage of this approach is that it does not require goods to take the form of wellbeing in order to index them to times and or lives. However, this approach still requires justification. It must be possible to sensibly assign the contribution of the pattern good to the desired index. For example, if the pattern good is to be indexed to times, then it must be in keeping with the concerns expressed by the pattern good that the goodness it contributes could be conceived of as belonging to one or more specific times. For convenience, I will henceforth call this approach *indexing*.

As with dispersion, indexing involves calculating the contribution of the pattern good normally (this can be done with either replacement or additive style representations) and then dividing the resulting quantity into parts and allocating those parts to one or more specific indexes. Once this step is complete, the overall contribution of the pattern good can be determined simply by taking the sum of these parts. If they are of the right sort to be used as input, they can also be input into another pattern good function so that the two pattern goods interact.

Furthermore, in the same way that the dispersed parts of a pattern good add to the existing quantities of wellbeing at each locus, the indexed quantities of a pattern good are summed with any preexisting quantities that occupied the same loci in the distribution of whatever type of goodness that pattern good represents. So, if two separate pattern goods represent time and person indexed general good after indexing, their contributions at each locus will be summed together to give overall time and person indexed general good. It is important that the contributions add in this way so that, if these outputs are used as input for

another function, there is only one quantity occupying each locus. If instead, indexing allowed pattern goods to maintain distinct contributions at a given locus, extra rules would need to be introduced to determine how the two quantities at this locus would be used as input by any other pattern good. This approach would not necessarily cause problems, but for the sake of simplicity I will stipulate that contributions from different pattern goods sum when they occupy the same locus. The quantity of the good at that locus then represents the total amount of goodness contributed by that locus after taking all goods into account.

To justify the use of indexing for a particular pattern good requires two steps. Firstly, it must be argued that the pattern good represents the sort of ethical concern that can be meaningfully attributed to the dimension or dimensions within which it is to be indexed. So, if one wishes to index a pattern good to lives, it must be intelligible and faithful to the character of the ethical concern on which that pattern good is founded to think of the pattern good as being a property of lives. The same holds true for times if one wishes to index across the dimension of times. For example, if one wished to index the output of a survival of humanity pattern good to times, one would have to argue that it made sense to think of this value as being comprised of parts distributed across one or more times. The second step is to justify the way that the parts of the pattern good will be allocated to loci. This involves answering questions such as which specific loci can the pattern good be attributed to, and what is the portion of the pattern goods total contribution that belongs to each locus. If, after indexing, the indexed pattern good is used as input to another pattern good, then the choice of how to distribute the first pattern good can have consequential effects on the overall evaluations made by the axiology, so it is important that this distribution be well founded.

I will use the example of Prioritarianism to illustrate how these steps might be justified. I assume that it is intelligible to think of the output of the Prioritarian function being attributed to times. Standard Prioritarianism adjusts the amount that a person's lifetime wellbeing contributes to overall goodness. It seems reasonable to say then that the adjusted contribution can be attributed to a time or

times within that person's life. At least with this pattern good, there is a reasonably clear range of times to which the good could be attributed. The second task is to decide precisely which times within a person's life will have the contribution attributed to them, and what portion of the total contribution they will each be allocated.

I claim that the most appropriate option is to distribute the contribution across all times in the person's life in such a way that the greater a person's wellbeing is at a given time, the larger the portion of the person's Prioritarian contribution that would be allocated to that time after indexing. To mathematically represent this approach the following steps can be taken: (1) for each time in a person's life, determine the proportion of his or her total lifetime wellbeing that was contributed by his or her wellbeing at that time; (2) use the standard Prioritarian function to calculate the contribution to overall goodness that the total wellbeing across that person's life will make; (3) for each time in the person's life, index a value to that time that reflects what the wellbeing at that time contributed to overall goodness. This means that, whatever percentage of the person's total wellbeing that time accounted for, the indexed quantity should be an equivalent percentage of the person's contribution to overall goodness (as determined by the Prioritarian function (2)). After this is done for each time in a person's life, summing the indexed values for each time will yield a result equivalent to applying the Prioritarian function to the person's lifetime wellbeing. The approach I have just described can be performed by expression (11), shown below. The function $g(i, t)$ in expression (11) represents the contribution that time t in person i 's life makes to overall goodness after accounting for Prioritarianism.

(2)

$$G = \sum_i P \left(\sum_{t \in N(i)} w(i, t) \right)$$

(11)

$$g(i, t) = \frac{w(i, t)}{\sum_{t \in N(i)} w(i, t)} \times P \left(\sum_{t \in N(i)} w(i, t) \right)$$

Indexing in the way I have described would produce something like the example distribution in Table 2 if the distribution in Table 1 were used input.

Table 1. Person and Time Indexed Wellbeing

	T1	T2	T3
Person 1's Wellbeing	1	2	1
Person 2's Wellbeing	2	2	2
Person 3's Wellbeing	1	1	3

Table 2. Hypothetical Person and Time Indexed General Good

	T1	T2	T3
Person 1's Prioritarian Contribution	0.87	1.74	0.87
Person 2's Prioritarian Contribution	1.41	1.41	1.41
Person 3's Prioritarian Contribution	0.77	0.77	2.31

The values at the loci in Table 2 each represent the amount that a person's wellbeing at a given time contributed to overall goodness after adjusting for Prioritarianism. These quantities are proportional to the quantities in the first table, so they maintain the information about how the person's wellbeing levels (and therefore Prioritarian adjusted contribution to overall goodness) varied over the course of her life.

I pause here to note that Uniformly-Discounted Prioritarianism, which I discussed in Chapter 1, bears significant similarity to the approach I am advocating. Uniformly-Discounted Prioritarianism effectively involves indexing the output of the Prioritarian function to the dimension of time. Unlike my proposed approach, it indexes the entire Prioritarian adjusted contribution of a person to a single time within that person's life (such as her birth year). As a consequence, the person's entire contribution will be discounted at a single rate if discounting is applied. I

claim that my proposal to distribute the Prioritarian contribution across times in proportion to the person's wellbeing levels at those times is superior to this. My approach avoids the arbitrariness in picking a single time to represent a whole life's contribution to goodness, and, unlike Uniformly-Discounted Prioritarianism, my approach does not favour inequality among people of different ages who are living contemporaneously since this inequality results from each person's contribution being indexed to a single time (which differs between people) and then discounted.

VI. Combining Discounting with Other Pattern Goods

If the output to a standard Prioritarian function can be used as input for a time discounting function, then the combination of these two pattern goods would satisfy the criteria I discussed in Chapter 2. This would mean that an axiology could represent both Prioritarianism and time discounting without risking a money pump and without the drawbacks of Time-Slice Prioritarianism and Uniformly-Discounted Prioritarianism.

At this point I have defended the claim that Prioritarianism can justifiably be represented as function that outputs person and time indexed general good. Furthermore, this output can easily be converted to time indexed general good by summing the contribution made by each person at that time. In order to make Prioritarianism (along with any other pattern goods that can output time indexed general goodness) compatible with discounting, I need to justify using time indexed general goodness (as opposed to time indexed wellbeing) as input for the discounting function.

If discounting can justifiably take time indexed general good as an input, then it can be represented by an expression such as the one below.

(12)

$$G = \sum_t R(t) \sum_{i \in N(t)} g(i, t)$$

This is identical to the basic discounting function I described in section IV of Chapter 2, except that rather than taking the sum of each person's wellbeing at time t and then discounting it, this expression takes the sum of each person's contribution to general good at time t , and then discounts this value. (The function g represents general good indexed to person p and time t .) This process is then repeated for each time, and the results for each time are summed to determine general goodness (across all time) from the perspective of the present moment. This approach does not generate the money pump I described in Chapter 1, because the pattern good that makes times interdependent (Prioritarianism) has its output used as input to the discounting function and not the other way around.

Discounting is implicated in most of the problems that can arise when expressing multiple pattern goods in an axiology. Therefore it is worth investigating how easily it can be represented in a way that avoids these problems. If discounting can take time indexed general good as an input, then it may be possible to combine discounting with a number of other pattern goods whilst still following the criteria laid out in Chapter 2, and thus avoiding problems like the money pump. If it turns out that discounting can't be represented in any way that allows it to successfully combine with other pattern goods, this may constitute an argument against discounting.

A discounting function outputs evaluations of general good from a specific temporal perspective (assuming that it employs relative discounting; with neutral discounting, the output would be general goodness that is not perspective indexed). Determining the appropriate *input* for the discounting function is somewhat more complicated. There are a number of different arguments in favour of discounting, and these arguments implicitly support different types of inputs. In what follows, I will give an overview of a number of the most prominent arguments for discounting and discuss and the kinds of inputs that they would support (assuming that they do make a compelling argument for discounting). I will also briefly survey criticisms that have been made in response to these arguments.

VII. The Two Main Types of Arguments for Discounting

The arguments in favour of discounting can be partitioned into two categories those which argue that future wellbeing is intrinsically less significant than present wellbeing, and those which make no claim about the intrinsic value of wellbeing at different times, but instead argue that discounting should be used for reasons of practical necessity.

Claiming that future wellbeing should be treated as less important amounts to an argument to use wellbeing as the input to the discounting function, whereas arguments for discounting on the basis of pragmatic reasons are more open-ended concerning what the input to the discounting function should be; importantly though, it is consistent with pragmatic discounting to claim that general goodness at a time should be discounted (as opposed to just wellbeing at a time). In fact, depending on what considerations beyond wellbeing are at play in an axiology, it might be necessary for a pragmatic approach to use time indexed general good as an input so that the undiscounted value of future goods besides wellbeing does not generate the problems that discounting was intended to prevent.

If one of the arguments on pragmatic grounds is successful, then discounting can take time indexed general good as its input, and can therefore be made compatible with a number of pattern goods, including Prioritarianism. If arguments for future wellbeing mattering less are the only successful ones, then discounting can only take wellbeing as its input, and it cannot be represented in a way that is compatible with pattern goods such as Prioritarianism. Of course, if one is unconvinced by either form of argument for discounting, then the problems around combining discounting with other pattern goods will be of little consequence.

VIII. Standard Discounting Expresses Two Kinds of Pure Time Bias

Arguments in favour of discounting on the basis of future wellbeing mattering less face a unique challenge. Discounting, as it is typically represented, actually embodies two distinct kinds of views regarding the importance of future wellbeing. It expresses both the view that, within each person's life, the wellbeing they will experience at future times is less important the further it is into the future. It also

expresses the view that the wellbeing of people who are born further into the future is less important than the wellbeing of people who are born near to the present.

Bias regarding people who live at different times involves assigning less weight to people's wellbeing the more temporally distant their lives are from the present moment. So the wellbeing of people who live at later times would count for less than the wellbeing of people who live at times close to the present. For the sake of convenience, I will call this *Intergenerational Time-Bias*.

Bias regarding the times within a life involves valuing the wellbeing at different times in a person's life differently. This commonly manifests as a bias people display in their prudential decision-making; for instance there is a substantial body of empirical research demonstrating that people often prefer to receive a medium sized benefit soon rather than a larger benefit later. In this case people seem to be valuing the wellbeing that occurs later in their lives less highly than the wellbeing that occurs earlier, or closer to the present moment. This attitude can also be expressed in axiology by adjusting the contribution that the wellbeing across each person's life makes to overall goodness such that wellbeing that occurs closer to the present moment is more valuable than wellbeing that occurs later. For the sake of convenience, I will call this *Within Life Time-Bias*.

Shane Frederick (2006) argues that: "Although these two types of time preference are often confounded, there is no close connection between them. One person may weight their own future utility the same whether it occurs in the near or distant future, but have no concern about consequences that affect others who will be alive after they die" (p. 675). Although I agree with Frederick that the views that underpin these two types of discounting are independent, it seems to me that, in practice, it is difficult to formulate a satisfactory method of discounting that does not express both of these kinds of time bias at once.

Suppose one wishes to create an axiology that expresses Intergenerational Time-Bias but not Within Life Time-Bias. One way to represent this view would be to hold that, depending on the number of years a person will be born after the present moment, that person's utility is given a diminished weight. This could be applied

backwards and forwards in time so that prior generations were weighted differently to present generations, or it could be applied only from the present moment forward. Either way, a person's utility would receive the same weighting throughout her entire life, as determined by her birth year. This is effectively the same as Uniformly-Discounted Prioritarianism, which I discussed in Section V of Chapter 1, except that it omits the application of a concave function to wellbeing. Uniform Discounting, as I shall call it, can be represented by the following expression:

(13)

$$G = \sum_i R(b_i) \left(\sum_{b_i < t < d_i} w(i, t) \right)$$

Each person's lifetime wellbeing is determined by summing his or her wellbeing at each time that he or she is alive. This lifetime wellbeing is then discounted at a factor set by a single time in that person's life, for example the year of birth.

This form of discounting has the same drawbacks as Uniformly Discounted Prioritarianism (discussed in Chapter 1); the particular time within a person's life that is set to determine the discount rate seems arbitrary, and this approach consistently favours inequality among people who are alive at the same time but who were born at different times. It will give greater weight to the wellbeing of the oldest people alive at any given time. Consequently, this does not seem to be a satisfying way to represent the view that wellbeing is worth less in lives that begin later but not at later times within a life.

Alternatively, suppose an axiologist wishes to represent only Within Life Time-Bias, and not Intergenerational Time-Bias. This approach also seems to be difficult to satisfactorily express in mathematical terms. The standard approach of determining the discount factor for each time based on its relation to the present moment cannot be used here, since this would introduce intergenerational discounting (people who live at later times would be more significantly affected). One option is to discount everyone's lives individually, in a way that is not relative

to the present time, so that the beginnings of their lives are not discounted at all and each person's personal discount rate gradually increases over the course of her life so that the ends of their lives are the most discounted. The first year of each person's life would receive full weight, and each subsequent year would receive less. Formally, this would involve discounting each person's wellbeing at each time individually (rather than aggregating across times first and then discounting). When discounting a person's wellbeing at a time, the discount factor would be determined by treating that person's birth year as $t = 0$, so that, irrespective of the time at which a person lives, the first year of each person's life will be assigned the same discount factor, and so on.

This approach has two unintuitive consequences. Firstly, it has the effect of making people's past wellbeing more significant than their future wellbeing. This is unintuitive by itself, but it might cause further problems by encouraging actions that mimic sunk-cost fallacy behavior. Some people believe that a person's wellbeing at a past time can be retroactively improved in some cases. For example, if at a previous time I had invested a great deal of energy in some project, the eventual outcome of that project might retroactively alter my wellbeing at the time I was working on the project. If the project is a great success, then the effort I spent on it in the past is vindicated, and my wellbeing at that past might be thought to be higher in accordance with this. Supposing something like this story is true, the method of discounting I have outlined might recommend forgoing more beneficial future projects in an attempt to make something of a relatively unpromising project one had previously invested in, because the past wellbeing that could be attained by vindicating previous efforts would be discounted less than any future wellbeing resulting from the actual completion of the projects. Another unintuitive consequence of this approach is that it will favour inequality among people living contemporaneously, except, unlike the only-intergenerational approach, it will give preference to increasing the wellbeing of the youngest people living at any given moment. Not only is this sort of inequality morally undesirable, it also means that

this approach has failed to express Within Life Time-Bias without introducing Intergenerational Time-Bias ³⁴.

Although I agree that Frederick (2006) and others ³⁵ are right to distinguish these two separate sorts of time-bias and the possible justifications that go with them, I do not think a theory that attempted to apply one but not the other of these forms would be appealing due to the difficulties I have discussed. It may be that a satisfactory representation of distinct forms of time-bias is possible, but I have not been able to identify one in the literature or devise one myself. Until such a time as one is proposed, standard discounting seems to me the better alternative. It seems that, at least for the moment, the most appealing version of discounting available is one that makes these two kinds of time bias inextricable. This places extra strain on pure time preference arguments for discounting as they must be able to justify treating wellbeing at later times within each person's life as less important *and* justify

³⁴ At present I have only encountered one example of a theorist attempting to devise formal representations of discounting that keeps these two kinds of time-bias distinct. See the following quote from Partha Dasgupta:

In work under preparation, I have tried to construct a framework that builds an intergenerational welfare economics admitting the idea of selfhood. The model I have constructed permits someone to discount his own future felicities in any way he likes (that's the demand of his "self"), but requires of him to give a weight to the lifetime well-being of each of his children that equals the weight he gives to his own lifetime well-being. The model would seem to reconcile the widespread finding from consumption behaviour that people do discount their future felicities at a non-negligible positive rate ... and the philosophical injunction that many people would seem to adhere to, namely, that they should not discriminate against their children's futures (2008, p. 147).

The approach proposed here by Dasgupta would make it possible to represent Within-Life Time-Bias without introducing Intergenerational Time-Bias. However, as far as I am aware, this work has not yet been published, so I am unable to comment on the success of the formal representation Dasgupta has in mind.

³⁵ See for example Greaves (*n.d.* pp. 28-29).

treating the wellbeing of people who live later as being less important. As we shall see, the standard arguments in favour of pure time preference do this with varying degrees of success.

IX. The Democratic Representation Argument

The first argument I will consider in favour of discounting is typically aimed at axiologies that are to be used to determine the actions of a government or similar organization. It holds that policy decisions should be selected on the basis of aggregating the preferences of citizens, including preferences regarding the relative importance wellbeing at different times. There is a substantial body of empirical research that suggests that people do in fact discount future wellbeing, treating benefits and harms in their far future as less important than benefits and harms in their near future³⁶. On the basis of this, it is argued that an axiology intended to democratically represent citizens' preferences should also discount future wellbeing.

There are two commonly cited sources of information on people's preferences regarding discounting. One is the data accumulated on what people choose in studies designed to test the extent to which people discount their own future wellbeing. The other is market interest rates, which purportedly reflect the people's willingness to save for the future.

People frequently express attitudes and make choices that suggest that they discount future wellbeing within their own lives. This attitude might be described as impatience or a desire to enjoy benefits sooner rather than later, even at some cost to the size of the benefit. The empirical research on this subject demonstrates that people are often willing to sacrifice a certain benefit in the further future in order to gain a smaller benefit in the nearer future. However, as I will discuss further below, there is also a great deal of variety in the specific discount rates that people's choices suggest. This makes it more difficult to base an axiology's discount rate on such preferences.

³⁶ This empirical literature is surveyed in (Frederick, Loewenstein, & O'Donoghue, 2002).

Within Life Time-Bias can be observed in people's observed preference for receiving a smaller benefit sooner rather than a larger benefit later. The substantial evidence that people do hold Within Life Time-Bias could perhaps justify representing this kind of bias in public decision-making. However, this by itself is not enough to justify discounting as it is typically construed. As discussed above, the standard representation of discounting expresses both Within Life Time-Bias and Intergenerational Time-Bias. Therefore, for the democratic representation Argument to justify its use, it would also need to be the case that people expressed time-bias not just within their own lives, but also towards future people. I have not yet encountered any empirical literature on this, though it seems likely that people do tend to hold Intergenerational Time-Bias.

There are serious difficulties with attempting to create a decision-making framework that is representative of people's attitudes towards time. As Frederick describes:

It is unclear what rate should be used, because the descriptive research in this area offers little guidance. Over the last two decades, psychologists and economists have conducted dozens of empirical studies attempting to estimate individual's implicit discount rates. The implicit discount rates in these studies are spectacularly variable, ranging from negative to several thousand percent per year (see Frederick, et al, (2002). Moreover, researchers are often agnostic as to what, exactly, the implicit discount rates imply – whether they reflect time preference, or some of the other considerations ... (e.g., uncertainty, expectations of greater future wealth, or the perception of opportunity costs) or both. The descriptive studies rarely attempt to isolate or assess the relative effects of these different considerations (2006, p. 675).

Another problem with using observed time preferences to inform policy is that empirical studies typically reveal people to have attitudes that match what is known as hyperbolic discounting. Hyperbolic discounting involves using a discount rate that declines over time such that the difference in discount factor from year to year starts off high and gradually decreases. In real terms, this means that the function can lead to preference reversals over time. In particular, it can lead to the nefarious

sort of preference reversals that involve changing one's preferences regarding events that have not yet come to pass. This is the same type of preference reversal that allows the money pump I described Chapter 1 to occur. Frederick, et al, (2002, p. 360) provide an excellent summary of the evidence for the fact that people display hyperbolic discounting.

Although there is strong evidence that people typically display a preference for hyperbolic discounting as opposed to the form of discounting I have been discussing up until now (exponential), advocates of discounting on democratic grounds nonetheless propose using an exponential discounting function in axiology. Presumably this is because hyperbolic discounting is generally regarded to be irrational and because it can lead to money pumps. By electing to use exponential discounting rather than hyperbolic discounting, proponents of the democracy argument avoid these problems. However, this move does somewhat undermine the claim that the reason discounting should be used is because it is important to democratically represent people's preferences. However the defender of this approach could respond by saying that exponential discounting is actually a truer representation of people's preferences since it is the form of discounting they would endorse if they were properly informed about the negative consequences of hyperbolic discounting.

Even non-hyperbolic discounting is often considered irrational. Many theorists attribute the discounting of wellbeing to a cognitive illusion or a weakness of will. They argue that rationality requires impartial concern for all times in one's life. Some of these theorists argue that public decision-making ought not to reflect such irrational attitudes, even if people do hold them (see, for example Goodin (1982); Groom, Hepburn, Koundouri, & Pearce, (2005).

Another argument against the democratic approach directly attacks the notion that ethicists and economists should recommend discounting on the basis of peoples' observed preferences. This argument holds that the recommendations of such theorists are best construed not as a summary of the democratic process, but rather as a further input to it. The contributions of these theorists can help to inform public opinion concerning arguments for discount rates that may diverge from the ones

they express via their behavior (see Cowen & Parfit (1992, pp. 145–6) and Broome (2008)).

Finally, it has been argued by a number of people that it would be wrong to devalue the wellbeing of future people on the basis of the current generation's preferences. In fact, some argue that it is actually part of the role of democracy, properly construed, to act as a trustee for future generations and guard against such selfish and shortsighted behavior on the part of the present generation. The different positions on this issue are well represented by Stephen Marglin, who argues that the state should represent only the present generation, and Arthur Pigou, who holds that the state must also reflect the interests of future generations.

I want the government's social welfare function to represent only the preferences of present individuals. Whatever else democratic theory may or may not imply, I consider it axiomatic that a democratic government reflects only the preferences of the individuals who are presently members of the body politic (Marglin, 1963, p. 97).

There is wide agreement that the State should protect the interests of the future *in some degree* against the effects of our irrational discounting and of our preference for ourselves over our descendants. ...It is the clear duty of Government, which is the trustee for unborn generations as well as for its present citizens, to watch over, and, if need be, by legislative enactment, to defend, the exhaustible natural resources of the country from reckless spoilation (Pigou, 1932; emphasis in original).

X. The Special Relations Argument

The second prominent argument for treating future wellbeing as less important holds that we owe special consideration to the people who are temporally near to us. The thought goes that, just as it seems reasonable to have special concern for the wellbeing of friends and relatives, so too the public decision making of a group of people should give preference to the wellbeing of those people who are temporally

near to them. As Parfit (1987) puts it, this view would not involve discounting time so much as degrees of kinship.

This approach would justify discounting the wellbeing of people who are born later, so this approach supports intergenerational time bias. However this approach does not seem to justify within-life time bias. The justification centers around having differing levels of concern or different people, not differing concern for a person at different times in his or her life. Perhaps this justification could be combined with the argument from democracy to support standard discounting, but by itself, this argument does not seem to be able to support the Within Life Time Bias that is a consequence of discounting.

XI. Over-demandingness Objection

The remaining arguments for discounting that I will discuss take the form of pragmatic arguments. These arguments hold that a discounting function should be used, not because future wellbeing is necessarily less important, but because discounting is necessary to avoid absurd results. As such, these arguments have no particular need to justify the two kinds of time bias I described earlier. In a sense these arguments go over the top of concerns about time bias by arguing that, whether or not the two kinds of time bias are the most morally correct positions to hold, it is necessary to espouse them to avoid significant problems.

The first argument I will consider, and one of the most popular and compelling defenses of discounting, comes from the fact that, if a maximizing theory is applied without discounting, it will demand enormous sacrifices from the present generation on behalf of people who will live in the distant future. Frederick summarises the rationale as follows:

If the return to capital is positive into the indefinite future, and utility is increasing in consumption, then any amount of utility that the current generation sacrifices by foregoing consumption can be more than offset by greater utility gains to distant future generations (2006, p. 676).

Hilary Greaves highlights the parallels and differences between this objection that the over-demandingness objection that utilitarianism in general faces:

It is instructive to compare this “excessive sacrifice” argument in the context of intertemporal ethics with one that often crops up in the assessment of utilitarianism as a moral theory for the intratemporal case. According to utilitarianism, rich people (like ourselves) should give away the vast majority of our wealth to the desperately poor. Many (e.g. (Scheffler, 1982)) object that this implication is too demanding, and conclude that utilitarianism is false.

The difference in the intertemporal case is that the fully impartial value function asks us to make these large sacrifices for people who are already richer than us. Thus, in practice the ‘excessive sacrifice’ argument is more compelling in the intertemporal case than in the intratemporal case. (The difference arises because of the possibility of generating a larger benefit in consumption terms for others than one’s own sacrifice, thanks to the phenomenon of investment. This effect has no analog in the intratemporal case.) (*n.d.* pp. 15-16).

As a response to this concern, some commentators, such as Arrow (1999), argue that discounting is a necessary part of axiology.

Others have responded by arguing that discounting is not an appropriate way of addressing this concern since it fails to deal with the root problem, which is that consequentialist moral theories can dictate that a person sacrifice too great an amount in order to benefit others. A number of theorists have thus proposed alternate ways of dealing with this problem. For example, Parfit (1987) has proposed a solution to the problem of excessive sacrifice whereby some limit is set on the minimal acceptable amount of utility for any generation to have. An axiology that incorporated this feature would not recommend that a generation continue to deprive itself for the sake for future generations once it had reached this limit. If the limit were set high enough, this might be sufficient to allay concerns about excessive redistribution. Shane Frederick (2006) also places the fault with

consequentialist frameworks, but he argues that these frameworks should be abandoned altogether in favour of something like a Rawlsian approach.

XII. Optimal Investment

Posner and Weisbach (2010) argue that discount rates used in policy making should be the same as the market interest rate. In their approach, discounting is not a way to express any particular ethical position, but rather a pragmatic means of taking into account the opportunity cost of investment. The claim made by Posner and Weisbach is that if the discount rate is not set at least as high as the market interest rate, an axiology will recommend taking actions to benefit future people that are strictly less beneficial to them than investing at market rates. The idea is that one could expend some resources to enact a process that would bring some benefit to a future person, however one could also simply invest that same amount of resources such that the future person would receive the invested resources plus interest. They argue that the discount rate must be set sufficiently high to rule out any actions for the sake of future people that do not generate at least as high a cost benefit ratio as investing the equivalent resources³⁷.

This argument doesn't require that a particular sort of good be used as input: what it does require is that discounting will be done in such a way that the axiology will never recommend pursuing a course of action that provides a worse benefit to cost

³⁷ This argument makes the assumption that money received after an investment period can be converted into a benefit equivalent to that which would have been gained by pursuing the 'less efficient' project. In the context in which the argument was made (debates among economists regarding the appropriate way to discount policy responses to climate change) it is common to assume that all costs and benefits can be described in terms of consumption. This assumption makes it trivial to claim that a monetary investment now can produce the same kind of benefit as spending that money on some other kind of project to benefit future people. In the framework I am discussing though, costs and benefits are described in terms of a range of goods, chief among which is wellbeing. In this context it is not so clear that a monetary investment could be converted into equivalent benefits, so the defense of discounting from market interest rates may not be as compelling here.

ratio than investing equivalent resources at the market rate. This could be achieved by using time indexed general good as the input to the discounting function.

Fleurbaey and Zuber argue that the justification of discounting as a way to avoid sub optimal investments is misguided. They claim that this justification of discounting misrepresents the true purpose of the discounting function, and the context in which policy decisions are made using an axiology. They write:

The purpose of the discount rate ... is to make consumption levels or monetary values comparable across time. It makes it possible to compute the net present value (NPV) of any change to the status quo. If the NPV is positive, the change is an improvement. But this does not mean that this particular change is optimal. In order to choose the best policy or project, one must compare the NPV (computed with whatever discount rate seems appropriate) of *all* options, including ordinary market investments. Clearly, with this methodology, if one option costs less today or pays more tomorrow (or both) than another option, it will be deemed preferable, whatever the discount rate.

There is therefore no danger that adopting a lower discount rate than the market rate could induce inefficient (in other words, dominated at each period) choices. It will only imply making different choices among the efficient (in other words, undominated) options. With a lower discount rate, one will choose to invest more for the future, but one will never be tempted to invest at a low rate of return when a higher rate of return is possible (2012, p. 574).

XIII. Infinite Times Paradox

There is a problem that arises when applying a time impartial axiology to assess competing options that extend into the infinite future. All options will have infinite goodness so there is no best outcome. These results have been established by, e.g., Koopmans (1960), Diamond (1965) and Epstein (1986). In response, some have argued that some form of discounting is necessary in order to nullify the

contribution of goods in the very distant future so that states of affairs will have finite, and therefore comparable, goodness. Since this argument stems from pragmatic justifications and not any particular ethical stance about the value of future wellbeing, it would be appropriate to use time indexed general good as the input for a discounting function intended to solve this problem

This objection depends on the assumption that there will be wellbeing to be enjoyed infinitely far into the future. However, this assumption is implausible given, for example, the eventual heat death of the universe. Even if we were to accept the assumptions of infinite times at which goods such as wellbeing occur, the problem that arises can be resolved by using an arbitrarily small discount rate (see Dasgupta 2008, p. 157). An arbitrarily small discount rate would have negligible impact on any times, even into the distant future, that we are likely to be concerned with, but it would be sufficient to prevent an infinite sequence of goods contributing to overall goodness. So the argument from infinite times does not provide any compelling reason to use a discount rate that would have a meaningful effect on any of the times we are likely to be concerned with when setting policy.

XIV. Conclusion

In order for discounting to justifiably take general goodness as an input, at least one of the pragmatic arguments (Over-demandingness; Optimal Investment; Infinite Times Paradox) must be successful. If one of these arguments is successful, it will allow discounting to be paired with a faithful representation of Prioritarianism without generating a money pump. Specifically, it will be possible to use a combination of the Time Indexed Prioritarianism function (11) discussed in Section V and the discounting function which takes general good as its input (12) as discussed in Section VI. These two expressions can be combined quite naturally by using the output of the function $g(i,t)$ in expression (12) equivalent to the function $g(i,t)$ in expression (11)

(11)

$$g(i, t) = \frac{w(i, t)}{\sum_{t \in N(i)} w(i, t)} \times P \left(\sum_{t \in N(i)} w(i, t) \right)$$

(12)

$$G = \sum_t R(t) \sum_{i \in N(t)} g(i, t)$$

It also follows that discounting would be able to be paired with any pattern good that could justifiably be represented as outputting overall goodness at a time. However, if one does not consider any of these arguments for discounting to be compelling, then it would seem that it is impossible to create an axiology that simultaneously represents both relative time discounting and Prioritarianism without either generating a money pump or using an approach like Time-Slice Prioritarianism that changes the spirit of the original Prioritarian concern in order to make these two views compatible.

The reader will notice that although there are exceptions, time discounting is implicated in most of the problematic results such as money pumps that I have discussed in this thesis. I think a reasonable conclusion to draw from these results is that the arguments presented here constitute a new criticism of discounting; it turns out to be quite difficult to construct an axiology that expresses both discounting and one or more other pattern goods. Given the appeal of many pattern goods, this may persuade some to avoid using discounting, especially since there are alternative measures that can be used to address some of the concerns that discounting is touted as a response to. For example, it is possible to respond to the over-demandingness problem more directly by introducing limits on what a person or generation can be required to give in aid. Another possible solution is to introduce Prioritarian or Egalitarian concerns that would prevent a generation beggaring itself to create high return investments for the sake of wealthy future generations.

Reference List

- Arrhenius, G (2000), 'An impossibility theorem for welfarist axiologies', *Economics and Philosophy*, 16(02), 247-266.
- Arrow, K (1999), 'Discounting, morality, and gaming', in Portney, PR & Weyant, J (eds.), *Discounting and intergenerational equity*, Washington, RFF Press.
- Atkinson, AB & Stiglitz, S (1980), *Lectures on public economics*, New York, McGraw-Hill.
- Brock, DW (2004), 'Ethical issues in the use of cost effectiveness analysis for the prioritization of health care resources', in Anand, S, Fabienne, P & Sen, A (eds.), *Public health, ethics, and equity*, Oxford, Oxford University Press.
- Broome, J (1991), *Weighing goods*, Oxford, Basil Blackwell.
- Broome, J (1992), *Counting the cost of global warming*, Cambridge, White Horse Press
- Broome, J (2004), *Weighing lives*, Oxford, Oxford University Press.
- Broome, J (2008), 'The ethics of climate change', *Scientific American*, 298(6), 96-102.
- Broome, J (2015), 'Equality versus priority: A useful distinction', *Economics and Philosophy*, 31(2), 219-228.
- Cowen, T & Parfit, D (1992), 'Against the social discount rate', in Laslett, P & Fishkin, JS (eds.), *Justice between ages groups and generations*, New Haven, Yale University Press.
- Crisp, R (ed.) (2014), *Aristotle: Nicomachean ethics*, Cambridge, Cambridge University Press.
- Dasgupta, P (2008), 'Discounting climate change', *Journal of Risk and Uncertainty*, 37(2-3), 141-169.
- Davidson, D, McKinsey, JCC, & Suppes, P (1955), 'Outlines of a formal theory of value, I', *Philosophy of Science*, 22(2), 140-160.
- Diamond, PA (1965), 'The evaluation of infinite utility streams', *Econometrica: Journal of the Econometric Society*, 33(1), 170-177.
- Dietz, S, Hepburn, C, & Stern, N (2008), 'Economics, ethics and climate change', in K. Basu & R. Kanbur (eds.), *Arguments for a better world: Essays in honour of*

- Amartya Sen. Volume ii: Society, institutions and development*, Oxford, Oxford University Press.
- Dorsey, D (2015), 'The significance of a life's shape', *Ethics*, 125(2), 303-330.
- Dougherty, T (2011), 'On whether to prefer pain to pass', *Ethics*, 121(3), 521-537.
- Dworkin, R (1981), 'What is equality? Part 1: Equality of welfare', *Philosophy and Public Affairs*, 10(3), 185-246.
- Epstein, LG (1986), 'Intergenerational preference orderings'. *Social Choice and Welfare*, 3(3), 151-160.
- Fleurbaey, M & Zuber, S (2012), 'Climate policies deserve a negative discount rate', *Chicago Journal of International Law*, 13, 564-595.
- Frankfurt, H (1987), 'Equality as a moral ideal', *Ethics*, 98(1), 21-43.
- Frederick, S (2006), 'Valuing future life and future lives: A framework for understanding discounting', *Journal of Economic Psychology*, 27(5), 667-680.
- Frederick, S, Loewenstein, G, & O'Donoghue, T (2002), 'Time discounting and time preference: A critical review', *Journal of Economic Literature*, 40(2), 351-401.
- Gert, B (1990), 'Rationality, human nature, and lists', *Ethics*, 100(2), 279-300.
- Glasgow, J (2013), 'The shape of a life and the value of loss and gain', *Philosophical Studies*, 162(3), 665-682.
- Goodin, RE (1982), 'Discounting discounting', *Journal of Public Policy*, 2(1), 53-71.
- Greaves, H (2014), 'A reconsideration of the Harsanyi–Sen–Weymark debate on utilitarianism', *Utilitas*, 1-39.
- Greaves, H (n.d.), 'Discounting and public policy: A survey', forthcoming in *Economics and Philosophy*.
- Greene, P & Sullivan, M (2015), 'Against time bias', *Ethics*, 125(4), 967-970.
- Groom, B, Hepburn, C, Koundouri, P, & Pearce, D (2005), 'Declining discount rates: The long and the short of it', *Environmental and Resource Economics*, 32(4), 445-493.
- Hare, C (2007), 'Self-bias, time-bias, and the metaphysics of self and time', *The Journal of Philosophy*, 104(7), 350-373.
- Hare, C (2008), 'A puzzle about other-directed time-bias', *Australasian Journal of Philosophy*, 86(2), 269-277.
- Hare, C (2013), 'Time: The emotional asymmetry', in Dyke, H & Bardon, A (eds.), *A companion to the philosophy of time*, New York, John Wiley & Sons, Inc.

- Hirose, I (2015), 'Separability', in Hirose, I & Reisner, A (eds.), *Weighing and reasoning: Themes from the philosophy of John Broome*, Oxford, Oxford University Press.
- Kagan, S (1992), 'The limits of well-being', *Social Philosophy and Policy*, 9(2), 169-189.
- Koopmans, TC (1960), 'Stationary ordinal utility and impatience', *Econometrica: Journal of the Econometric Society*, 28(2), 287-309.
- Marglin, SA (1963), 'The social rate of discount and the optimal rate of investment', *The Quarterly Journal of Economics*, 77, 95-111.
- Ord, T (2015), 'A new counterexample to Prioritarianism', *Utilitas*, 27(03), 298-302.
- Ottersen, T (2013), 'Lifetime QALY Prioritarianism in priority setting', *Journal of Medical Ethics*, 39(3), 175-180.
- Parfit, D (1987), *Reasons and Persons*, Oxford, Oxford University Press.
- Parfit, D (1995), *Equality or Priority?*, The Lindley Lecture, University of Kansas.
- Parfit, D (1997), 'Equality and priority', *Ratio*, 10(3), 202-222.
- Parfit, D (2012), 'Another defense of the priority view', *Utilitas*, 24(03), 399-440.
- Pigou, AC (1932), *The economics of welfare: Volume 1*, London, MacMillan & Co.
- Posner, EA & Weisbach, D (2010), *Climate change justice*, Princeton, Princeton University Press.
- Ramsey, FP (1928), 'A mathematical theory of saving', *Economic Journal*, 38, 543-559
- Rawls, J (1971), *A theory of justice*, Cambridge MA, Harvard University Press.
- Sidgwick, H (1890), *The methods of ethics*, London, Macmillan.
- Slote, M (1984), 'Goods and lives', in his *Goods and virtues*, Oxford, Oxford University Press.
- Temkin, L (1993), *Inequality*, Oxford, Oxford University Press.
- Temkin, L (2011), *Rethinking the good*, Oxford, Oxford University Press.
- Velleman, D (2000), 'Well-being and time', in his *The possibility of practical reason*, Oxford, Oxford University Press.