A Subjective Wellbeing Approach to Some Major Problems of Welfare Economics

Timothy Scriven

A thesis submitted to fulfil requirements for the degree of Doctor of Philosophy

Faculty of Arts and Social Sciences
Sydney University

December, 2023

This research reported in this thesis was supported by the award of a Research Training Program scholarship to the PhD Candidate.

Acknowledgments

For a discussion of the truly outrageous behaviour of the University of Sydney in relation to this thesis, particularly in imposing minor corrections after suggesting that it would be passed without corrections through a process of numerous procedural and substantive failures, see: https://philosophybear.substack.com/p/the-assessment-of-my-phd-thesis-by or Google: *The assessment of my PhD thesis by the University of Sydney and the way I see our world*

I would like to thank:

My ever accommodating and hardworking supervisors Michael Beggs & Ryan Doody for the comments, corrections, suggestions, administrative support, and kindness.

My colleague and comrade Kieran Latty who provided vast and critically important comments, suggestions, and corrections on my thesis, and effectively acted as an additional supervisor, meaning I was very generously apportioned for supervisors indeed given how hard Mike, Ryan and Kieran worked.

My parents Diane and Ken Scriven for their love, support and encouragement.

John Broome for very kindly agreeing to read part of my thesis and giving very encouraging remarks in conversation.

Numerous of my friends who gave feedback and guidance. I will not list them because I would inevitably leave someone off the list by accident. Riki Scanlan, however, deserves a special mention.

I would also like to thank *Ursus Arctos* and the clade *Ursidae* in general, and a certain nosey and mischievous friend.

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged

.

Signature:

Name: Timothy Scriven

Abstract

I outline an approach to applied welfare economics: 1. Grounded in the psychometrics of subjective well-being. 2. Focused on providing *information* about the effects of policy useful to discussion by the democratic public and decision-makers. I argue this approach can help solve many of the traditional problems of welfare economics such as interpersonal comparison, cardinalisation, the proper role of values in welfare economics and the question of where to draw information about an agent's preferences. Concerning the problem of interpersonal comparison, I argue that a moderate form of philosophical functionalism, combined with a belief that psychometrics does what it claims to do can resolve the problem. Concerning the problem of cardinalisation, I argue that several potentially viable strategies already exist for cardinalizing psychometrics and common claims regarding the dangers of treating psychometrics as *already* cardinal may be overstated. Concerning the problem of values in welfare economics I argue that providing information on the likely effects of policy on SWB metrics avoids both the problem of being morally and politically irrelevant by not saying anything of ethical or political significance and the problem of ethical sectarianism- adopting strong value premises that make one's work irrelevant or repugnant to those who do not share those premises. Regarding the problem of getting information about an agent's preferences, I argue that psychometrics are likely to be superior to revealed preference approaches because, among other issues revealed preference approaches are insensitive to the distinction between *instrumental* and *intrinsic* preferences. Finally, I conclude by considering a variety of other issues- wireheading, arguments against weighted cost-benefit analysis and directions for future research.

One Mite wrung from the Labrers hands
Shall buy & sell the Misers Lands
Or if protected from on high
Does that whole Nation sell & buy

- William Blake, Auguries of Innocence

“Fresh from his ashram
deep in the Himalayas,
        barefoot,
        robes flowing,
        incense burning,
…. beads in one hand, well-being budget in the other”.

- Found poetry from the words of Josh Frydenberg, former Australian treasurer

Front matter

Table of contents

Table of figures

Glossary of mathematical terms

Description	Notation	Definition	Notes
Income	y		
Mean income	$\bar{y}$		
Welfare/utility	u		Although we use u for welfare, it is important to note that welfare need not be defined in such a way as to correspond with the economist's decision utility
Subjective wellbeing	u_{swb}		With the caveat that there are many forms of subjective wellbeing
Isoelastic/CES utility function	$u = f(y\|\eta)$	$u = \frac{y^{1-\eta} - 1}{1 - \eta} + c$	
Marginal welfare/utility of income	$\frac{\partial u}{\partial y}$		The change in welfare or utility from an additional unit of income. Generally thought to decrease in income (e.g. an additional dollar means more to the

			poor than the rich).
Income elasticity of marginal welfare/utility of income	η	$\eta = -\frac{\partial \ln \frac{\partial u}{\partial y}}{\partial \ln y}$	$\frac{\partial u}{\partial y} = y^{-\eta}$
Individualistic SWF	$W = \sum_{i=1}^{n} v_i$		In an individualistic SWF, total social welfare is the sum of valuations v of individual income or welfare.
Value function (income)	$v = f(y)$		Defines valuations over income (not, say, welfare)
Value function (welfare)	$v = f(u)$		Defines valuations over welfare (not, say, income)
Utilitarian value function	$v = u$		Valuation of a person's welfare by some utilitarian SWF is simply equal to their welfare, and the welfare function is individualistic, as utilitarian SWF equals the sum of individual welfares.
Isoelastic value function (income)	$v = f(y\|\varepsilon)$	$v = \frac{y^{1-\varepsilon} - 1}{1 - \varepsilon} + c$	Value is as isoelastic function of income. This valuation need not have any welfarist basis.
Epsilon	ε	$\varepsilon = -\frac{\partial \ln \frac{\partial v}{\partial y}}{\partial \ln y}$	The income elasticity of marginal valuation of income. A measure of income inequality aversion.
Welfare weight (income)	w_y	$w_y = \frac{\partial v}{\partial y}$	Under an isoelastic value function on income, we have $w = y^{-\varepsilon}$
Welfare weight (utility)	w_u	$w_y = \frac{\partial v}{\partial u}$	Welfare and value are linearly related in utilitarianism. In

			prioritarianism value is concave in welfare.
Prioritarian value function	$v = f(u)$	$\frac{\partial \frac{\partial v}{\partial u}}{\partial u} < 0$	The function is concave because welfare has declining marginal value in prioritarianism.
CES priority function	$v = \frac{u^{1-p} - 1}{1 - p} + c$	$\frac{-\partial \ln \frac{\partial v}{\partial u}}{\partial \ln u} = p$	The only possible scale invariant prioritarian value function. Cannot handle negative utilities.
Atkinson index	A_ε	$\begin{cases} 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\bar{y}}\right)^{1-\varepsilon}\right]^{1/(1-\varepsilon)} & \varepsilon \neq 0 \\ 1 - exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{y_i}{\bar{y}}\right)\right] & \varepsilon = 1 \\ 0 & \varepsilon = 0 \end{cases}$	
Atkinson SWF	$W_A = f(\mathrm{Y}\vert\varepsilon)$	$W_A = \bar{y}(1 - A_\varepsilon)$	The welfare from a given income distribution, expressed as a level of income such that if everyone had that income, social welfare would be identical to that produced by the given income distribution.

An ordinary language introduction to the thesis and why it matters

Welfare economics is the study of economic policies and economic arrangements affecting welfare, where welfare is understood as wellbeing or the degree to which a person's life is going well for them.

Welfare economics is often used to evaluate policy, and in that role, it has traditionally faced a problem: what to do when a policy, affects different people differently? How can we compare Oscar's harms to Jennifer's benefits? But even if we could collate all the different benefits and losses, and make them comparable to each other, how could we *evaluate* them against each other, let alone decide which gains and losses we should care about more? The enormous prestige of economists and the occasional taste for technocracy means that many policy makers would like economists to be able to do this for them- to deliver a neutral ground truth on what the right policies are, or at least which policies increase human welfare.

There are a number of obstacles. Firstly, there is the issue of interpersonal comparison- what does it mean to say that Jennifer's benefit (or loss) from a state of affairs is greater than Oscar's benefit (or loss) from the same or another state of affairs. How could we discover this?

The second is the issue of magnitudes- sometimes called the problem of cardinality. How can we quantify benefits as if they were temperatures on a thermometer? Even if we restrict ourselves to one person, what does it mean to say that x was very much worse for Oscar than y and not merely worse? Or that the difference between 25,000 a year and 50,000 dollars a year matters at least as much to Oscar as the difference between 50,000 and 100,000 dollars a year?

Finally, there is the problem of value judgements. Even if we could quantify all the benefits and harms Oscar and Jennifer face and show that policy P does more benefit than policy Q some people might still suggest that policy Q is superior. Let's say policy Q, despite having less benefit overall, benefits Jessica more. Some people might prefer policy Q if they feel that Jennifer is harder working, if they feel that Jennifer is worse off and therefore should be prioritized for benefits, or if they feel that Jennifer has been historically oppressed and must be compensated. How can economics, often held to be a positive science, possibly choose between these value frameworks?

Right now, welfare economics has methods of resolving each of these problems in practice. However, the solutions used are subject to many critiques. To restrict myself to a selection of criticisms I agree with:

1. Current methods often unjustly favour the rich over the poor.
2. Current methods rely on psychologically implausible models
3. Current methods make very strong assumptions about what it means to live a good life

4. Current methods are in tension with a democratic approach to economic decision making.

Meanwhile, there have been efforts to psychologise welfare economics for a long time- to combine it with the burgeoning literature on wellbeing psychology. My argument in this thesis is that a particular way of psychologising welfare economics can solve- or at least has extremely good prospects for solving, all three of those problems I listed above- the problem of interpersonal comparison, the problem of cardinality and the problem of the proper role of values in applied economic science. It is also not subject -or at least it is much less subject- to the critiques of current welfare economics I outlined above. In other words, I argue that a certain type of psychological approach to welfare economics can solve many of the paradoxes and antinomies that welfare economics faces.

This approach consists in 1. Estimating the effects of policies on the level and distribution of various psychometrically measured forms of wellbeing. 2. Providing this information to policy makers and the public as opposed to making value judgements ourselves. Although conceptually simple I will argue this can get around the three problems I list.

Addressing these issues matters, because arguably there is no single area of practice in the humanities or social sciences that has such a direct impact on our lives and wellbeing as welfare economics, applied through the policies of treasuries, agencies, infrastructure funds, international institutions, non-government organizations and many others.

0 Introduction and precis

0.1 General considerations

This is a thesis about *applied welfare economics* -a phrase some might contend is, or ought to be a tautology and others might say, in exasperation, is an oxymoron. In this thesis, applied welfare economics refers to welfare economics intended to directly shape or inform policy. Examples of what we are calling applied welfare economics include cost-benefit analysis (R. H. Frank, 2000; Adler & Posner, 2006), designing tax systems through optimal tax theory (Mirrlees, 1971, 1976, 1986; Boadway, 2012), designing institutions, planning intertemporal tradeoffs (Ramsey, 1928; Cass, 1965; Koopmans, 1965; Ponthière, 2003; Summers & Zeckhauser, 2008; Y.-K. Ng, 2011; Dasgupta, 2019), devising measures of economic and social progress (Hamilton, 1998; Kubiszewski et al., 2013; Fitoussi & Stiglitz, 2011; Stiglitz et al., 2015) and of income inequality (Dalton, 1920; Aigner & Heins, 1967; Atkinson, 1970; Osmani, 1978; A. K. Sen, 1982; Anand & Sen, 2000; Atkinson & Brandolini, 2015) for use as policy targets or policy evaluation, and the economic analysis of law (R. A. Posner, 1977; Kaplow & Shavell, 2002, 2000). We use cost-benefit analysis almost exclusively for our examples, but *mutatis mutandis,* we believe that the approach we sketch in this thesis could be applied to almost all of these-although showing this in detail would require additional work to deal with the subtleties of each area.

Many criticisms have been made of the various approaches that make up orthodox welfare economics (Stigler, 1943; Little, 1950; Streeten, 1953; Chipman & Moore, 1978; A. K. Sen, 1979b; Cooter & Rappoport, 1984; Blaug, 1992; Scarantino, 2009; Hands, 2014; Backhouse, 2016), but these criticisms have sometimes languished for want of an alternative paradigm.

We aim to address this problem, vis a vis applied welfare economics. In this thesis, I claim to locate an alternative paradigm in already existing research- the study of subjective-wellbeing applied in the context of economic policy (B. S. Frey, 2008; B. S. Frey & Stutzer, 2002; Layard, 2005; Y.-K. Ng, 2003b; Y.-K. Ng & Ho, 2006). Such research already exists and has been proposed as an alternative basis for evaluating economic policy (Diener & Seligman, 2004; Frijters et al., 2020; Kahneman et al., 1997; Kahneman & Sugden, 2005; Y.-K. Ng, 1999, 2003a; Oishi & Diener, 2014; Stiglitz et al., 2015). However, my purpose is to defend the idea that a subjective-wellbeing oriented approach can address many of the needs we want welfare economics to address while at the same time offering a resolution to many of the traditional problems and limitations of welfare economics. I call this approach by various names including the *psychometric approach* and the *subjective-wellbeing effect estimation* (SWBEE) approach. For clarity, my proposition is not that this approach to welfare economics is the only possible approach that resolves the problems I consider, but only that our approach can well be applied to both the traditional uses, and the traditional difficulties, of welfare economics and is thus a promising paradigm.

My approach in this thesis is somewhat unusual. This is a thesis on the philosophy of welfare economics, but my goal is to describe a potential working consensus in the field. It is then, somewhat strangely for a work of philosophy, about trying to piece together a strategy and conceptual apparatus that is robust enough to serve as a working approach. This means having to put aside many valid criticisms and additional research questions, and treat many philosophically controversial processes and ideas, but particularly the psychometrics of wellbeing, as tools, rather than, primarily, objects of inquiry in their own right.

In this kind of practical enterprise, the right question is not “is the method robust to all criticism”, but rather “is the method more robust than its plausible rivals”. It is my impression that part of the reason the New Welfare Economics has remained so obdurate, despite its conceptual deficiencies and inegalitarian tendencies, is because its critics have been concerned to criticise it, rather than describing a workable alternative that can be rolled out by civil servants. The sciences, especially the applied sciences, cannot always wait for all philosophical questions in an area to have been answered to proceed, and that includes the social sciences. It would be unfortunate to keep using flawed rules like Kaldor-Hicks for want of a perfect alternative- it would be to make the perfect not just the enemy of the good, but the ally of the bad. This is in no way intended to undermine criticisms of the concepts the approach I describe relies upon, and as time goes on, it is my hope that these concepts and practices will be further refined by criticism.

Relatedly there is a growing dissatisfaction with the idea that a technocratic happiness economics can replace a technocratic new welfare economics (Fabian et al., 2023; Singh & Alexandrova, 2020). In broad terms, I agree. I do not think policies should be selected algorithmically by technocrats trying to maximise happiness or life satisfaction, I agree that many have tried to frame happiness economics in such terms- as replacing the numéraire of the social planner- with happiness instead of dollars- and this is regrettable. I believe, as I will outline in chapters 7 & 8, that there is a better alternative- treating welfare economics as an information gathering step before public debate, always responsive to public needs and criticism, and not a decision-making procedure. However, I also think that there is a need for an agreed upon manual of policies and procedures for civil servants which, while not taking the place of ethical judgement, prevents the applied welfare economist from having to start from the beginning each time. Applied welfare economists usually do not have the time to conduct extensive consultations on axiological questions with the public and with ethicists, political philosophers and other experts in values. Technocracy is the replacement of democratic decision making with expert rule, not a degree of standardisation in gathering information before public debate.

In this thesis, we talk about subjective wellbeing. This phrase tends to imply we are talking about wellbeing as experienced by the subject. We do so because this is the language that is often used in the literature. However, our main emphasis is on psychometrically quantifiable forms of wellbeing. There are quantifiable, psychometrically defined forms of welfare that are not subjective- or that are only partially subjective. For example, certain aspects of the concept of eudaimonia. References to subjective well-being should thus be read as references to psychometrically quantifiable welfare, including aspects of welfare which are not purely ‘subjective’ where the additional breadth would make a difference unless context indicates otherwise.

0.1.1 Chapter one: The aggregation problem

In this chapter, I briefly consider traditional criticisms of new welfare economics- especially in the area of applied welfare economics. These criticisms include its formalistic character, often weak conclusions, heroic assumptions and tendency to adopt arguably implausible ethical standpoints in an attempt to avoid taking on any controversial ethical assumptions at all, for example in the case of the Kaldor-Hicks (Kaldor, 1939) criterion.

The best way to understand many of the problems of welfare economics is to understand them as constituents of a broader problem- the problem of how-to best weigh and balance interests when a policy affects different groups and their interests differently. In this thesis, we call this problem the *problem of interpersonal aggregation*. It is important not to confuse the problem of interpersonal aggregation with the problem of interpersonal comparison, although a solution to the interpersonal aggregation problem may require a solution to the interpersonal comparison problem. These problems are distinguished in our usage in the thesis as follows. The interpersonal aggregation problem is the problem about how to compare either overall goodness, or perhaps just the welfare-related portion of goodness when a menu of policies has costs and benefits accruing to different people. The interpersonal comparison problem -as we use the term in this thesis- is the problem of how to compare the magnitude of a mental state of some sort- perhaps a desire, emotion, “utility” (conceived of as a mental state whatever that may be) or even a personality trait between two or more persons.

We argue that three separate problems, viz: 1) the interpersonal comparison problem, 2) the problem of cardinal measurement of mental states, and 3) the problem of value judgements in welfare economics can all be seen as components of the interpersonal aggregation problem. Without comparability, cardinality and some answer to the question ‘what do we value and to what degree” (or some principled way to evade the problem), coming to an overall evaluation of the effects of a policy with disparate effects on disparate people is all but impossible.

0.1.2 Chapter two: a typology of welfare economics

In this chapter I present a taxonomy of many different approaches to welfare economics, organized by their relation to the interpersonal aggregation problem. I argue that existing approaches tend to either suffer from limited scope, limited relevance to the problem of selecting economic policy, a presumptuousness about ethics that is unlikely to be persuasive in a democratic and ethically pluralistic society, or some combination of these problems.

Having considered the problems with traditional approaches to applied welfare economics, we begin to lay out our alternative approach over the next two chapters.

0.1.3 Chapter three: an introduction to the psychometrics of subjective wellbeing

First, we provide an explanation of standard psychometric techniques and concepts that will be relevant to this work. This includes an introduction to the concepts of validity (& validation) and reliability.

Then we briefly review the state-of-the-art, both philosophical and scientific, on psychometric measures of happiness. Although our primary aim in this thesis is not the evaluation of the psychometrics of wellbeing as a field of study, my general conclusion is that the science of subjective wellbeing is reasonably robust.

There are a variety of different welfare concepts in the philosophy of wellbeing. These include *desire satisfaction*, *the balance of pleasure over pain* and *flourishing.* I briefly review special philosophical and methodological questions in the psychometric measurement of each. I argue that psychometrics can evaluate each of these types of wellbeing so there is no need for our methodology to be sectarian about ethics, although psychometric measurements of desire satisfaction are sadly underdeveloped at present.

0.1.4 Chapter four: subjective-wellbeing effect estimation

In this chapter, I lay out my preferred approach to welfare economics. The goal of the applied welfare economist is not to make *judgements* about what policy is optimal but to provide *information* to decision-makers and the public about the probable effects of policy on a variety of different metrics of subjective wellbeing (especially the distributional parameters and average of subjective wellbeing). Application of the method will require judgment, especially in the initial stages as the canons of procedure are still being developed. Our thesis is mostly concerned with *theoretical rationale* and I do not aim to provide an exact manual for the methodology I describe in broad terms.

I call this approach Subjective Wellbeing Effect estimation. Interestingly, SWBEE provides a rationale for something like traditional (weighted) cost-benefit analysis (Kelman, 1981; Brent, 1984; Harberger, 1978; Johansson-Stenman, 2005; Adler & Posner, 2006; Adler, 2016)) as one method of pursuing the goal of subjective-wellbeing effect estimation.

However, it does so without requiring us to adopt the value judgments inherent in a specific social welfare function, or even a range of social-welfare functions. ***Instead of representing a value judgment as to the social worth of a marginal dollar going to an individual, elasticities represent a best attempt to estimate the probable effects of a marginal dollar on a particular form of psychometrically measured well-being.*** We seek that information not because of endorsement of a utilitarian social welfare function but because information on the wellbeing effects of a policy is useful almost regardless of one’s political or ethical beliefs. Thus our approach can lead to a rationale for weighted cost-benefit analysis without a social welfare function. The rationale is that it can be constructed so as to inform us of changes in SWB.

I then proceed to address one problem in welfare economics per chapter which traditional approaches have struggled with, but which I believe a psychometric approach can resolve. As noted earlier, three of the four problems I investigate can be seen as tributaries to a larger problem- the interpersonal aggregation problem- or how to quantify and evaluate overall gains and losses when different people or groups gain or lose. One other problem we consider is not connected to the interpersonal aggregation problem, at least not directly, but is interesting in its own right: that is the question of the best way to infer desire satisfaction.

0.1.5 Chapter five: the problem of interpersonal comparison

The first problem we consider is the problem of interpersonal comparison of mental states. This problem is a broader problem than the traditional problem of interpersonal utility comparison. Choose some mental attribute, say happiness, anger or desire. How can we compare X’s degree of this attribute to Y’s degree of this attribute? If we define a person’s good in a way that is tied, in at least some degree to their mental states, resolving this problem will likely be necessary to resolving the interpersonal aggregation problem. This problem has traditionally bedevilled economics, in the form of worries about the interpersonal comparison of utility. Thus, if the psychometric approach can resolve this, so much the better for psychometrics.

I argue that psychometrics, when combined with even moderate functionalism in the philosophy of mind, gives us a meaningful way to empirically compare degrees of states such as life satisfaction and happiness between people. Functionalism says that like functional states are associated with like mental states, and psychometrics shows us through the process of construct validation that like scores tend to be associated with like functional states. We can make this result appealing to more people by weakening functionalism from a metaphysical to an epistemic principle. That is to move from saying like mental states are associated with like functional states to saying that it is reasonable to hold that mental states are like when functional states are like- a methodological rule I call weak functionalism. Accepting weak functionalism gives us reason to affirm that comparison of psychometric scores allows for interpersonal comparison of mental states.

We can further supplement this combination of functionalism and construct validation in the psychometrics of subjective wellbeing with concepts drawn from a consideration of Lerner’s principle of equal ignorance. Our argument then becomes not so much that like mental states must be associated with like functional, or even that it is reasonable to assume such, as that assuming this has the lowest expected error of any view.

0.1.6. Chapter six: The problem of cardinality

Even if we could compare X and Y on their degree of happiness, or desire fulfillment, establishing whether one is greater or both are equal, could we quantify the difference between them? At least for many common ways of approaching social welfare, this is going to be of key importance to resolving the interpersonal aggregation problem. For example, if we think the sum or average of wellbeing matters, or if we think a weighted or otherwise modified sum or average is what matters, such cardinal information will be essential.

Welfare economists have traditionally been troubled by the issue of the cardinality of welfare, with the advent of ordinalism being seen as a major innovation in consumer theory, but a major difficulty for traditional welfare economics. The problem also arises in the context of the relationship between underlying wellbeing, and wellbeing scores, and whether the former is linear in the latter. My purpose in this chapter will be to argue that the problem can be solved in the context of the psychometric approach to welfare economics. Thus, it is not a stumbling block to our approach.

To see the problem in a psychometric context, imagine a questionnaire like Cantrill's ladder. Participants are asked to rate their happiness on a scale from 0 to 10. Now suppose that we have three responses: 6, 8 and 10. We might therefore reason that the average happiness is an 8, but unless we have established cardinality this is impermissible. It might be, for example, that the gap between a 6 and an 8 is much larger than the gap between an 8 and a 10, so taking the arithmetical averages of the responses won't really tell us about the average of happiness as such.

I argue that while there are difficulties in establishing that psychometric measures are truly cardinal measures of wellbeing, the following considerations give us reasons for confidence:

1. There is preliminary evidence that subjective wellbeing is linear in subjective well-being scores

2. If scores of measured subjective wellbeing are not already linear in subjective-well being, there are many promising directions for future research on this topic that could uncover the real relationship between subjective-wellbeing scales and underlying subjective-wellbeing.

3. There is evidence that derived statistics are relatively robust to perturbations in the underlying relationship between wellbeing and wellbeing score.

0.1.7. Chapter seven: The problem of the role of values in welfare economics

There has long been a problem about the proper role of values in science, and this problem has often troubled welfare economists in particular. This is linked to the overarching problem of interpersonal aggregation, because to make any headway in cases where what is good for one person diverges from what is good for another it will be necessary to make a judgment of value. For example, we might establish that Susan's subjective wellbeing is improved more by a policy than Bob's loss of subjective wellbeing, but a judgment of value will be needed to decide whether that matters more (and even whether a particular form of subjective well-being should matter at all). Thus, welfare economics seems committed to making such value judgements, but one might worry that this is not the proper place of the scientist- this is our problem.

I suggest that the problem is most profitably divided in two. There is a concern that values may *contaminate* science invalidating its empirical method and positive focus, and there is a concern that, in a pluralistic and democratic society, proceeding on the basis of any one set of values may be inappropriately *sectarian*. In this chapter we will be concerned solely with the problem of *sectarianism* rather than the problem of *contamination*, as we regard the contamination problem as a general problem *within the specific context of an applied science which is intended to be action guiding for policy* and thus not something we wish to give specific comment on in a thesis narrowly applied to welfare economics. We are not concerned with trying to eliminate all values from welfare economics- perhaps an impossible task, but avoiding value sectarianism. In the context of this thesis, we conceive of the problem of sectarianism in a way that is more practical than normative, although of course our work has normative implications. Our problem can be stated so: given that welfare economics is often paralysed by a lack of normative agreement, how may it continue in a way that garners broad social assent?

I argue that the best approach in the context of welfare economics for solving the problem of sectarianism is for welfare economics to see itself not as *prescribing solutions* but instead as *providing ethically relevant information to inform decision makers and/or the democratic public*. An approach to welfare economics grounded in the study of subjective wellbeing is capable of performing this role because it can give us information that almost everyone cares about knowing, even if it will not form the whole basis for their decision making. *Everyone* or almost everyone wants to know about how policies affect subjective well-being. It thus manages to be ethically engaged, without insisting on a specific set of political values, or a specific social welfare function.

Such an approach does not remove us from the necessity of appealing to values in welfare economics, but it gets us out of ethical sectarianism by making the ethical appeals much less controversial. All one must assent to now is that it is worth having information about how different policies will affect subjective well-being before we make a decision. Welfare economics is conceptualized not as making policy recommendations, but as providing information for decision makers and the democratic public. No specific social welfare function or framework of rights etc. needs to be appealed to.

0.1.8 Chapter eight: Cost Benefit Analysis and values in Welfare Economics

Moving to chapter eight, I argue that this kind of strategy is not so readily available to approaches in welfare economics outside the psychometric approach. Conceptualizing the results of investigations such as cost-benefit analysis -whether weighted or unweighted- as mere information is difficult, because it is unclear what that information is, stripped of a normative recommendation, and of what basis there is for caring about that information. Approaches in new welfare economics, e.g., unweighted cost benefit analysis have attempted to avoid controversial value judgements through appeal to democracy. I find these attempts fail, largely because unweighted cost-benefit analysis is not, on examination, a democratic method. I give multiple reasons for thinking this, but at base our main contention is that these methods are more oligarchic than democratic.

Weighted CBA based approaches might be more promising, but these prompt the question “weighted on the basis of what?”. If we weigh on the basis of one set of ideas about what is distributively appropriate, we have run aground on ethical sectarianism again. If we weigh so as to estimate wellbeing effects, than we have effectively come to SWBEE.

0.1.9 Chapter nine: Psychometrics vs self-report in measuring desire satisfaction

How should we measure preference satisfaction? We consider two options: psychometric measures of preference satisfaction based on verbal self-report or *revealed* preference measures of preference satisfaction like willingness to pay, plus assessments of hypothetical buying and selling behaviour through stated preferences. The desire satisfaction theory of welfare has been seen as a strong suit of behavioural approaches to welfare- thus if the psychometric approach proves superior even here, this is a boon for our approach. There is a major problem- at present, no psychometric measure of perceived overall levels of desire satisfaction exists, but as far as I can tell, there are no fundamental barriers to the creation of such a measure.

Although this debate over the right method to measure welfare is not directly connected to the interpersonal aggregation problem like the other three issues we have considered, it does share a conceptual link: when we are contemplating changes which make things better for everyone, it is often less necessary to know the details of what people want or feel- simple models like “everyone prefers to be richer, to have more status, and to have more rights” will do. However, when a change may divide the population, knowing exactly what people want and/or how they will feel with a particular outcome often becomes more important.

I argue that revealed and state preferences measures suffer from indeterminacy- there are multiple ways of deriving them, but unlike psychometric measures, there is no natural way to get at the question of which is ‘more true’ to the subject’s real preferences.

Further, I argue that psychometric measures of overall desire satisfaction several advantages, perhaps the most notable of which is that they can get at *whether the subject has what they intrinsically desire* whereas traditional economic methods tend to assess achievement of goals that are instrumentally desired.

0.1.10 Chapter ten: conclusions, future research and selected topics

We briefly consider an objection that has been raised against many attempts to use happiness as a policy benchmark - *won't this lead to wireheading*. We argue, among other defences, that our approach is very robust to the wireheading objection, because it encourages the use of multiple indicators to feed into ethical debate, some of which (like eudaimonia) are very unlikely to be positively responsive to wireheading, if studied correctly.

We consider the argument sometimes made that *any* standard of evaluation other than unweighted cost-benefit analysis will lead to inefficiencies, for example, that valuing the happiness of the poor as much as the rich in decision making effectively amounts to an increase in the progressivity of the tax rate.

We contemplate a methodological question regarding welfare economics- should we expect welfare economics to give us a singular criterion for assessing policy, that can be followed in all cases (with maybe a few extraordinary exceptions)? Should we expect *any* such criteria to exist, whether coming from welfare economics or not, or is agency decision making best done through a series of “balancing tests” that draw attention to a number of factors, but do not quantise them against each other. We very briefly defend the latter view.

We urge an empirical character to future research in the philosophy of welfare economics. Experimental philosophy -surveying the folk about their philosophical beliefs- has previously been justified through a *positive* program (finding out more about what the folk believe) and through a *negative* program (destabilizing the very notion that there is any such thing as a clear folk position). In the philosophy of welfare economics, we propose a third rationale- a normative rationale, based on the right of the public to expect input into decision making procedures. Philosophers can facilitate this by exploring folk views on procedural and substantive justice in economic decisionmaking. According to this program of research, The folk deserve to have their views on certain philosophical questions known because they deserve to have them taken seriously in making political decisions. Though related to the positive program, the normative program is different insomuch as we are not trying to support philosophical truth claims, but rather support claims about what the views of the folk, in matter of fact, bind a state aiming to be democratically responsive to do. Welfare economics is tremendously important to decision making in many contemporary democracies, but its details are often unknown to the public. Welfare economists, in turn, have little idea what the public would make of their often-arcane processes were they made aware of them. Concern for democratic legitimacy alone should lead us to want a better sense of how the public feel about the specificities of processes like CBA.

We end with a consideration of various directions for future research. We have given some ideas for how the cardinality problem could be attacked, but further theoretical and experimental research is necessary. The process we have described is only given at a high level of abstraction- extra work would be necessary to transform it into a practical manual. Additional work would also be needed to take what we have written primarily with cost-benefit analysis in mind, and turn it into a process for other areas, like optimal tax, and the economic analysis of law.

1 The Aggregation problem

1.1 Introduction

For good or ill, much orthodox welfare economics exemplifies the characteristic features often attributed to contemporary economics by its critics. Formalism, a preference for rationalism over empiricism, individualism and simple models to generate provable results that do not require computational treatment. Perhaps the most damning charge that has been brought against welfare economics is that it covertly makes very strong assumptions about values (Little, 1950; de Van Graaff, 1957; Robinson, 1964; Chipman & Moore, 1978; Hands, 2014; Myrdal, 1953, 1970). For example Bromley (1990), Little (1950), Nath (1969), and Sen (1979b) have argued that the reliance on the concept of Pareto efficiency is not value neutral. Similar concerns have been raised about Kaldor-Hicks efficiency (Streeten, 1953) and the implicit assumption that the marginal utility of income is constant (Krutilla, 1981; Mishan, 1960; Medin et al., 2001; Hands, 2014).

It was this tendency that Machlup (1969) was inveighing against when he complained that he had no problem with free trade between the land of facts and the land of values- it was smuggling that was at issue.

Mainstream consumer theory, with a close relationship with welfare economics, particularly via utility theory, has also been the subject of criticism along these lines. For example, Bruni and Sugden (2007) argue that the ordinalist revolution involved an excision of psychology from economics and a corresponding focus on the revealed preferences framework, which Drakopoulos (1990) argues involves a value laden conflation of revealed preferences with welfare. Hands (2010, 2011) argues that although there was a shift away from mentalist psychology and towards behaviourism, this was never fully carried through, as revealed preferences were still associated with volution in order for them to retain a normative relevance and hence relevance to the ordinalist and New Welfare Economics (Hands, 2010).

Even within the revealed preferences approach, certain behaviours have been rarely analysed, arguably in a manner that is value laden. For example, there has been a recurrent neglect of relative income concerns, (Drakopoulos, 2008, 2010, 2011, 2012, 2016; Hudík, 2014) perhaps because an open or implicit value judgment that motives of “envy” should be neglected in welfare analysis has been made, either directly (Robertson, 1954; F. M. Fisher, 1956; Feldstein, 1998, 1999; Milanovic, 2007) or on the instrumentalist grounds that taking such considerations into account will have an unacceptably large negative impact on efficiency (Weinzierl, 2017), or because it presents analytic difficulties (Claveau, 2009; Drakopoulos, 2016) that are implicitly not considered worth the effort to address.

Many critiques have been made of welfare economics as a whole, and we cannot do justice to them all here nor do we need to, given our goals. Picking one central topic, the two so-called fundamental theorems of welfare economics show four limitations, any one of which is arguably fatal to their direct use in economic decision-making, viz:

1) In general, showing that something is *Pareto optimal* is of limited probative value in assessing policy because a Pareto optimal outcome can be far from optimal from the point of view of almost any social welfare function. One (selfish) individual receiving everything is Pareto optimal in many scenarios.

2) Lump sum non-distortionary transfers are usually not possible in the real world (Mirrlees (1969) in his Nobel lecture notes such and argues this may be at base an information problem). This means that distortionary policies are called for, but the fundamental theorems provide little guidance here.

3) We do not live in an undistorted market, and there are no guarantees that local movement towards an undistorted market will be an improvement, as in the general theory of the second best (Lipsey & Lancaster, 1956; Lipsey, 2007; Boadway, 2017), for an elegant discussion see Morton (n.d.)

4) Many of the things that people value are not well modelled as commodities (Lancaster, 1966), and many of the commodities people value are only valued instrumentally (Yew Kwan Ng personal communication 2023), with many non-commodity factors mediating whether their instrumental value is realized. Thus, modelling welfare with a focus on commodities may be mistaken.

Of course, there are many traditions in welfare economics, which is, after all, a field and not a theory. But we are concerned in this thesis with a fairly narrow sliver of welfare economics. This kind of welfare economics is "at the pointy end" of making policy decisions, not models of ideal behaviour under heroic assumptions. Perhaps its best example is cost-benefit analysis, but other examples include the analysis of optimal tax (Mirrlees, 1971; Boadway, 2012) as a guide for the policymaker, of intergenerational tradeoffs (Ponthière, 2003; Dasgupta, 2019) and the economic analysis of law (Mishan, 1967; R. A. Posner, 1977; E. A. Posner, 1996; Kaplow & Shavell, 2002, 2000).

We are concerned with the making of economic decisions for the common welfare by the government (and occasionally non-government organizations like charities), especially where these decisions benefit some, and harm others, at least in the sense of opportunity cost- and with the role welfare economics plays in advising policymakers regarding these decisions.

We deal with various approaches to this problem in the next chapter, but some possible options include: *Throwing one's hands up*- denying that economics should say anything about such cases *The Kaldor-Hicks criterion*- asking whether those winners of a change could, in theory, compensate those who lose out from the change and still be better off overall and *Traditional social welfare functions*- in theory, a ranking of all possible social states, in practice this may be reduced to a valuation on an additional marginal dollar for each income level on the basis of moral intuition or the preference of a decision maker. Later in this thesis, we will propose our own alternative to existing models of how to deal with interpersonal tradeoffs.

1.2 Three problems of welfare economics

Here are three core problems that have troubled welfare economists more or less since the beginning. Broadly these problems can be split in two, and one of them can be split in two again. The first half is the problem of providing a rule for choosing between alternatives, the second half is the problem of providing the information the rule needs to be applied to.

The problem of providing information to apply a policy selection tool to can be divided into many parts, but in this thesis, we focus on two parts in particular, *cardinality* and *interpersonal comparison*.

1.2.1 The problem of cardinality

We begin with the problem of *cardinality*. For an overview of this topic, (one with a definite agenda) see Andreas (2010). It may be relatively easy to work out whether one state of affairs benefits a person more than another state of affairs. This gives us *ordinality* in assessing benefits. However, ordinality alone is not enough to perform mathematical functions like averaging. For this, we need some metric of benefit to be defined up to a *cardinal* scale.

A measure is *ordinal* if it can tell us the ordering (ties are permissible) between some set of measurements (x, y, z…). A measure is *cardinal* if it can tell us not only the ordering of some set of measurements, but the *distance* between each. Consider, for example, two ways of assessing education. On one we code people on the basis of whether or not they have not completed high school, completed high school, started college and finished college. Each category signifies ‘more’ education than the last, but one couldn’t say on the basis of such a measure that the gap between A and B’s education was twice that as the gap between C and D’s education. On the other hand, a measure of *total years in education* would allow us to make such judgements (assuming the validity of measuring education in years!).

Within the category of cardinal measurements, there is a distinction between *ratio* and *interval* measurements. A ratio scale includes a defined zero point, whereas an interval scale needn’t- and need only define the intervals between variables. For an example of an interval and a ratio scale, consider a thermometer before and after the discovery of absolute zero.

Ethical theories on which the consequences for welfare matter (a much larger class than just consequentialist theories) have requirements for information at different strengths of measurement (Y.-K. Ng, 1984b; Blackorby & Donaldson, 1982; Blackorby et al., 2002). I call this the problem of the informational basis of social choice (Fleurbaey, 2003; A. Sen, 2011). Where population is invariant, utilitarianism requires interval information. Thus the data generated by achieving at least interval measurement has been important to utilitarians in welfare economics (Pigou, 1912, 1951). Other ethical views, and utilitarianism outside the fixed population case, require ratio information, or potentially even stronger information (Adler, 2012; Brown, 2007), as summarised in Table 1.

While in this thesis we will only discuss cardinality in detail (in chapter 6), since part of our object is to avoid ethical sectarianism and the exclusion of views that require stronger information, we would like definition up to a ratio scale. At the end of chapter 6 I will give a few notes on ways forward for constructing a ratio scale.

A common solution to the problem in traditional welfare economics is to use von Neumann Morgenstern utilities (von Neumann & Morgenstern, 1944). Whether this solution is adequate or falls to one of the various objections against it, is beyond the scope of this thesis. However, one of the most persuasive objections to this approach is that there are alternative, equally conceptually sturdy methods of quantifying utility from behaviour and these methods can and do give at least somewhat contradictory results- for example see Latty (2011) for a literature review of various methods producing varying parameter values. There is no obvious way to resolve which method is “right”, and no obvious sense in which one must be ‘”right” and critically, no clear sense of which quantification of utility from preferences reflects welfare where they contradict- as Andreoni & Sprenger's (2012) put it in the title of their work, “Risk preferences are not time preferences”.

The problem reoccurs outside the revealed preferences framework in self-evaluation of wellbeing. Suppose that a dozen people are asked to rate their well-being on a scale from 1 to 5. It may be reasonable to think that any given individual is likely having a better day if they put a 5 instead of a 4 but is it reasonable to take the average of the group, and in so doing implicitly assume that the gap in the amount of wellbeing between a 2 and a 3 is the same as the gap in wellbeing between a 4 and a 5?

Of course, the problem of the informational basis of social choice we have described also presupposes *interpersonal comparison*, insomuch as it involves comparing utility across multiple people, our next problem.

Table 1: Strength of welfare measurement, permissible transformations, and supported SWF

Measurement strength	***Permissible statistics***	***Permissible transformations***	***Supported SWF/evaluative procedure***	***Example***
Ordinal	*Median, percentile*	*Monotone increasing (order (<))*	*Pareto improvement, a 'Rawlsian' max-min SWF*	*A scale that goes "bad, average, good, excellent"*
Interval	*All of the above: plus mean and standard deviation*	*Positive linear (affine)*	*All of the above, utilitarianism granted the assumption of a fixed population*	*Measurements of temperature before the discovery of absolute zero*
Ratio	*All of the above plus geometric mean, harmonic mean, coefficient of variation and generalised entropy indexes*	*Multiplication*	*Utilitarianism with variable population. Prioritarianism but only with a CES value function*	*Weight, height*
Absolute		*None*	*Prioritarianism, sufficientarianism, generally all major social-welfare functions*	*Number of books on a shelf (zero point and definite increments)*

Table notes: table adapted and expanded from Andreas (2010, Table II).

1.2.2 The problem of interpersonal comparison

We might all be willing to say that A has greater well-being if one thing happens rather than another. We might even be willing to say that the difference between state A and state B is twice the difference between state B and state C. Many will insist that a much harder question is how we can establish that Janet's welfare is greater than Li's, or more modestly, that changes to Janet's welfare are greater than Li's. Although the issue is easier to explain than the problem of dinality, the philosophical issues go very deep indeed and touch on epistemic problems like the problem of other minds, although surprisingly most classical studies seem not to make the link. I speak here of what is often called the problem of interpersonal welfare comparison, (Robbins, 1938; A. K. Sen, 1970; Elster, 1993; Binmore, 2009), for reviews see Drakopoulos (1989) and Hammond (1991). There are both epistemic issues here (how could we know the internal experience of two people well enough to compare them) and ontological issues as well (is the wellbeing between persons even the sort of consistent quantity that *can* be compared)?

1.2.3 The problem of values in welfare economics

Even if we could quantify changes to welfare, however defined, there would still be problems about the right ways to integrate values into welfare economics. Why choose one particular welfare definition and not another? How should we value the welfare of persons relative to each other. Should we prefer to maximize mean welfare? Prioritize the welfare of the worst off? Prioritize the welfare of those who are seen as worthy in some sense? The problem is one of finding a way to navigate these issues, without simply resorting to a preferred ethical framework "just because" or alienating from your conclusions those who do not share your values. We call this *the problem of values in welfare economics* - for a discussion see Hands (2014) and Baujard (2014).

1.2.4 All three problems are part of a larger problem, the problem of interpersonal aggregation

These can all be seen as part of a larger problem- what results is the welfare economist to report when measures that make one person better off make other people worse off? How can we then talk about the welfare of society as a whole, when the interests of its constituent members run in different directions? The problem of interpersonal comparison makes it difficult to compare the harms and benefits done to one person to the harms and benefits done to another. The problem of cardinality means that even if we could compare harms and benefits done to different individuals, saying that one was larger than another, it would be unclear how we should add them up. The problem of the role of values in social sciences means that even if we could add up harms and benefits done to different people it would be unclear what function we should try to maximise. Should we try to maximize the sum or average? Or maximize a weighted sum or average based on some moral criteria? What about maximising some concave transformation of welfare, on the basis of the prioritarian view (Harsanyi, 1975; Parfit, 1995, 1997; A. K. Sen, 1979a) that increases to welfare matter more to the worse off.

How could we resolve these questions without taking positions on controversial matters of justice and fairness that would automatically make our conclusions suspect or irrelevant to many people with different moral views? Is there some way we could avoid a function to be maximised altogether?

One clarification on weighting is required. When we talk about what the welfare economist should report when different effects run in different directions, we do not necessarily mean that they must report a single aggregated figure- much less a univocal verdict. Rather our expectation is more modest than this- they must report *something* which throws light on the overall composition of harms and benefits. This might be a few different figures- the change in welfare, plus the change in inequality of welfare, for example- or even many different figures. However, except perhaps in cases with a very few involved parties, it cannot simply be a relisting of the disaggregated harms and benefits to each person especially if those harms and benefits are in merely monetary terms and thus incomparable between persons with different income levels.

To say that these three problems are aspects of this larger problem of aggregation is not to say that they are exhausted by the larger problem. Even without dealing with differences in interests between individuals these problems can arise. For example: the problem of cardinality haunts us if we want to make decisions under risk. The problem of values might arise even without a divergence between individual welfare in the context of paternalism- should welfare economics aim at what is good for people, or at what people want? What should we do when these clash? The problem of interpersonal comparison could arise without a conflict between what increases the welfare of one individual and what increases the welfare of another. For example- in the context of looking at two societies and trying to decide in which society people are better off for the purpose of policy selection etc. Nevertheless, these three problems do intersect in this larger context of clashing human needs.

Another problem that troubles welfare economics which we will consider: *where to source information about welfare and/or preferences?* This is not wholly disconnected to the questions we asked earlier. Clashes in interests and preferences- the aggregation problem- is related to this problem of how to learn about preferences because when we are contemplating changes which might make things better for everyone, we often need not concern ourselves overmuch with the details of what people want. When, however, we are considering policies that will hurt some and harm others, it becomes of vital importance to know what people want in detail and to know the degree to which they want them.

1.3 Conclusions for chapter one

A primary problem, perhaps *the* primary problem, of welfare economics is to aggregate harms and benefits to different individuals. We call this the interpersonal aggregation problem. This problem includes within its aegis many of the classical problems of welfare economics: The problem of interpersonal comparison, the problem of cardinality, and the problem of the role of values and ethical judgements in a science. In this thesis, we will be arguing that a particular approach, grounded in psychometrics, which in chapter 4 we will name Subjective-Wellbeing Effect Estimation (SWBEE), is uniquely placed to resolve these problems.

2 A typology of welfare economics

2.1 Introduction

If everyone's welfare rose and fell conjointly, welfare economics and probably everything else in this world would be far less controversial. However, different people's welfare rises and falls independently- or even in opposition to each other all the time in economics and questions of policy. This gives rise to a problem of interpersonal aggregation- a problem with both positive and normative components. One could say that the problem of interpersonal aggregation (as opposed to the problem of interpersonal *comparison* which we defined more narrowly earlier) is the problem of how to form a judgment about which policy out of a menu is better, based on its differing effects on different people. However, we will formulate it somewhat more cautiously- how can the welfare economist, qua welfare economist, say something of definite value as a contribution to selecting between these policies on the basis of the policies differing effects on people. The reason for this more cautious definition will become clear in chapter 7, where we consider the problem of values and the question of value neutrality in welfare economics, and ultimately argue that the economist should not aspire to *recommend* policies, but instead aim for the much more modest policy goal of saying something important regarding policy selection.

Answers to the question of how to deal with interpersonal aggregation give rise to a number of different approaches to welfare economics. In this chapter we illustrate many of these different approaches, showing their weaknesses for the purpose of motivating our alternative approach.

To that end, I will now present a typology of approaches to welfare economics categorized by their approach to interpersonal aggregation. This typology is not exhaustive, nor are the categories always exclusive. This typology is just *a* way of understanding responses to the problem, it is by no means the only way. It does not represent the way participants in these debates understood themselves. It is a philosophical-conceptual typology, not a historical typology, much less a social one. Although I hope to show problems with other approaches to motivate our approach I do not pretend to "disprove" or decisively defeat any alternative here.

2.2 The Paretian approach

Now we will consider the Paretian approach. The Paretian approach tries to deal with our problems by being minimally controversial. Judgements involving tradeoffs may be beyond us, but at least we can say that increasing the degree to which one person's preferences are met, without penalizing anyone else would be an improvement. In terms of a money metric, we might say that one person having more income is better than that person having less income, all other things being equal- so if at least one person has more income, and no one has less income, and nothing else is changed, (and, as is often tacitly assumed, no one has preferences over anyone else's income) this counts as an improvement.

The idea of the Pareto improvement is very well established in welfare economics- the so-called *two fundamental theorems of welfare economics* are both about Pareto optimality (c.f. Blaug (2007) for a statement and overview of their importance) and sometimes seen as the only needed “widely acceptable ethics” (see for example (Archibald, 1959; Bergson, 1938; Little, 1950; Mack, 1952; Mishan, 1957)).

However, I was not able to find in the literature a clear and contemporary, example of someone adopting what I call the Paretian approach- insisting that welfare economics is and must be grounded in the search for Pareto improvements- eschewing bolder ideas like Kaldor-Hicks optimality or social welfare functions. Nonetheless, given the power the concepts of Pareto improvement and optimality have in the field, it is important to consider Paretianism as a possible position, and one might say it is “in the water”. There are certainly authors acting as if this is true- see, for example, Thurow (1971) who considers the criterion of Pareto optimality important enough that he even tries to apply it to the question of the income distribution, at first glance a paradigm case of something beyond the Pareto criterion’s reach. Why? Presumably because he sees the Pareto criterion as holding profound ethical significance.

The primary limitation of the Paretian approach is obvious. It is limited because it only provides a partial ordering of different states of affairs (Kemp & Ng, 1976), indeed the vast majority of plausible changes will leave at least one person worse off.

The second limitation is that while this approach may be less controversial than others in some regards, it does not buy one entirely out of controversy. For example, it is far from uncontroversial that every person should be made better off, if possible, regardless of factors like their history. Some philosophers have argued that retribution- which is thought to usually necessitate a reduction in material conditions- can be an intrinsic good (Walen, 2015).

There are also plenty of egalitarians who would argue that a Pareto improvement can make things worse off if it takes the rich further away from the poor- any egalitarian who bites the bullet on the leveling down objection (see Temkin (2003) for an exposition of the debate), falls in this camp.

Blaug (1978) cited in Hands (2014) identifies an additional ethical assumption in the Paretian framework- that individuals know what is best for them. It seems reasonable to think that sometimes people want things that will make their lives worse- e.g., a heroin addict, but in identifying welfare with preference fulfillment, the Pareto criterion makes no allowance for this. We could of course remove this assumption, for example, by making a Paretianism defined purely over welfares, but then Paretianism loses it’s unanimity property that supposedly made it ethically non-controversial - people might object to changes intended to raise their own welfare!

But let’s suppose that giving people more money generally increases *their* welfare, and moreover, assume the truth of what we might call minimal welfarism:

> *Minimal welfarism:* The total goodness of a situation is always at least somewhat improved by an increase in someone's welfare with no decrease in anyone else's welfare.

Even so, there are grounds to think that Paretianism remains ethically controversial as an approach because welfare economics of this kind requires a numeraire. If the numeraire is money- if we make our judgements of better or worse directly on the basis of one person having more income or wealth and no one having less income or wealth- then the case that Pareto improvements can make things worse is very strong, because relative income effects (the effects of one person's income on the wellbeing of another) suggest that one person increasing in income can make others worse off, as pointed out by Reder (1952).

Alternatively, if we don't count monetary Pareto improvements as real Pareto improvements unless they are also Pareto improvements in terms of psychologically defined welfare, then Pareto improvements will become very rare indeed. For example, if some people get richer, and we leave even one person just as they are, it is quite likely that there will be a negative psychological effect on that person through relative income effects. For a discussion of this problem in relation to Pareto improvements, see Hansson (2004).

We might say that there are three possible varieties of Paretianism. Please note though that we intend this as a division of possible positions in logical space, rather than as a statement of how actual scholars are organized or view their own work. The varieties of Paretianism we could call *Strong, moderate and weak*:

2.2.1 Strong Paretianism

One way to cash out the Paretian view is to interpret it very literally. Strong Paretianism holds that all welfare economics can properly do is engage in a search for Pareto improvements. Such a view would be that any evaluation of whether a gain to one person is more important than a loss (or opportunity loss) to another person is a matter of value judgments that economists cannot properly make in their capacity as an economist, but recommending Pareto improvements is perfectly proper.

To which I would say "The one who says something can't be done shouldn't get in the way of the one who wants to do it". The strong Paretian owes us an account of why it is not possible or why it is not desirable for the welfare economist to appeal to any value judgements except those inherent in the idea of a Pareto improvement. It can't be that the Paretian has a general objection to values in science since, as we have already established, the Pareto criterion itself involves values. What is the basis for thinking that *these* value judgements are fine but other value judgements are not?

It could be a simple preference for only the (putatively) minimal values embedded in the Pareto criterion, but this value judgment is by no means obvious or self-justifying. The partisan of the Pareto criterion has no clear right to tell others proceeding on the basis of other values that they are wrong for advancing their investigations under the banner of those values. Perhaps they think the use of any other system of values will lead to some sort of disaster? But if so, let them present the argument.

It is worth noting that other applied sciences have not taken the view that value judgements must be constrained to the endorsement of Pareto improvements or other ‘thin’ moral judgements- consider conservation biology and public health:

Conservation biology: Though the matter is controversial, many feel conservation biology can and should deploy strong value judgements. Consider for example Barry & Oelschlaeger (1996) who do not discuss the Pareto criterion explicitly but appear to be speaking about value judgements that go well beyond it. Charpentier (2015) takes an unusual approach that is worth mentioning. The role of conservation biology, for Charpentier, is not so much in helping us *get* to the Pareto frontier, as in helping us understand the tradeoffs along that frontier, moreover, Charpentier thinks we can understand those tradeoffs partly in terms of biological-ecological concepts.

Public Health: Ng and Rueger (2011) discuss value judgements in the administration of public health noting:

> *The exercise of government powers for the health of its population raises ethical issues, such as public welfare, individual autonomy and freedom, privacy and confidentiality, just distribution of benefits and burdens, transparency, and public accountability. These ethical concerns sometimes conflict, pitting values against one another. How they should be balanced will vary on a case-by-case basis.*

Insomuch as the academic field of public health plays an advisory role to government role of public health, the academic field will necessarily be embroiled in value judgements. Harvard et al. (2020) make the argument for an area very closely related to our own- modelling in health economics.

Looking ahead to the rest of this thesis, the Paretian might respond with a *tu quoque* why am I here so intent on insisting that valuable judgements can play an honourable role in policy oriented social science, only later to outline a system that seems to aim to minimise controversial value judgements? My answer is that our aim is not to banish value judgements altogether, but to avoid value judgements which are sectarian in a democratic and pluralistic society. Partisans of the Pareto criterion sometimes claim that achievement for themselves, but even if this is true, it does not mean that no one else has found such a basis,

2.2.2 Moderate Paretianism

Moderate Paretianism suggests that we should prefer looking for Pareto improvements where we can, though there might be alternative approaches welfare economics can adopt should looking for Pareto improvements no longer be possible or fruitful.

While I concede that there is power in finding Pareto improvements where they exist, I think a reasonable critique might be developed of this position that it is potentially *ideological*. It counsels that we direct much of our attention to that which is harmless to existing interests. An overemphasis on Pareto improvements might prevent us from finding opportunities to do more good. It is possible, for example, that, from an even moderately utilitarian or egalitarian point of view, the lowest hanging fruit is egalitarian redistribution, which is the furthest from the Pareto ideal as can be.

Another example of what could be low hanging fruit (Ramsey, 1928; Dobb, 1940; Lange, 1936, 1937; Cowen, 2004; Yew-Kwang Ng, 1989; Cohen & Parfit, 1992; Cohen, 1992; Greene & Sullivan, 2015) but is rejected by the Paretian, might be reducing the discount rate on welfare- what is sometimes called the pure time discount rate. Such a change would likely reduce welfare modestly for some period but lead to large welfare gains over the future.

We might worry that the idea of unobjectionable improvement has taken up space that would have been better used in seeking improvements which are, from the point of view of many plausible social welfare functions, larger. Bromley (1990) argues a similar line that in treating efficiency as an objective, and using it as a decision-making criterion, but refusing to say anything about any other goods, the Pareto efficiency concept privileges efficiency at the expense of other criteria.

To illustrate the plausibility of this, suppose that someone in public health or conservation biology suggested something similar to the moderate Pareto criterion- that most of our energy should go into interventions that are offensive to no one’s interests. People would reject this view because there are important forms of good that can be done that harm some people and interest groups- regulations on cigarette advertising and limitations on habitat reduction to pick two relevant examples.

2.2.3 Weak Paretianism

Weak Paretianism would hold only that uncovering possible Pareto improvements, where they exist, is an especially important contribution of welfare economics, but may not be the only contribution. It is unclear that this is a theory of welfare economics per se, rather it is a much more modest statement.

As we have previously mentioned, is arguably not even true that discovering Pareto improvements is an important contribution. Assume for a moment a social-welfare function like prioritarianism or utilitarianism. If the numeraire by which an outcome is assessed as a Pareto improvement is money, relative income effects might mean that a putative Pareto improvement is a wash or negative.

But even if the numeraire is subjective well-being, we need to have caution- there could be long-term effects on subjective well-being that are not well captured in our calculus. For example, let's suppose I create a policy that will make everyone richer to such a degree that a first-order calculation suggests they will be better off in subjective welfare terms. However, it will make the rich, much, much richer than it will make the poor richer, and in a way which is widely perceived to be unfair. In the long run, it is plausible that such policies might undermine civil society etc. in a way that may cause more harm than good, but which will be hard to quantify in the context of an economic analysis of apparent Pareto improvements even if they are denominated in something that is thought to be a direct measure of welfare.

The point is that there are no "clean" Pareto improvements. Many operational definitions of a Pareto improvement are thin enough that, in practice, a Pareto improvement might prove to be a backwards step. A definition thick enough so as to rule this out in principle- at least in the eyes of many- e.g., "everyone will be better off or no worse off throughout all time because of this policy" goes beyond what can reasonably be established by social science. The concept of Pareto improvement in practice always involves compromises with reality despite its high-minded motivation.

Thus, even the seemingly innocuous statement that Pareto improvements, where they exist, should be identified through economics and then pursued by states turns out to be a more complex than it first appeared. Finally, we reiterate, even were weak Paretianism correct, it still wouldn't be a complete approach to dealing with the sorts of problems we are interested in welfare economics, and it admits as much.

2.2.4 Kaleckian inefficiency

Our final concern about Paretianism relates to all forms of Paretianism. *The assessment of whether or not something is really a Pareto improvement requires a deeply limiting consideration of all future possible improvements, as well as the strategic situation the players find themselves in.* This is well illustrated by Latty's (2014) concept of *Kaleckian inefficiency.*

Intuitively In a situation of Kaleckian inefficiency, it is not in the interests of one party to a negotiation to accept an efficiency-increasing measure, because doing so reduces their control over some process. If they lose control over that process, the process may be used to their disadvantage and the advantage of another player in the future. The concept builds upon Kalecki (1943). Latty (2014, p. 8) defines the concept as follows:

> *A policy is Kaleckian inefficient if:*
>
> 1. *There exists some alternative policy which is potentially Pareto (Kaldor-Hicks) or more specifically Little superior (i.e. we could simultaneously raise output and reduce inequality)*
> 2. *Policy choice affects bargaining power or distribution, so that lump sum or other redistributions that could make a policy strictly Pareto optimal is politically impossible.*

3. *The potentially Pareto optimal alternative is not viable because it is an inferior choice for some important equality averse policy actor(s).*

Latty's main example is taxation- a mode of taxation without deadweight loss might seem appealing to all involved but suppose that the rich have more control over the *structure* of taxation, but the poor have more power over the *rate* of taxation. Latty argues that under these conditions it is in the interests of the rich to design a taxation system riddled with inefficiency and deadweight loss. Why would we think that power over taxation is divided so? Plausibly, voters (and thus the more numerous and less wealthy citizens) have power over the larger details of politics, through the exercise of their franchise. This includes the overall scale of taxation. However, the rich have power over the finer details, through their connections and access to lobbyists etc. Thus, it might be in the interests of the rich to avoid implementing efficient taxation, because if efficient taxation were introduced, one of the most important barriers to raising taxes much higher would be removed.

In Kalecki (1943) the main example was full employment rather than taxation- in the Keynesian model, full employment is theoretically to the advantage of all involved, however as Kalecki notes it may be unacceptable to the employing class for political reasons- because it empowers politically and industrially the working class. The problem of Kaleckian inefficiency raises an important problem for Paretianism- what is and is not really a Pareto improvement can be very hard to judge, and require a comprehensive awareness not just of money, not even just of material factors, not even just of intangibles like prestige, but also of power and institutions. The ubiquity of political factors likely limits how many changes are to the benefit of all involved in a comprehensive sense.

More generally, every Pareto improvement sets a new baseline, below which future Pareto improvements may not lower recipients, or else cease to be Pareto improvements. If a rule "only pursue Pareto improvements" is accepted, this gives apparently harmless Pareto improvements deep strategic significance, since once given, something cannot be revoked (except in exchange for something desired even more). In future research, I intend to discuss how limited the idea of "Pareto improvement" becomes in real world situations where strategizing over future changes and other forms of political calculation are a factor.

2.3. Realism

Now we will consider the realist approach to welfare economics. Realism generally is a convoluted word in philosophy, and its specific meanings become even more complex. By realism here though, I mean a family of views which hold that, for some definition of welfare which the realist in question holds to be the right one, there is a fact of the matter about the interpersonal aggregation of levels of welfare that can be discovered empirically, and this should form the 'core' of our approach to evaluating policy in some sense. Different species of realism will differ on the basis of what empirical truths they believe enable us to uncover and compare how policies affect the welfare of different people.

2.2.5 Psychological realism (or psychometric realism)

According to psychological realism, the approach that most informs this thesis, the following hold true:

A) It is sometimes legitimate to treat subjective well-being as welfare, at least for some important purposes. Alternatively, it is okay to treat subjective-wellbeing as a proxy for welfare- and as an indicator, we can use to assess, at least in part, how well people's lives are going and B) Subjective wellbeing is a real psychological variable that is best measured with the traditional tools of psychometrics.

Although the movement is nascent, and there are debates about the details, there is an emerging body of scholarship by authors who favour this approach - see, for example, Ferrer-i-Carbonell (2013) *Happiness Economics* and MacKerron (2012) *Happiness economics at 35,000 feet*. A good example of a very direct advocate of this approach is (Bronsteen et al., 2013) who suggest that wellbeing analysis (WBA) could take the place of cost-benefit analysis.

One of the most important challenges to psychological realism is concern over the role of values in science. For example, there is disagreement about the extent to which values are implicated in the study of subjective wellbeing because it involves what appears to be *mixed terms* (Alexandrova, 2012)- terms that involve both positive description and evaluation. Implicit in this debate are further concerns about the degree to which using values in science, openly or otherwise, prevents us from taking a realist perspective on that science. This question goes beyond our aims in this thesis, as we largely take the science of wellbeing as a given. We might identify three possible views on the matter:

1. Beyond the value judgment that something is worth studying, there are no value judgments involved in the study of subjective-wellbeing, for a given subjective-wellbeing concept (although of course, the creation of concepts of subjective-wellbeing reflects values)
2. There are value judgements involved in defining the detailed and operationalizing specific concepts of welfare, but no value judgements need to be involved once we have specified these welfare concepts and head out to the field to administer them.
3. There are value judgements involved at every stage of the study of welfare, even in comparing and measuring its levels between persons, once we have a specific welfare concept.

Alexandrova (2016) accepts at least 2 and possibly 3, whereas Nagel (1961) defends one (as a general principle about all mixed claims and not solely in relation to welfare). The psychological realist will have to defend 1, or they will have to accept 2 or 3 but argue that the kinds of value judgments involved do not prevent us from being realists about the science of subjective wellbeing. One might offer a variety of defences of this position- for example, it is often thought that to describe a condition as a disease requires a value judgment - see Reiss and Ankeny's (2022) overview of the philosophy of medicine, but few would challenge or refuse to accept the scientific credentials of epidemiology. One might bite the bullet and defend realism about mixed claims, arguing that science can reveal truths with a moral component and is thus, at least at times and in part, a form of moral inquiry.

We flag the issue here for completeness as a potential threat to psychological realism, but we will not discuss it further in the thesis, except for brief allusions in chapter 3, as we are not, for the most part, concerned with methodological critique or examination of the psychology of subjective wellbeing, but rather expounding upon what I take to be its philosophical implications if accepted.

2.2.6 Representational theory of measurement realism

The representational theory of measurement holds that:

> Simply put, a measurement scale is a many-to-one mapping—a homomorphism—from an empirical to a numerical relational structure, (Tal, 2020)

In the context of economics, the representational theory of measurement has been combined with the revealed preferences methodology to map preferences onto choices that agents actually take. For an agent to increase in welfare then is simply for that agent to transition into a situation the agent would choose over their previous situation.

The first problem to note here is that any approach tied to revealed preferences will presumably have to embrace a preference theory of welfare. Any objections to the preference theory of welfare will thus attach themselves to this form of realism (unless it adopts something like the indicator or anti-theory of welfare we discuss later in this section on realism).

But even leaving this aside, articulating a framework for interpersonal aggregation using this approach has not proven easy. A procedure for creating a mathematical description of an agent and its preferences may only take us so far.

2.2.6.1 Ordinalism

The ordinalist wants to measure utility on an ordinal scale. This makes interpersonal aggregation especially difficult due to reasons associated with Arrow’s theorem (Arrow, 1950, 1951b) which shows that any ordinal social welfare function must violate at least one of several highly desirable conditions. Roberts (1980b, 1980a) shows that the situation is still difficult even if interpersonal comparison of ordinal preferences is allowed, and Parks (1976) and Ng (1982) give reason to think that both interpersonal comparison and cardinality are necessary to resolve the problems of Arrows theorem.

Of course, the impossibility theorem or its extensions do not mean that there can’t be rules which give sensible results in most cases. McKenna (2008, p. 33) in an article in the popular scientific press (*New Scientist*) quotes Arrow as saying “Most systems are not going to work badly all of the time. All I proved is that all can work badly at times.". A famous example in which group ordinal decision-making works better than expected is single-peaked preferences (Black, 1948). There have also been various attempts to grapple with Arrow’s conditions and soften the result (Morreau, 2019).

Probably a more fundamental difficulty than even Arrow’s theorem is this- any reasonable approach to policy must account for the intensity of preferences, but by definition, ordinal approaches cannot do this. Sam, Jessie and Tash are discussing what to eat for dinner. On a whim, Jessie and Tash would like to try lizard meat, and it just so happens that Sam owns a lizard. Sam loves her lizard dearly and will be devastated if it is eaten. It seems that an ordinal interpersonal aggregation rule is going to state that the lizard should be for dinner. We could of course make the case even more grisly should we want to, yet no amount of intensity on Sam’s part will flip the result- majority rules due to a restriction on cardinal information. Even if this is fair as a democratic decision-making method, it is not correct as a way of working out the aggregate well-being effects of policies. It is important to be able to work these things out separately, if possible, for example because people might want to make their voting choices on the basis of knowledge of how different policies will affect total well-being.

2.2.6.2 Cardinalism

The cardinalist approach, as I talk about it here, is the strategy of trying to read cardinal strength preferences into behaviour. Without going into the technical details of how some of these approaches work, we will consider them in overview.

Possibly the most famous way of deriving cardinal utilities from behaviour is considering an agent's preferences over different hypothetical lotteries or other risky propositions. This method, sometimes called expected utility, was first proposed by von-Neumann and Morgenstern (1944).

When it was first proposed, it came at an odd historical juncture (Andreas, 2010). Robbins (1932) had only fairly recently launched his canonical attack on Cardinalism and interpersonal comparison, casting doubt on non-ordinalist approaches in welfare economics.

Von-Neumann and Morgenstern’s demonstration proof of the possibility of deriving cardinal preferences from behaviour under risk was thus something of an embarrassment, and there were less than convincing attempts (Baumol, 1958) to dismiss some of its practical implications, as discussed by Andreas (2010).

In practice, the vNM approach can be applied in the lab, based on hypothetical gambles, or based on pricing in insurance markets among other options. There is no theoretical guarantee these will give the same results, and some empirical evidence that they diverge. Latty (2011) shows this through an overview of different derived values of η. As far as I can tell, there is no obvious means of picking between different results, should there be divergence.

It is also possible to generate cardinal utilities from the tradeoffs between consumption and savings across varying interest rates and income levels, given the assumption that agents wish to maximise utility from consumption across the lifespan (Besley & Meghir, 1998; Latty, 2011).

A prominent approach in the literature (Cowell & Gardner, 1999; Evans, 2004, 2005; Evans et al., 2005; Evans & Sezer, 2004; Lopez, 2008; Percoco, 2008) is to derive estimates of the elasticity of the marginal utility of income on the assumption that taxation imposes an "equal absolute sacrifice" on citizens. However, there are many concerns one might have about such a model- firstly it seems to assume a rather benign model of state, wherein we can read off the nature of welfare from state policy, one hardly needs to say why such an approach would be controversial.

Latty (2011 pp 35-36) sums up the literature by pointing out that the equal sacrifice model:

1. Is incompatible with huge variation in the tax schedule across time and place.
2. Misses that there are political and ethical reasons why governments might depart from equal sacrifice, both towards greater and lesser redistribution
3. Gives no reason to think that governments are better at accessing facts about what constitutes an 'equal' sacrifice than anyone else.

If the state is benign, it is far from clear that it will aim at equal absolute sacrifice. The equal absolute sacrifice model of justice in taxation has a venerable history perhaps originating with John Stuart Mill (1848/1909) however it has never been unchallenged. For example, if you think that, relative to a state of nature, the wealthy gain more utility from the maintenance of government and the social and economic order than the poor, you might think this gives reason to hold that the rich should make a greater sacrifice in utility terms than the poor (Murphy & Nagel, 2004). Despite John Stuart Mill's own utilitarian views on ethics, there is little reason, at least prima facie, for the utilitarian to embrace the equal sacrifice model. If more good can be derived from an extra unit of sacrifice from the rich than from the poor, the utilitarian has, at least prima facie, a reason to support greater sacrifices from the rich than the poor.

These are not the only approaches, for example there is the consumer demand for a preference independent good model (Fisher, Fellner, and Frisch method), see Fisher (1927), Fellner (1967) and Frisch (1932, 1959)). Another approach calculates η on the basis of the income elasticity of the value of a statistical life (Eeckhoudt & Hammitt, 2001; Costa & Kahn, 2004; Kaplow, 2005; Somanathan, 2006) and there are other methods besides.

Regardless of the method by which they are obtained, cardinal preferences "thicken" the data in a way which avoids Arrow's theorem and allows us to consider the intensity of wants and needs, but it does not avoid all troubles.

The first major problem that remains with this type of approach is that human preferences can be inconsistent (McFadden et al., 2000; Lucas et al., 2011) and behaviour often reflects misconceptions about how things are as much as true preferences. If we can't know what an agent prefers from their behaviour, a fortiori there can be no "safe" interpersonal aggregation of preference-based welfare derived from behaviour.

The second main problem is directly related to interpersonal comparison. What does it mean to "compare" these kinds of purely behaviourally defined preferences? I would choose a new wallet over a new shirt and would choose either over dying. You would choose a new shirt over a new wallet, and would also choose either over dying, good and well, yet if your choosing is identifiable with your preference, what does it mean that your preference for a wallet "exceeds" mine? Or to bring in strictly cardinal information, what does it mean to say that the gap between two of the outcomes for you is twice as large as the gap between the same two outcomes for me? If your preference for a wallet is a mathematical construction from your behaviour, and my preference is a mathematical construction from my behaviour what does comparing these constructs really mean?

There are approaches that have been developed to allow comparison of preference utility. These include the zero-one proposal (Hausman, 1995; Weintraub, 1996), and those based on an assumption of equal intensity of preference for some common commodity (Friedman & Savage, 1948; Sidgwick, 1874; Vickrey, 1960).

The zero-one proposal suggests that we treat each agent's worst-case scenario and each agent's best-case scenario as equal in their disagreeability and agreeability respectively. It is not my intention in this thesis to prove that no such approach could work, but even were the zero one proposal to work, there is reason to doubt that the view would exonerate the representational approach, for the resulting system still has the air of a convention, rather than any substantive comparison of persons, their lives, and experiences. It is instructive to consider Hausman (1995), and his contention that the preference satisfaction theory of welfare will make the zero one proposal true, but in a way which is both trivial and somewhat disturbing.

A final problem I would draw attention to is less developed in the literature, but we might call it the "embarrassment of riches" problem. What should we do if different plausible techniques for constructing cardinal preferences from behaviour give different results? How should we pick which figure is the true figure for the value of η and more ontologically, why should we assume that there even is such a thing as "a" true value of η with respect to welfare? It would be one thing if we were trying to estimate an underlying psychological quantity- with each behaviourally inferred number a different estimate- but η *is a purely behavioural construction* hence the existence of multiple contradictory constructions leaves us with the question, if one should guide ethical policymaking, which is it?

For example Latty (2011) compares values of η obtained using a variety of methods - but this can only work on the assumption that we already have good reason to posit a single underlying value which represents "the" diminishing marginal value of income. In truth, different values of η might be implicit in different behavioural contexts, *and we must remember this is a theory about mathematical representations of behaviour, not what may or may not underlie behaviour*. Barron et al. (1984) argue that both behaviourally and conceptually there is a difference between utility derived from preferences under risk and the intensity of true preferences, which they call "value". Andreoni and Sprenger (2012) show a difference between time and risk preferences. The choice of which form or aspect of behaviour to read preference intensity of is thus a question of ethical orientation or pragmatic choice, not of underlying psychology because the representational theory is not a claim about the psychological mechanisms underlying behaviour- only duelling mathematical constructions over different aspects of behaviour.

It might be thought that the psychometric approach that we will later defend faces the exact same challenge- are there not many different tests and theories for various psychological constructs? However, since the decline of the popularity of operationalism (Chang, 2009), most psychometricians have thought that any given measure of a construct- e.g., a test- is an attempt to get at a ground truth about that construct. Different tests can be more or less accurate as construct measures, and as we discuss in chapter 3, there are various strategies of assessing fidelity. It is not clear that the revealed preferences approach, based on the assignment of numbers to behaviour and not an attempt to "get at" an underlying hidden variable, has a way of determining which cardinalisation of preferences based on behaviour corresponds to the real underlying intensity of preferences. Rather different measures seem to have their own- domain limited- legitimacy, and function, and it is far from clear that the question of which one is "right" is meaningful.

Finally though, I wish to flag, but not investigate further in detail in this thesis, the possibility of a hybrid approach. This hybrid approach defines utilities not just on the basis of behaviours, but also on the basis of the neuronal reward system and its relationship with behaviour. Such an approach, by seeking underlying variables behind observed behaviour commonly ascribed with cardinal utility functions, would combine a traditional cardinalist approach, and what we earlier called a psychological approach.

Stauffer, Armin and Schultz (2014) found that the dopaminergic system reflected the revealed cardinal utilities of monkeys making decisions under risk: "Critically, the dopamine prediction error responses at the time of reward itself reflected the nonlinear utility functions measured at the time of choices" (2014, p. 2491). Hybrid approaches that bridge *psychology, biology* and *formal constructions on behaviour* seem like a promising direction for future research- one that could solve, among other problems, the embarrassment of riches issue. A Biorvix preprint by Matsumori et al. (2021) focusing on interpersonal comparison further develops a similar and exciting line of research, using a much larger sample of human subjects they deployed: "A method based on brain signals that correlated with changes in expected utility weighted by subjective probability. The signal was larger for participants with lower household incomes than for those with higher household incomes and their ratio coincided with the estimates by "impartial spectators"". In other words, they found that estimates by third parties of comparative utility changes between persons corresponded to the difference in magnitude of the activation of brain regions associated with reward.[1]

By coordinating multiple forms of behavioural utility, questionnaires, behavioural cues etc. it may be possible to bridge the gap between the psychometric and revealed preferences approaches to utility, ultimately treating utility as a psychometric construct, to be investigated by the standard psychometric methods, in conjunction with revealed preferences. Such an approach would not be a formal construction on behaviour, but an attempt to measure its wellsprings.

2.2.7 The material school as a form of realism

Cooter and Rappoport (1984) argue that pre-ordinalist welfare economics had its own approach to the problem of interpersonal comparison based on "material welfare" - in essence, the measurement of physical health *especially as it related to productive capacity*:

> *Essentially, goods were seen as having utility if they contributed to a person's physical well-being, which was conceived as nearly equivalent to productive capacity. Physical well-being is objective, like the condition of a person's health, not subjective, like a person's enjoyment of a good meal. Furthermore, productive capacity of people is closely related to economic efficiency, which is also objective. (Cooter & Rappoport, 1984, p. 509)*

The approach described is practically oriented and intended to inform public policy. The focus was on the measurement of the material well-being of different classes, nations, regions etc. rather than an interpersonal comparison of individuals:

> *Comparisons between two named individuals were declared impossible or infeasible. However, this was not considered to be of particular importance for policy. (Cooter & Rappoport, 1984, p. 518)*

But inasmuch as health, suitably conceptualised, could be an interpersonally measurable and real quantity which could guide policy (e.g., expected remaining QUALYs), this approach is a species of realism as we are defining it here.

The question of the degree to which Cooter & Rappoport's (1984) characterisation of the history of welfare economics, and posit of a "material school" is accurate need not concern us here. The more interesting question for our purposes is *to what degree would such an approach be feasible, especially in today's intellectual, moral and political context*.

[1] The proceeding paragraph to this footnote was quoted on my Substack, Philosophy Bear although ironically the original post no longer seems to be indexed. For a discussion of the disgraceful conduct of the University of Sydney in relation to this, see: https://philosophybear.substack.com/p/the-assessment-of-my-phd-thesis-by

It is widely understood that there are very significant differences in health outcomes between classes, and other groupings of special policy concern such as race and gender (Veenstra, 2011). Thus, there remains scope to continue investigations within the tradition of the material school albeit with a different emphasis- looking at how different groupings economic and social status affects their wellbeing with an eye to correcting health injustices.

A natural extension of the approach, and one that would bring it closer to the approach of psychological realism I champion in this thesis, would be to consider the epidemiology of mental as well as physical diseases. Kristoffersen (2015a) explores the links between measures of mental health and measures of subjective-wellbeing for the purpose of establishing the cardinality of subjective-wellbeing. Thus she might be said to be doing something in the conceptual vicinity of the material approach- although her ultimate purpose is to support a more psychological approach. There are also numerous studies exploring links between the epidemiology of mental health problems, and economic factors such as inequality and poverty (Murali & Oyebode, 2004; Wilkinson, 1996; Wilkinson & Pickett, 2017b, 2017a; Buttrick & Oishi, 2017; Ribeiro et al., 2017; Case & Deaton, 2020).

Perhaps then the material school approach is ripe for revival, although likely as a partial approach, or adjunct to other methodologies. However insomuch as the approach is focused on increasing the productivity of people conceived of as workers, it would be morally imperative to modify the approach to include those who cannot, and will never again, work- for example elderly retirees and some disabled people. The welfare of such people still matters, but it is hard to see how the material welfare approach as described by Cooter and Rappoport (1984) would appropriately value such welfare.

2.2.8 Question mark realism and evidential realism

There's another kind of realist view that I want to flag here. The view is somewhat different to the others, although once grasped the idea is simple enough. In combination with a psychological approach, it is close to what I myself believe, although it will not be the topic of the thesis.

Earlier I defined realism as a class of views that holds that, for some concept of welfare which the realist holds to be the right one, there is a fact of the matter about the interpersonal comparison of welfare that can be discovered empirically. Suppose that you *believed* in welfare, but you were sceptical of the human capacity to *define it* or capture it in any terms that weren't vacuous. Or suppose that you believed it was definable in principle, but that it wasn't wise to get into the business of defining it in the context of welfare economics, or that its final definition had not yet been found. In other words, for the "some concept of welfare" part of the definition above, you pleaded no contest.

Nonetheless, you held that we could make a reasonable guess at welfare *levels* without defining welfare. Suppose you thought, for example, that willingness to pay- perhaps appropriately adjusted for baseline income- was such an estimator.

This view seems to me, in spirit at least, to be closest to realism of all the taxonomic categories we consider, and hence I have classed it as one. Welfare is held to be a real, interpersonally comparable thing which we can obtain empirical evidence about, although precisely what it is left open. This is by no means a purely hypothetical view. Adler and Posner (1999) and Hausman (2011, 2012b) have both advocated this view, or at least something very close to it. For comparison, see also discussion of interpersonal comparison as “undecidable non-normative” as discussed by Klappholz (1964, p. 104). The ideas in this thesis, are compatible with this approach, but do not depend on it.

2.2.9 Concluding thoughts on realism

In our discussion thus far, we have focused on issues in the interpersonal comparison of wellbeing and in the quantification of wellbeing (i.e., in relation to the issue of cardinality). However, even if the realist can resolve these problems, there remains another difficulty related to the problem of interpersonal aggregation- that is the problem of how to aggregate different changes in interpersonal wellbeing for the purposes of assessing policy. At first glance, it might seem that the natural strategy is to sum together well beings, however, defined and calculated, but this is simply to assume the truth of something like utilitarianism- if not as the truth about all ethics, then at least as the truth about how to aggregate that portion of the good represented by human welfare. Utilitarianism is tremendously controversial, and it is unlikely any approach baldly asserting it could win widespread assent. The problem might seem irresolvable without adopting strong and controversial ethical premises, most likely a social welfare function of some sort. We will return to this topic of aggregation in light of values disagreement in Chapter 7, where we will argue there is a way out- viz treating the information gathered not as a direct judgment on which policy is optimal, but as a package of information intended to aid and inform decision makers and the public.

2.3 Stipulationism

Now we’ll consider a view I name Stipulationism. By a “stipulationist” I mean someone who takes some function on goods, services, income or consumption and declares that to be welfare. The solution to the interpersonal aggregation problem here is to pick an input for our function which can be calculated more easily than traditional candidates for what welfare is, like happiness or satisfaction.

History’s first consequentialists, the Mohists (Fraser 2023) had something of a stipulationist flavour to their work. The Mohists most important work is the *Mòzǐ* (Mòzǐ, 5th–3rd centuries BC/2020). They articulated a form of consequentialism according to which the fundamental goods were peace, the material essentials of life, and population increase (Fraser, 2023). Admittedly, the Mohists don’t seem to have described methods for quantifying these things, let alone quantifying them all into a single function- at least not in the texts that survive to our day, as many Mohist texts were lost.

Archibald (1959 pp 317) gives a classic and influential statement of his own version of the idea:

> *"The sensible procedure in welfare economics appears quite simply to be this: we take, as an interesting criterion, the choice-system of the individuals, and ask how different arrangements alter the available choices"*

A good example of a contemporary approach which would count as Stipulationism is announcing, not on the basis of an empirical investigation but on the basis of your own broad sense of what is fair, or what you take to be a common set of values or norms, or, as Archibald says, the choice system of some individuals, that you will be assuming that economic welfare is equal to some function on income. In general, there is a tradition of economists arguing that interpersonal comparison is just a direct intuition about the goodness of transferring resources e.g. (Boulding, 1952; Myrdal, 1953; Reder, 1958)

Creedy (2007) effectively argues for ethical Stipulationism in a negative way- he argues against an approach to welfare economics which implicitly accepts utilitarianism, and then conflates empirical measures of η (eta) with the level of inequality aversion ε (epsilon) in a SWF or simply takes the social discount rate from individual behaviour. Instead, argues Creedy, economists must recognise that the degree of inequality aversion or time preference for us in policy assessment are based on value judgements. However, Creedy (2007) doesn't think that the solution (or at least not the primary solution) is economists stipulating their own values, but rather considering what the implications of a variety of different possible stipulations would be. Creedy thinks the stipulations should come instead from the decision-makers, or perhaps through the public via way of leaky bucket experiments which Creedy himself has carried out in the past (Amiel et al., 1999).

The first problem with Stipulationism is that it is unclear why anyone would care about some arbitrarily chosen transformation between income, or any other material goods, and "goodness". A superb ethical argument would be needed to motivate care, and even superb ethical arguments have dissenters so controversy looms.

The second problem is that, as we will argue in chapter 7, there is a difference between tracking some variable that is widely agreed to be of ethical interest, versus declaring that a specific social welfare function is correct. One tends far more than the other to cross the "proper" boundaries of science in a democratic society by relying on much more controversial value judgments. One might argue in response that the value function could come from outside social science- e.g. from a political decisionmaker.

However, it may seem improper to act as if democratic decision-making should be reduced to the operation of a value function. To put it a little differently, it could be argued that acting as if there were a single social planner with a social welfare function ignores that many polities are meant to be democracies, or at least not dictatorships (Nyborg & Spangen, 1997). Democracy should not be something that happens only in elections. Admittedly, it may be possible to deal with this objection to some degree by using multiple plausible social welfare functions and exploring the results they generate (Creedy, 2007) but social welfare functions can differ from each other in numerous different ways, so creating a fair or representative" sample of "plausible" social welfare functions might be difficult, and perhaps most critically ordinary people's values when it comes to distributive justice probably aren't well captured by any social welfare function (Colander, 2018). Consider, for example, rights-based constraints which are not well captured in a SWF framework- whether they be rights to redistribution or rights against it- both of which some people believe in.

Perhaps the most critical objection is that it is far from clear that non-specialists have the tools to assess the meanings of concepts involved- for example variables such as η (eta) or the inequality aversion parameter ε (epsilon). One can present as many models based on different SWFs defined over income as one likes, but they likely will not have as much meaning to the public as our preferred alternative, models of the effect on various sorts of subjective wellbeing by policy. We critique attempts to derive stipulated SWFs in a 'public friendly way' through the use of leaky bucket experiments in chapter 10.

2.4 Divertism and proxyism

Now we come to divertism, closely related but conceptually distinct from stipulationism. The divertist argues that we should study something else other than welfare in welfare or normative economics. For example, aggregate willingness to pay. The thing in question should be *relevant* to welfare- with more of the thing in question (likely) leading to greater welfare. However, this approach is unlike psychological stipulationism in that there is no attempt to equate the tracked variable with welfare. This is a fine distinction, and it is likely that many approaches are ambiguous between divertism and stipulationism. The solution this approach offers to the problem of interpersonal aggregation is simple and mirrors that of the stipulationist- pick a variable that can be quantitatively compared.

There are at least three lines on which we could defend divertism. The first is epistemic- we cannot know about welfare itself, so this other variable, that we have theoretical reasons to think will be related to welfare, will have to do as an object of study. Pigou, though not a divertist, argued that within economics "the range of enquiry becomes restricted to that part of social welfare that can be brought directly or indirectly into relation with the measuring rod of money" (Pigou, 1920/1932).

The second is ethical. Maybe you think that we obviously care about wealth because it's linked to welfare, it's inappropriately paternalistic for the government to attempt to increase welfare directly. Does not that venerable slaveowner's manifesto declare the right to "life, liberty and the *pursuit***"** of happiness? If we don't want the government in the bedroom, why do we want it on the psychologist's couch etc. A good example is commentary by a senior Australian politician, Josh Frydenberg, who gave a blithe dismissal of the idea that the government should measure and try to maximize happiness "fresh from his ashram deep in the Himalayas, barefoot, robes flowing, incense burning, beads in one hand, well-being budget in the other". (Liuzzo, 2020).

The final reason is an argument from democracy, by way of the second fundamental theorem of welfare economics. If we can just expand the pie, democracy can be used later, through the tax and transfer system, to achieve any system of distribution desired.

A number of special arguments have been made in favour of the Kaldor-Hicks compensation principle which does not apply to other forms of Divertism. For example, an argument has been advanced (Hicks, 1941; Polinsky, 1972; Zerbe & Scott, 2014) that if the Kaldor-Hicks compensation principle is followed consistently in every case, while each individual choice might make this or that individual worse off, it will tend to increase the welfare of everyone in expectation. While individual policies that are Kaldor-Hicks optimal may not be Pareto-improvements, overall, the choice to adopt a Kaldor-Hicks approach *will* be a Pareto-improvement, or at least near enough in some sense.

The problem with this reply, as noted by Little (2002a) and Streeten (1953, 1981) is that it is not clear why we should think that the benefits of Kaldor-Hicks improvements are distributed randomly- with everyone just as likely to be a winner as a loser. If, for example, Kaldor-Hicks improvements systematically favour the rich (as may be likely because their willingness to pay is higher) application of the Kaldor-Hicks principle may increase income inequality. The intellectual historian Cook (2022) argues that it has done this *in fact* as part of a global movement towards neoliberalism where:

> Kaldor-Hicks efficiency emerged as a central pillar of a new, interventionist, wealth-maximizing and market-based form of depoliticized technocratic governance that not only marginalizes distributive concerns but actively exacerbates the problem of global inequality.

Further, the problem becomes even worse if some degree of selectivity is exercised in picking *which* Kaldor-Hicks improvements are made. In particular, if the wealthy have a veto (or even probabilistic veto) over which Kaldor-Hicks improvements are carried out for political economy reasons, then even if the total universe of Kaldor-Hicks improvements favours neither the rich nor poor, the Kaldor-Hicks principle can lead to not just relative, but even absolute immiseration if the rich are in the driver's seat, as the rich systematically select proposals that make them better off at the expense of the poor. Perhaps even a relatively modest bias towards the rich in the generation or fulfillment of projects for Kaldor-Hicks improvements, whether that bias was political and/or economic in its basis, could compound over time, as greater inequality further strengthened the bias iteratively.

In the general case, outside the Kaldor-Hicks case, the divertist faces a similar worry to the stipulationist: motivating their particular alternative of concern to welfare.

One notable case is modified measures somewhat similar to GDP. For example, Nordhaus and Tobin (1972) initiated a tradition of altering GDP to try and better capture "true" welfare with their measure, the MEW (Measure of Economic Welfare). A more recent approach is the Genuine Progress Indicator (GPI) (Hamilton, 1998; Kubiszewski et al., 2013). However, such sophisticated approaches, while valuable, face a conceptual gap- as we add bells and whistles to our measures of economic welfare, the question of what it is, really, we are trying to capture becomes more urgent. Something like "true consumption of resources for our own benefit" perhaps? But philosophically this seems like a rather obscure idea laden with conceptual, empirical, ethical and accounting subtleties and every ad hoc adjustment we make to exclude things (advertising costs, defence costs, perhaps environmental degradation) raises additional methodological questions. The existence of a way to resolve these questions *without referring to some deeper concept of welfare and thus at least partially undermining the Divertist project* is unclear, at least to me, though I am uncertain either way.

2.5 Non-welfarism

So far, we've looked at approaches that generally accept the idea that policy aims at welfare, even if, like Divertist, they do not think the welfare economist should aim at it directly. One alternative option is simply to deny the premise of the question and argue that policy should not be planned through predictions of effects on welfare, nor even through predictions of effects on something thought to be a proxy for welfare, but instead through something else entirely. This is far more radical a view than denying welfare is *all* that matters- in order to sustain this view, one must, hold that welfare has little or no relevance to policy. On this view, welfare, if it even exists, is not of interest to the economist, insofar as her interest is in *guiding economic policy*. The problem of interpersonal aggregation, as a problem of concern to the policy-oriented economist, is simply dissolved.

Some variations of libertarian normative theory do this. The policy of the state must follow the dictates of an extremely demanding series of constraints. After these constraints are considered, there is scarcely any room left for objectives like maximizing welfare. For example, Nozick (1974) and Rand (2011) advocate a small state with very minimal responsibilities focused on law, policing and the military.

There may be some room for welfare analysis in the construction of law (Kaplow & Shavell, 2000, 2002), but even the directions that law is developed may end up highly constrained by libertarian principles. The result is a normative framework for running the state with, at least, a reduced need for welfare analysis and perhaps no need at all.

Other than objecting to such throughgoing forms of non-welfarism by attacking their underlying philosophies, there are a few lines of refutation here. One is that it is surely implausible that welfare will *never* matter in economic decision-making. One can imagine, for example, that a non-government economic planner of some sort might want to increase welfare among their (voluntary) charges- for example, the manager of a large voluntary residential community with an internal economy. Even a libertarian must acknowledge that it is proper for such a decisionmaker to be concerned with welfare and that they have a large number of economic levers through which they can influence welfare, thus we arrive back at the desirability of a solution to the interpersonal aggregation problem.

2.6. Little's criterion

Little (1950) proposed a criterion - a policy is to be recommended if it satisfied both the Hicks and Kaldor criteria, as in the Scitovsky criterion (1941a), and it does not make the distribution of income worse (we will leave aside subtleties in the definition of 'total income' here).

The main problems with little's criterion are:

> A) It is an incomplete rule. Multiple proposals can meet the criterion- yet none of them may be strictly dominant over the other- for example, one might be better for efficiency and the other for distribution, even though both improve both. (Kennedy, 1952; Meade, 1959).
>
> B) In order to apply Little's criterion we need to specify a normative measure of inequality. This at least implicitly gives us a social welfare function (Aigner & Heins, 1967; Chipman, 1974; Blackorby & Donaldson, 1978; Osmani, 1978; A. K. Sen, 1982; Dagum, 1990; Atkinson & Brandolini, 2015; Brandolini & Carta, 2016; McGregor et al., 2019). To do so in a principled way- through evaluation of welfare effects- will most likely lead us to realism, to do so ion the basis of direct judgements will lead us to Stipulationism. If the SWF associated with the normative measure of inequality is utilitarian, then the measure of inequality implies the existence of a cardinal and interpersonally comparable measure of welfare (Arrow, 1951a, p. 1928) thus many critics asked of Little's approach (Kennedy, 1952; Reder, 1952; Robertson, 1954, 1962) why not just use the information to calibrate a SWF instead?
>
> C) Is it really plausible that there could never be a good policy that made the distribution even slightly worse? Or that slightly reduced income but greatly improved the distribution?
>
> D) The ways one can bundle different policies interact with Little's criterion. Proposals necessarily involve different components. Consider component A which greatly decreases equality, but also greatly increases efficiency, and component B which greatly increases equality- even more than component A decreases it- but somewhat decreases efficiency. A combined proposal of A&B passes little's criterion, but neither A nor B alone passes it. By altering the values, we can get the result that A&B or A alone pass, but not B alone, that ABC passes, and A passes, but not AB or AC and so on. The proposals might not even relate to each other substantively, but be bundled purely for the sake of passing the criterion Whether or not this is a disadvantage may be a matter of taste, but it seems like it might be constraining on a government that genuinely wanted to improve welfare but didn't want to have to worry about how it packaged every proposal it made.

We can deal with objections A & C by retreating a little and making it clear that Little's criteria is not an all-round assessment criterion- it's just a useful principle to keep in mind along with others- but this takes us back to step one, trying to work out some more comprehensive guidelines for reviewing harms and benefits to different individuals.

2.7. The social welfare approach and income inequality aversion

Social welfare functions (SWF) are a widely used welfare economic approach that can complement many of the approaches we have described, including realism and stipulationism. We introduce them here because we will refer to them throughout the thesis. SWF can be interpreted as the ethical preferences of a particular social planner, or the point of view of a particular ethical theory. Alternatively, and more rarely acknowledged in the literature, they can represent an ethical *fragment* in the sense of constituting a judgement over the welfare related aspects of a prospective action, but not a judgement on the desirability of the action as a whole. Popular SWF are often individualistic, meaning that social welfare can be written as the sum of the valuations of each individual's welfare. The general form of such an individualistic SWF is:

$$W = \sum_{i=1}^{n} v_i(y_i)$$

where W is social welfare, and v_i is the valuation of the total income or money metric utility y_i of the i'th individual.

It is common for the value function $v_i(y_i)$ to be isoelastic, that is it assumes a constant income elasticity of the marginal valuation of income. This elasticity is denoted by ε (epsilon) and can be interpreted as a measure of the income inequality aversion of the social welfare function. Epsilon is defined as follows:

$$\varepsilon = -\frac{\partial \ln \frac{\partial v}{\partial y}}{\partial \ln y}$$

Giving the following weights on income increases:

$$w = \frac{\partial v}{\partial y} = y^{-\varepsilon}$$

implying the following value function:

$$v_i(y_i|\varepsilon) = \frac{y^{1-\varepsilon} - 1}{1 - \varepsilon} + c$$

But how are we to select the inequality aversion parameter ε? We can use stipulationist methods to select ε, for example, choosing a round or plausible seeming figure. Alternatively, we can adopt an ethical stance (e.g. utilitarianism or prioritarianism) in combination with an estimation- whether behavioural or psychometric- of the function between income and welfare.

Let us explore the second option. We can fix ε by making v a function of utility or welfare

$$v = f(u)$$

In utilitarianism, the function is the identity function, thus v equals u. On this approach, assuming welfare is isoelastic in income, both our welfare *and* our value function will be isoelastic. We then have:

$$v = u = \frac{y^{1-\eta} - 1}{1-\eta} + c$$

where η is the income elasticity of the marginal welfare of income. Thus estimation of η then acquires special importance for policy assessment for the utilitarian (Evans, 2004, 2005; Evans et al., 2005; Evans & Sezer, 2004, 2005; Latty, 2011) as this allows for the determination of ε.

Such an approach incurs objections. First of all, utilitarianism is far from an obvious truth, at least to many, and imposing it in the context of policy assessment may be ethically sectarian (Creedy 2007). Secondly, our assumptions about the relationship between income and welfare may be simplistic, for example, the welfare individuals derive from income might be affected by the income of others, or policies might have broader effects than those we can model directly via changes in income.

2.7.1 Applied weighted CBA- the case of the HM Treasury Green Book

Weighted cost-benefit analysis is an attempt to implement these ideas in welfare economics by multiplying costs and benefits accruing to an individual estimated in monetary terms, by income weights, where these weights decline with income. An example of applied weighted CBA is the procedure recommended in the UK HM Treasury Green book. The Green book calls for the use of weighted CBA in cases where policy notably affects the income distribution:

> Where redistribution is a policy objective such as payments under the welfare system or if it is highly significant in terms of the impact on incomes and welfare of those affected then a weighted and equivalised income distribution analysis may be justified . . . In weighted analysis, financial benefits for lower income households are given a higher social value than the equivalent benefits for higher income households. Weighted estimates should be presented alongside unweighted estimates to demonstrate the impact of the weighting process. (HM Treasury, 2022, Section 11.A3)

A social welfare function and a corresponding set of weights is adopted, and in the case of the Green Book, the approach is utilitarian (leaving aside concerns like relative income), and ε is therefore set to η, where η is derived primarily from analysis of economic behaviour, rather than psychometrics. The Green Book (2022) recommends using $\varepsilon = 1.3$ and gives a justification for this approach that seems to assume utilitarianism, by citing evidence that $\eta \sim 1.3$:

> The basis for distributional weights is the economic principle of the diminishing marginal utility of income . . . Higher estimates of the [income elasticity of] marginal utility of income will mean the value of an additional pound declines more quickly relative to increases in income . . . A review of international evidence provides an estimate of the [income elasticity of] marginal utility of income at 1.3. (HM Treasury, 2022, Section 11.A3)

Occasionally the term weighted cost-benefit analysis will be used to describe weighting by some other criteria than income- for example, weighting in favour of a particular constituency. Unless specified otherwise, when we talk of weighted CBA we refer to the use of income weights to favour the poor.

2.8 Conclusions for chapter two

We have reviewed a variety of approaches to welfare economics. Our method of taxonomy is not exhaustive, includes overlapping categories and does not necessarily correspond to the classical self-conceived fault lines of the field, but seems to me to provide an interesting framework for conceptualising the topic. Finally, we briefly introduced some of the machinery of Social-Welfare Functions, which play a prominent role in welfare economics, and which we will discuss frequently throughout the thesis.

3 An introduction to the psychometrics of subjective wellbeing

3.1 Introduction

Our argument in this thesis is that a psychometrically informed approach to practical welfare economics can resolve or alleviate many of the traditional problems of welfare economics. In a key paper Angner (2013) argues that the widespread belief among economists that happiness (and presumably other kinds of subjective wellbeing) cannot be measured must be attributed either to a misunderstanding or ignorance of the psychometric approach to measurement or to a disagreement with that approach to measurement. Dealing with the latter problem- except in cursory detail- is beyond the scope of this thesis, although we do consider a few choice objections. We are not primarily aiming in this thesis to defend or exonerate psychometrics. I will use this chapter to combat the former problem by laying out a methodological sketch of psychometrics. I review psychometrics as a discipline, both in basic terms for those unacquainted with it, and then in terms of a selection of the deeper philosophical and practical problematics that have been raised around it.

The purpose of this chapter is to lay out certain aspects of the methodology of the psychometrics of subjective wellbeing in a relatively brief form. This chapter is intended as much as a reference for understanding other parts of the thesis as an original contribution, though I do give thoughts on the subject. As a guide to the scientific state of the field, we largely take as authoritative the *OECD Guidelines on Measuring Subjective Well-being* (2013)*.* In terms of our philosophical approach to construct validation in psychology, we concur on many points with *Is Construct Validation Valid?* By Alexandrova and Haybron (2016) - with the proviso that they tend towards a somewhat more sceptical approach than mine, and I have concerns about the practical difficulties associated with their proposal of integrating philosophical ideas into ongoing work in psychometrics.

Psychometrics, in the sense that it is used in this thesis, is the attempt to quantify difficult to quantify psychological traits, states and other psychological features of a person. It is both a *field of study* which is -at least in principle- not tied to any methodology and a collection of methodologies that have been developed in pursuing this field of inquiry. Psychometrics could perhaps be said to boast its own (domain specific) *theory of measurement* - at least in the limited sense that it possesses a unique collection of rules, practices and standards developed by the field that allow the assignment of numbers to its objects of study, and (putatively) justify the use of mathematical operations on those numbers. Angner (2013) more modestly calls it an *approach* to measurement.

3.1.1 The basic concepts and a trip to Babylon

We will define a number of terms in quick succession.

By *construct* I mean a theoretical psychological entity. Examples include joy, creativity and extroversion. Folk psychology is a rich source of constructs, but constructs do not have to be entities that appear in folk psychology- some, such as attachment styles, originated as concepts in psychology itself.

By *measure* is meant a procedure- designed to measure a person's level of a construct.

A *reliable* measure is a measure that, roughly speaking, will produce the same outcome consistently (OECD, 2013, p. 44). There are multiple measures of reliability. Two of the most important are split-half reliability and test-retest reliability. Split half reliability is an assessment of the reliability based on the degree of correlation between an (ideally randomly selected) half of the questions, and the other half. Test-retest reliability is based on the level of correlation between the measure given to the same subject at time A and time B, (where the gap between A and B is small enough that we think that the true underlying score should not have shifted for theoretical reasons.

A *nomological network* (Cronbach & Meehl, 1955) is a postulated series of relationships between constructs. This postulated series of relationships implies observations about the relationships between different measures, should those measures validly track their constructs. It is postulated on the basis of theory, whether folk, scientific or both.

By *validity*, we mean the extent to which the measure truly measures the concept/construct it aims to. Now we face a difficulty- how to measure the validity of a construct? The construct is what we are trying to measure, how can we assess the degree to which the measure correlates with the construct? When it comes to psychological constructs, we don't always have an unambiguous behavioural definition which would allow us to access the construct separately from the measure. We could of course use another measure developed independently (and this indeed is one strategy for validation) but for obvious reasons, it's not clear whether this gets us out, or simply broadens the vicious circle. The solution psychology has developed is, in its broadest sense, an application of the scientific method. Make a number of predictions about what we would expect to see- or at least would probably expect to see if the measure is indeed a good measure- and then see if that prediction holds up. There are different kinds of predictions, thus there are different kinds of validity.

If a test measures happiness, one thing we might predict is that the questions will be about happiness directly or things that are related to happiness. We would be sceptical if we opened up the test and found a series of questions about the respondent's opinion of highway building programs in the United Kingdom, instead of questions related in some way to whether the respondent was content, joyous or fulfilled, for example. This is the issue of *face validity.* Face validity is often measured non-quantitatively, but there are certainly ways it could be measured quantitatively, using panels of experts, non-experts or both. Face validity, as you may well imagine, is a deeply controversial subject because, in a certain light, it appears vague and perhaps even unscientific. As an aside, my own opinion is that face validity is absolutely key to psychometrics, that even when we claim not to be relying on face validity we implicitly are and the practice of assessing face validity is methodologically indispensable. Thus, attempts to banish face validity really only send it underground. I plan to expand on this in future research.

A good measure of something should measure all aspects of it, or at least target a variety of different important aspects. *Content validity* is related to, but distinct from, face validity. Content validity, sometimes called logical validity, is concerned with the degree to which the content of the measure appears, often to experts, to cover all the different facets of the posited construct (Salkind, 2010). To take an example from Alexandrova & Haybron (2016) It would be wholly appropriate for a measure of happiness to contain questions about whether the respondent is frequently excited, but if it contained nothing but questions about this, the construct would lack content validity as a measure of happiness.

Psychometrics is typically concerned with constructs that should, at least in theory, be associated with measurable outcomes for a subject. Often these relationships are very direct and obvious. We would think that there was something wrong with a measure of happiness that had no correlation whatsoever with a subject’s tendency to laugh, especially sincere laughter, and engage in Duchene smiling (smiles that affect the muscles around the eyes- relatively difficult to fake, and taken as more sincere- e.g. (Ekman et al., 1990; M. G. Frank & Ekman, 1993; Gunnery et al., 2013). Thus, if we think a measure is a valid measure, we would predict that it would correlate with certain outcomes for the subject. These posited relationships between variables, observed and unobserved, are the nomological networks we mentioned earlier. This is known as criterion validity (McDonald, 2005).

Although criterion validity is usually treated as a unity, I taxonomize it further into three components: *behavioural validity, situational validity* and *biological validity*.

> ***Behavioural validity***- The correlation of a measure with certain relevant behaviours- and what can reasonably be assumed to be at least partially outcomes of behaviours- for the subject. These might include anything from Duchenne smiles for happiness to criminal convictions for aggression. Assuming we found such relationships, we would say that the measure displays *behavioural validity.*
>
> ***Situational validity***- The correlation of a measure with relevant factors in the subject’s environment. We would expect, for example, a painful event like a death in the family to be related to reduced happiness. Assuming we found this, we would say that the measure displays *situational validity*.
>
> ***Biological validity***- We have reason to believe that certain biological states and changes are related to certain types of emotions. For example, cortisol is a highly evolutionarily conserved stress hormone (Haase et al., 2016). It would be a surprise if people who were

happier tended to have higher levels of cortisol.

Note that in all of the above, we have been talking of *correlation* between the measure and the variable of interest, but more intricate statistical relations could be investigated. For example, we could examine the Granger causation or Autoregressive Distributed Lag relations between the measure and behaviours, or the measure and situations. Various more sophisticated approaches, such as path modelling can be used as well, as well as experimental designs in some cases.

Convergent validity is observed when two tests intended to measure the same thing are highly correlated (Chin & Yao, 2014). One might think that it is only possible to usefully deploy the concept of convergent validity when one already has a measure that has proven valid, in order to “pass that validity on” to another, perhaps newly developed, test. However, if two tests were developed independently, both intended to measure the same construct and displaying face validity, this may provide some evidence, even if not very strong, that this intercorrelation reflects a common capacity to measure the construct.

To establish *discriminant validity* vis a vis two measures is to establish the inverse of convergent validity- to show that two measures of constructs which are not thought to be related do not, in fact, correlate with each other (Hubley, 2014). Discriminant validity can be a difficult concept, insomuch as even if two things are not conceptually related, they might be closely tied to each other in the real world. If you did not yet know much about chemistry you might think that the concept of gasoline has little, if anything, to do with the concept of fire, yet in empirical reality, they are closely related to each other. Similarly, concepts that we expect to be unrelated in our study of discriminant validity may be tightly connected in any number of ways, without our being aware of it. Despite these difficulties discriminant validity is still considered an important part of psychometrics.

The concept of validity is dependent on theory- we must already have a number of ideas about how things should be related to begin the process of validation. This interplay- testing pre-existing assumptions against each other, rather than starting “from the ground up” has led some to comment that the enterprise has a sense of coherentist epistemology about it (Alexandrova & Haybron, 2016). This is often referred to in terms of a pre-existing body of theory (e.g. see Hagger, Guccicardi and Chatzisarantis (2017) which frames this ‘coherentist flavour’ in terms of “nomological validity” and nomological networks, but treats a nomological network as something very like a theory) but an interesting question is, is this really about pre-existing-scientific-theory, or is it really about leveraging preexisting folk psychology? I will leave the question open here, but my sense is that folk psychology, even more than scientific theory, is inextricable from many applications of the construct validation framework, *and this is a perfectly fine thing*. This is connected to the issues of face validity I raised earlier.

Regardless of whether we think of validity as informed by scientific or folk theory, the resulting methodology is quite different from what one generally sees in economics. Alexandrova and

Haybron's (2016) suggestion that psychometrics is coherentist is useful to think about here, especially when compared with the axiomatic approach that economists take to the definition and measurement of welfare, which might be seen as at least loosely corresponding to foundationalism in the sense epistemologists use the term. Angner (2013) aptly frames the tension as a clash between two theories of measurement, the *representational theory of measurement* deployed by most economists and the *psychometric approach to measurement* deployed by most psychologists. Since this thesis is an intervention in economics, it may be useful to bring in a comparison to what Dow (2012), in reference to the work of Keynes and Feynman, has called a distinction between *Babylonian* and *Euclidean* methodology.

Whereas the Babylonian methodology uses a variety of distinct yet overlapping models and arguments, none of which on its own is necessarily seen as decisive, the Euclidean methodology starts with one set of axioms and attempts to prove a conclusion using one decisive chain of reasoning. Dow also regards the Euclidean approach as *dualistic*- based on the use of either/or variables such as "either a variable is endogenous, or it is exogenous" and *atomistic* - starting with the simplest possible building blocks. By contrast, the Babylonian approach depends upon multifaceted systems. Variables have a complex, independent existence which we do not fully understand when we include them in our models. In a move that may surprise some readers, Dow also insists that it is the Babylonian approach- not the Euclidean approach- that is more likely to be *realist* in the sense that term has in the philosophy of science. That is to say the Babylonian approach treats the objects of its study as having a mind-independent existence and not being mere "useful fictions" but instead entities which exist apart from our mentations, that our mentations can more or less adequately describe. Superficially at least this is the opposite view to Friedman's (1953) contribution to economic methodology.

The question of whether the Euclidean or Babylonian approach is better, or whether the contest is even useful or ultimately meaningful, is well beyond the scope of this thesis, nevertheless, we flag the difference in mindset by way of introduction.

3.2 A note on circularity, coherentism and foundherentism

Either an economist or philosopher might worry about circularity at this point. How do we know these measures are valid? Why, in part through comparing them to related psychological constructs! Doesn't this make the approach viciously circular?

Recall our earlier reference to Alexandrova and Haybron's coherentist approach. They write:

> *We believe that this vision of measure validation is defensible. Its spirit is remarkably similar to the coherentist vision that characterizes recent work on measurement of physical quantities (Chang, 2004; Fraassen & C, 2008; Tal, 2013). These philosophers emphasize that the outlines of the concept in question, be it temperature or time, and the procedure for detecting it are settled not separately but*

> *iteratively, checking and correcting one against another. (Alexandrova & Haybron, 2016, p. 1102)*

This hints at a way of thinking about the problem which does not require us to embrace coherentism in any strict sense, as a global epistemological theory. What we are aiming at in this kind of psychometric procedure is best conceptualized as mutual support. To put it in vaguely Bayesian terms, we have some degree of prior confidence that each of our measures indeed measures what it purports to simply on its face validity. The existence of the correct relationships between these measures then puts the initial weak credence we had for each construct, in a relationship of mutual support. External factors- both behaviours and environmental factors- that we find correlated with our construct measures, add further confidence. I will note in passing that by making explicit the role of prior belief in networks of theoretical and observed entities, and adjusting from there, psychometrics may be *the* ideal case for a Bayesian case study in the epistemology of science. Additionally, it is worth noting that the mixture of external validation with internal co-support particularly calls to mind the flavour of Haack's (2009) foundherentism.

3.3. Just how advanced is the study of subjective wellbeing from an empirical and methodological point of view?

Many volumes could be written on the methodological fitness for task of contemporary subjective wellbeing studies and many have been (Bruni & Porta, 2005; B. S. Frey & Stutzer, 2002; Layard, 2005; Y.-K. Ng & Ho, 2006). It is not our intention to revisit this topic in this thesis, except in a few select aspects which I will discuss in further detail primarily for illustration. A statement of the position I assume in this thesis is given by the OECD (2013) in *OECD Guidelines on Measuring Subjective-Wellbeing*:

> *There is a large body of evidence on the reliability and validity of measures of subjective well-being and the methodological challenges involved in collecting and analysing such data. Indeed, given the academic interest in the topic and the challenging nature of the subject, the body of evidence on the strengths and weaknesses of measures of subjective well-being exceeds that for many measures regularly collected as part of official statistics (Smith, 2013). While measures of subjective well-being have some important limitations, there is no case for simply considering subjective measures of well-being "beyond the scope" of official statistics. Although subject to some methodological limitations, it is clear that for many of the purposes for which they are desired, measures of subjective well-being are able to meet the basic standard of "fit for purpose". (OECD, 2013, p. 12)*

With that said, let us briefly review some of the specific empirical aspects of subjective-wellbeing studies.

We have outlined the basis on which a mental state like happiness which is not directly observed could, in principle, be measured on a psychometric basis. In what follows I will briefly summarize empirical evidence for the adequacy of psychometric measures of happiness. I will often draw in my empirical presentation on the results discussed in OECD (2013)

(0) Reliability, the Ur condition

Reliability is not validity, but it places an upper bound on how valid a measure can be. In its summary of the literature, the OECD (2013) handbook suggests that multi-item subjective wellbeing measures have acceptable reliability and single-item measures most likely do as well, although here the evidence is less clear. It is worth noting that for many economic purposes what will matter will be the reliability at the population level, which will by necessity be higher than the individual reliability (though the same need not be true of validity).

(1) Face validity

As the OECD (2013) handbook notes, questions about happiness clearly intuitively relate to the idea of happiness- though c.f. Alexandrova & Haybron (2016) for philosophical concerns about certain measures.

The OECD handbook also argues that measures of respondent confusion about questions- like non-response rates and time to answer questions- can be viewed as a measure of face validity. On both of these criteria, measures of subjective wellbeing generally do well. However, I would argue that this line of reasoning is weakly supportive at best, because questions can "make sense" in a way that will lead respondents to answer them and answer them quickly without making sense *as questions for measuring happiness*.

(2) Convergent validity:

Multiple measures

One of the simplest forms of convergent validity is ensuring that multiple different measures of the same kind of subjective wellbeing, ideally created by different research groups, correlate with each other. This at least shows that what we are measuring is not centred on any "quirks" specific to one measurement. Another strategy is to look at the correlation between multiple measures each purporting to measure overlapping but distinct constructs- e.g. different forms of subjective-wellbeing. An example of multiple measures of subjective wellbeing intercorrelating is Clark and Senik (2011) who found a correlation greater than 0.6 between life satisfaction and happiness (despite transforming both into a binary variable, likely to resulting in a substantially reduced correlation). Where different measures of subjective-wellbeing- including different kinds of subjective-wellbeing, conceived of and created independently intercorrelate, this gives us some basis for confidence in both.

Related psychological constructs

If a measure is working, we would expect it to correlate well with measures intended to track other constructs which we believe the first construct should be related to for theoretical reasons. These theoretical reasons could be anything from prior research to clinical experience, or, often, folk psychology. While above we discussed the intercorrelation between measures of different kinds of subjective wellbeing, we would expect correlations to run further afield than that.

Examples of constructs which should be related to happiness include loneliness, autonomy, energy, mental health and suicidality. There is good evidence that many of these constructs correlate well with happiness. For example, Diener (2012) summarises research on the relationship between suicidal ideation and happiness and finds a correlation of about R=0.44. Bergsma and Veenhoven (2011) find that 90% of people without a mental disorder felt happy in the previous week, whereas less than 70% of people with a mental disorder did.

(3) Criterion validity

Behavioural criterion validity

Moving out of the realm of subjective assessments- we have the observation of behaviour as another support for validity. This could range from the period of time spent smiling during an interview to the number of suicides per year across differing populations. An example of evidence suggesting behavioural criterion validity is correlation between Duchenne smiles and SWB (Ekman et al., 1988, 1990) as discussed alongside other evidence by Di Tella and McCulloch (2006).

One type of behavioural criterion validity, which also might be classified as convergent validity, is third-person ratings. Just as there are good reasons to think that people have understanding and awareness of their own mental states, so there are good reasons to think that people have understanding and awareness of the mental states of others. Third-person ratings from friends, family members, clinicians and trained interviewers have all been found to correlate well with measures of subjective wellbeing. For example, correlations between 0.43 and 0.66 were found between self-assessment of life satisfaction and interviewer assessment in a review of the literature by (Pavot & Diener, 1993).

Situational criterion validity

Finding that the subject’s behaviours have the right relationship with posited constructs and their measures is only part of the story. Another important piece is finding the right relationships between measures and the subject’s situation. For example, one factor that we have reason to believe will be related to happiness, and the economist will likely be especially interested in, is income. Other obvious choices include recent stressors. With regards to the expected relationship between happiness and income we see mixed results- income appears to correlate with happiness at a given moment, but as Easterlin (Easterlin, 1974, 2017; Easterlin et al., 2010; 2020) showed, does not have such a straightforward correlation with economic growth in society over time. The relationship between stress and happiness is consistent, but perhaps unsurprisingly complex. Fabricatore, Handal and Fenzel (2000) find the relationship is mediated by spirituality. Hwang & Lee (2018) find the relationship is mediated by self-esteem and growth mindset.

However, these are only measures of subjective stress, measures of the relationship between objective stress and happiness- as measured, for example, by the Holmes-Rahe stress scale are surprisingly weak. For example, Lyubomirsky and Lepper (1999) find no significant relationship between the Holmes and Rahe stressful events scale and happiness in two different samples. Lyubomirsky and Lepper themselves profess not to be surprised by this, writing: “According to the subjectivist approach, constructs that theoretically should not be related to happiness include college grade point average, math and verbal ability, and stressful life events.” They thus argue that the failure of their happiness scale to correlate negatively with stressful events shows its discriminant validity! Lu (1995) also found a null result- no link between their measure of happiness and Holmes-Rahe stress (though links were found with poor mental health and life satisfaction). Kozma & Stones (1983) found a significant but small relationship between “life events” (the Holmes-Rahe Scale) and happiness. It may be that the Holmes and Rahe scale is too much of an aggregate, and a closer look at more specific factors would reveal important effects of situational stress. For example, Knabe and Ratzel (2010) found substantial effects of both unemployment (where prospects of finding another job are thought to be poor) and job insecurity on subjective wellbeing. Krauss and Graham (2013) find that unemployment is related to reduced subjective well-being, and so on (for a metanalysis on unemployment see Gedikli et al. (2023)).

Biological criterion validity

As we have previously discussed, the biological validity of a measure is the degree to which it correlates with the biological features that we would expect for theoretical reasons. Biological correlates of happiness might include hormone levels, neurological correlates, or potentially general health.

There is evidence of the biological validity of measures of subjective wellbeing. Kong et al. (2015) found various neurological correlates of subjective wellbeing. Birchler-Pedross et al. (2009) found effects of the circadian cycle and sleep deprivation on subjective wellbeing. Diener and Chan (2011) find evidence that subjective wellbeing is associated with living longer even with controls. (2008).

Subjective wellbeing is anti-correlated with stress related neuroendocrine responses, with elevated cortisol, fibrinogen, C-reactive protein, and other markers elevated in those with low SWB (Steptoe et al., 2005, 2007, 2008, 2009; Steptoe & Wardle, 2005; Dockray & Steptoe, 2010). For example, Steptoe, Wardle and Marmot, (2005) found that the lowest quintile of happiness had cortisol 32.1% higher than the fifth quintile of happiness. Another case study is the stress hormone fibrinogen, which tends to be elevated in those with low SWB, or to increase more in response to stressors in those with low SWB (Brunner et al., 1996; Kim et al., 2016;

Panagiotakos et al., 2004; Steptoe et al., 2003; Yang et al., 2014). Similarly, Blanchenflower and Oswald (2008) find that rates of hypertension are lower in countries with high psychometrically measured SWB.

3.4. Brief overview of selected relevant philosophical writings on psychometrics

Angner on welfare economics and psychometrics

Angner (2013) argues that there has been a lack of understanding between psychometrics and welfare economics. While Angner accepts the basic logic of the psychometric approach to measurement but does not think this salvages psychometrics *as an adjunct to welfare economics*, holding that while happiness *in the sense of the psychometrician* is indeed measurable, the measurement of happiness should not be a goal, or at least not the primary goal, of welfare economics, because welfare is not just a mental state. Exactly what he takes welfare to be, Angner does not quite say, but he argues that the equivocation between *happiness* in the sense that has moral significance and *happiness* in the sense of a particular mental state studied by psychometricians, is dangerous and pernicious.

We will argue later in this thesis in chapter 9 against this objection, on the grounds that psychometrics may be capable of measuring, with reasonable accuracy, welfare, even on non-mental state accounts of what welfare is like the desire satisfaction account. Indeed, we will argue that psychometrics can likely measure desire satisfaction better than revealed preferences measures.

But what if no existing account of welfare is satisfactory- what if welfare is something too elusive to be defined precisely enough for use by the psychometrician or even philosopher? Or what if we did accept Angner's premise that psychometrics depends on mental state accounts of welfare, and breaks down outside these, at least as a way of measuring welfare? In this case, the kinds of wellbeing-like constructs that the psychometrician can measure might still provide *evidence* about the nature, level and distribution of true wellbeing in a point similar to that (Adler & Posner, 1999; Hausman, 2010, 2011) make about cost-benefit analysis. Even if, say, measured happiness is not welfare, it is likely to be evidence of welfare, especially, at a population level, and especially across multiple different forms of subjective-wellbeing measurement.

Alexandrova and Haybron

Alexandrova and Haybron support the *idea* of construct validation but argue that, in practice, it is insufficiently responsive to philosophical theories of the constructs involved. For Alexandrova and Haybron, the construct validation process is sensible because the idea of testing a nomological network as first propounded in Cronbach and Meehl (1955) is in line with increasingly popular views in the philosophy of measurement which emphasize the iterative and coherentist logic of the process. Nonetheless they argue that in practice it is insufficiently *theoretical*. In particular, Alexandrova and Haybron think that construct validation should be responsive to philosophical theories of emotions, personality traits etc. An important type of theory in this regard is *normative* theories, and this is especially true in relation to philosophical theories of wellbeing. They argue our concepts of certain emotions are inherently normative- the very meaning of happiness is tied up with it being, either in itself inherently good for you, or at least, in itself, part of the way to being good for you, or tending to be related to things that are, intrinsically good for you. Thus, certain theories require what can be called "normative validity". Additionally, naive empiricism could lead to misleading results because what counts as a "plausible" series of correlations between the construct under investigation and outcome measures can be very open to disputation.

Alexandrova and Haybron give, as an example of a measure that has become impoverished by not being tied closely enough to theory, the PANAS which:

> *is commonly used to measure the affective dimensions of subjective well-being. This twenty-item questionnaire asks subjects to rate themselves on whether they feel enthusiastic, interested, excited, strong, alert, proud, active, determined, attentive and inspired etc. (Watson et al 1988). All these items have passed factor analysis and other standard psychometric tests. But note that absent from this list are cheerfulness, joy, laughter, sadness, depression, tranquillity, anxiety, stress, weariness – emotions that are intuitively far more central to a happy psychological state, and to well-being.(Alexandrova & Haybron, 2016, p. 1105).*

One might say that while the PANAS list seems to be very closely related to what we might term *vigour,* Alexandrova and Haybron's list conceives of positive emotions in terms that are arguably more central to the concept, a concept of *positive valence* that is agnostic on whether the subject is active or passive.

Alexandrova and Haybron have a diagnosis of how the PANAS came to be like this- they attribute it to an overreliance on factor analysis – a statistical technique often used to try to find patterns in data- rather than theory. A list of positive emotion English terms was used in factor analysis, and this was the result, but such a procedure is, on their account, really a way of avoiding the 'hard theoretical questions'.

It is instructive in this regard to consider what is sometimes called the crud effect. The crud effect, as defined by Orben and Lakens (2020, p. 241) is: "The epistemological concept that, in correlational research, all variables are connected through causal structures, which result in real non zero correlations between all variables in any given data set." Particularly in the human sciences, almost everything is causally related to everything else, whether directly, or indirectly, through several intermediate steps, in a web of dizzying complexity and endless mediating and moderating factors. This means that as one variable changes, so do others, and the existence of a correlation between variables can have very little theoretical significance. For example, a correlation between *joy* and *tranquillity* might represent nothing but the tendency of people who prefer positive descriptions of themselves to prefer both descriptions, or the tendency of people in a secure social and economic situation to feel both. That correlation does not necessarily reflect a direct relation is a well understood point, but in the context of the vastly complex manifold of different aspects of the human condition, it takes on new meaning through the crud effect where almost everything is correlated, at least to some degree and usually for many reasons.

I largely agree with Alexandrova and Haybron on all counts although I have some concerns about the practical difficulties of trying to get philosophy involved in the day-to-day practice of the science of Subjective-Wellbeing. The claim that the exact parameters of the definition of happiness (in the sense of happiness as a particular emotional state) might be set partly by normative concerns is plausibly true. However, we must concede that it creates problems for this thesis. Our goal is to imagine a form of welfare economics that can provide ethically relevant information, without needing to posit greatly ethically controversial premises. If Alexandrova and Haybron are correct about the normative roots of these psychological concepts, it becomes dubious whether we can achieve our aim. We can only plead, once again, that this is not a thesis on the philosophy of psychometrics, thus we cannot consider all the fallout of this debate. Still, it may prove the case that even if Alexandrova and Haybron are right that normativity is unavoidable in creating measurements of concepts of subjective wellbeing, nevertheless the kind of normativity involved in properly creating measurements of emotions like “happiness” and “life satisfaction” is less likely to break the consensus than the kinds of normativity involved in directly making judgements over the just distribution of wealth. We are not opposed to value judgements, rather we are trying to find a workable set of value judgements that will not breed too much sound and fury. Additionally, the use of a variety of wellbeing measures may help mitigate any ethical sectarianism.

3.4.1 Some prominent critics of psychometrics as a field

Although this thesis is not primarily on the philosophy of psychometrics it would be remiss not to mention the numerous critics of the field.

There are open philosophical questions about psychometrics and the philosophy of measurement that undergirds it. A prominent critic of the field as a whole is Joel Michell (Michell, 2000) who argues that psychometrics is a pathological science, in that it makes strong assumptions about the measurability of its subject but, much worse than that, conceals the strong assumptions it makes. There have been various responses, positive and negative to Michell in print, including (Borsboom & Mellenbergh, 2004) and (Kyngdon, 2008).

The field of psychometrics generally is also no stranger to high profile public criticism, although the measurement of subjective wellbeing has usually been one of the least controversial aspects of psychometrics in this regard. Attempts to measure intelligence, and to a lesser extent, attempts to measure personality, have been seen as very politically charged in their implications, and have been criticised for a wide variety of reasons (e.g., Stephen Gould’s “The Mismeasure of Man” (1996)).

The existence of these criticisms should not be surprising. Psychometrics:

A Deals with extremely controversial and politically relevant subjects- as well as subjects integrally wound up in people's self-image.
B Pushes our ability to 'quantify the unquantifiable' to its limits. And:
C Uses its own suite of statistical methods which are complex, involve many degrees of freedom, and are often idiosyncratic to the field, or rarely used elsewhere.

These are all valid grounds for concern and inquiry, but this is beyond the scope of this thesis. Except where indicated, we take psychometrics- at least in its well-established results- at its face and see what follows from that assumption.

3.5 Some general methodological thoughts on psychometrics and the philosophy of psychometrics

Although many philosophers in an absolute sense of 'many' have considered psychometrics It is perhaps surprising that there has not been more philosophical study of the field. Psychometrics is not only philosophically fascinating in its own right, but also a foundational concern in ethics and political philosophy, for any view which depends, in whole or part, on quantifying welfare and welfare gains, unless it takes a very non-standard view of welfare, or unless it is willing to use speculative proxies. Philosophical work here is required not just from the point of view of welfare economics, not just from the point of view of utilitarianism or generalised welfarism, but from the point of view of any applied ethical practice, especially at large scale, which values human welfare to any significant degree.

As with the special issues in the study of psychometrics generally, it is not our aim to resolve these philosophical problems of the psychometrics of wellbeing in this thesis, *except where they relate to already existing problems in welfare economics*. In this thesis, we treat the psychometrics of subjective well-being as something of a scientific black box, assumed to be well functioning. We introduce it here only to give a flavour of the field. Every area in science, not just social science, has its own special philosophical problems, and the resolution of those problems is not generally a prerequisite for the application of the field in solving other issues. Physics and biology should and do interact with philosophy, but they do not wait for the philosophy to solve their undergirding methodological problems before proceeding, nor should they.

Extending on Haybron & Alexandrova (2016) and Alexandrova (2017) - how can philosophers be integrated more fully into the process of analysing psychometrics? This is a profound question for future research and of the highest priority, but I might venture to add my own question. How can the views, understanding and knowledge of ordinary people, be brought into the process? After all, we all have folk concepts about many of the things in the nomological networks psychologists' study. How can we systematically gather the implicit and explicit knowledge of ordinary people about psychology?

In theory, something like our ordinary understanding of concepts is meant to come in during the process of determining face validity, but face validity is often considered low rent. A more rigorous alternative, called *logical validity* is sometimes used, engaging panels of psychological experts. However, we may worry that the emphasis on expert involvement, far from pre-theoretical, removes the contribution that the concept itself, outside particular theories and understandings created by the field, is meant to create. Without such a contribution, we are in danger of outcomes like those Alexandrova and Haybron describe in relation to the PANAS, for it seems to me that what was missing in that case was not a philosophers or psychologist's deft touch, but rather the ordinary, obvious understanding that *central to positive emotions are good feelings*.

Regardless, we must get clear on exactly what we are looking for in studies of face validity and/or logical validity- are we looking to introduce folk knowledge to the nomological network- as a starting point (though not necessarily ending point) of our understanding, or are we instead looking to introduce theoretical expertise, or all three at once? All answers come with their own epistemological and methodological problems.

Our project in this thesis will be about providing ethically relevant information whilst avoiding sectarianism, but Alexandrova and Haybron's concept of *normative validity*, whilst well defended, raises a difficult question of its own. How can we pursue normative validity without fracturing measurements of happiness and all other thick concepts into different philosophical camps based on different theories of welfare? A related problem is that, to put it somewhat ruthlessly, philosophy is seemingly not an epistemically progressive discipline, at least not in the way we want science to be. Do we really want philosophy to hold the whip hand in psychometrics?

We have given the reader a taste of some conceptual issues that are involved in psychometrics. One could just as easily give equally interesting philosophical critiques of any number of areas of science. There is an ongoing debate in the philosophy of science about the extent to which the philosophy of science should be *regulatory* in the sense of disciplining science or telling it what to do. This debate, sometimes called the normativity debate largely concerns the question of whether the philosophy of science ought to be a *descriptive* or a *prescriptive practice* (M. I. Kaiser, 2019). The question can be difficult to formulate, the philosophy of science is so wide and heterogeneous that the question of whether as a discipline it has a “warrant” to engage in an epistemological critique might seem to be so broad as to be meaningless.

Should we “put a hold” altogether on results from psychometrics because of concern about the philosophical critiques that have been made of it? Or more concretely, are these criticisms, insightful as they are, so devastating that until they are adequately addressed, we should not regard statistics about subjective well-being as fit for use in the context of policy appraisal?

All intellectual frameworks involve taking risks and that includes psychometrics. It is possible to ask devastating methodological and empirical questions about just about any approach in the social sciences. We are not convinced that the risks here are greater than the risks attached to other paradigms, e.g., more traditional approaches in welfare economics, which, after all, face their own philosophical critics many of which we will discuss. If you do not accept that psychometrics is a sufficiently mature science for this kind of public policy work, then simply treat the rest of this thesis as a conditional argument that: *if* the psychometric approach to the measurement of subjective wellbeing works, *then* these things follow from that. That said, we will touch upon this issue again in later chapters, including chapter 6 (in relation to the problem of cardinality) and chapter 9 (in relation to the question of self-report versus revealed and stated preferences).

3.6 Easterlin's paradox and other problems of comparison between different times, places and groups

One persistent source of scepticism in psychometrics - *particularly* the psychometrics of subjective wellbeing, is that even if interpersonal comparison is possible within groups, however, defined, the comparison may be impossible between groups due to significant cultural variation. Research has suggested that there may be cultural differences in how groups understand happiness scales (e.g., Lai, Cummins & Lau (2013)- c.f. also Tov & Deiner (2009). There are similar methodological difficulties involved in comparing historical differences (Jorm & Ryan, 2014).

Most famously, such comparison problems are highlighted by the Easterlin paradox. The Easterlin paradox is an observation owing to Easterlin (1974) that increasing income in a society over time does not appear to be associated with increased happiness across that society, even though wealthier individuals at a given time are, on average, happier.

One way of understanding the Easterlin Paradox- as an artifact caused by comparison is perhaps best put by John Quiggin (2014) in Aeon magazine with an analogy to height. To summarize, suppose that you established the age of all the children in a school, recorded the grade they were in, and then asked them to rate their height on a scale from 0 to 10. As you went up the grades, you found that subjectively estimated height increased little, if at all, however, age did increase. Within a grade, you found that the difference in age was correlated with the difference in height. How to explain this? The most likely answer is that the children are rating their height by comparing themselves first and foremost with their classmates, rather than those in other grades (and by analogy, other times).

Intuitively, as an explanation of Easterlin's paradox, this is sensible. It seems to me that this theory of what is going on is wholly compatible with the approach we outline in this thesis, at least for many sorts of problems, for which we only need to know whether people are becoming more or less happy, relative to some standard, to inform our decisions. It is not necessary that we, for example, be able to judge whether those effected by our policy in the present year will be happier or less happy than a specific 1980 population- comparing how much they have gone up or down relative to our present baseline is enough to determine whether the policy is positive. Returning to our previous analogy, if a handful of students at a time were given human growth hormone, the changes on their self-rated height would show up in our statistics. Of course, this approach will leave us with difficulties in applying welfare economics to societies over time, and to whole societies at once- e.g. in evaluating growth-equality tradeoffs.

Other approaches to Easterlin's paradox are more pessimistic. It is possible, for example, that happiness *really is* determined by relative, rather than absolute wealth and thus the growing wealth of society will do nothing or little for human happiness (and might even make it worse if inequality continues to increase). This is the approach taken by Clark. Frijters, Shields (2008). Such an approach would tend to suggest that there are no, or minimal differences in scale usage. In a richer society, it really does take more income to be happy.

Determining the relative contribution of real changes and scale use changes to different scores between the same society at different moments seems little different (in some ways) from the problem of intercultural comparison. The solution to this problem will be the careful comparison of validating measures between different contexts, especially biological and behavioural measures. An interesting study, for example, would be to determine how cortisol levels change over time in a panel of countries as income increases. We could also use behavioural comparators more readily, as concern over differences in how happiness is expressed between times is presumably not as great as the difference between cultures in how happiness is expressed, at least for relatively short periods.

Though it is perhaps unlikely if we do find that there is no basis to think that increasing incomes society wide are good for welfare, we can and should face this result honestly and reevaluate our policy priorities in light of it. Certainly, the role of economic growth would become much more ambiguous from a welfare perspective, and environmental conservation, even at the cost of reduced income growth, would be a greater priority. If differences in happiness are primarily driven by the *distribution* of income rather than the *level*- this could mean that we should focus far more of our attention on distribution, with less concern for efficiency effects. Researchers have suggested that the Easterlin paradox may give us *more* reason to care about happiness measures- e.g., Jayawickreme & Pawelski (2013, p. 7) "Determining public policy based on economic indicators alone has meant that growing economic prosperity has not been accompanied by an increase of happiness… Many SWB researchers have argued that this paradox needs to be resolved, with more emphasis being placed on the well-being of citizens."

Of course, it may depend on what kind of welfare we care about. Life satisfaction is known to be more closely linked with income than happiness (Diener et al., 2010; Kahneman & Deaton, 2010). Many sources (e.g., Easterlin (2017)) suggests that the paradox applies to both happiness and life satisfaction. However, there does not appear to be a great deal of literature on the Easterlin paradox in relation to Eudaimonia.

3.6.1. Mach's problem

There is a different sort of problem in this area, which is subtler to describe, but perhaps even more philosophically interesting. What if validating measures diverge fundamentally in the relation they suggest between different populations and subjective-wellbeing scales? Consider an analogy. The great philosopher of Physics, Ernst Mach, in his "Critique of the concept of temperature" (1903)" suggests that there is no uniquely correct scale of temperature. Suppose we have a material, for example, that expands in relation to temperature, and we use this to establish a thermometer, which outputs changes in temperature proportional to changes in the volume of the material. There could just as easily be another thermometer based on another material which gives quite a different temperature mapping.

Let us introduce what we will call *Mach's problem* by analogy, on the proviso that our problem is not conceptually or mathematically identical to Mach's considerations regarding the thermometer. Consider Cantrill's ladder, a zero to 10 measurement of happiness. The psychometrician tries to validate this measure through comparison to a variety of indicators. Suppose we found that for Turkish respondents, a 6 corresponded to the same degree of Duchene smiles, cortisol, and sincere laughter as an 8 in France (these figures are purely hypothetical). All is good and well- and so we declare that, at least on average, a Turkish 6 on Cantrill's ladder is equivalent to a French 8.

Now suppose instead we found, looking at the prevalence of Duchene smiles, that a 7 on the happiness scale in Brazil was equal to a 6 in the UK, in terms of the outcome measure. But then we looked at another outcome measure, say blood cortisol, and found something more like an 8 in Brazil is equivalent to a 6 in the UK. How then should we proceed in converting the scores from one culture to the scores in another?

It could be that as we continued to make comparisons using different outcome measures, a solid majority of them will end up using a similar conversion formula between culture A scores and culture B scores. But it could equally be that different validation measures all had different slopes with respect to their association between UK and Brazilian happiness figures. We cannot rule this out a priori, and if it were to happen, it would seem to render the figures incommensurable. Have we too hastily helped ourselves to the idea of an interculturally quantifiable amount of happiness, when in truth there are many different ways of quantifying happiness similar to Mach's thoughts on temperature? To be clear, the problem we are primarily worried about is not one of *evidence* but *ontology*- on part of what it means to be happy, one comparison of scores is appropriate between cultures, on another part of what it means to be happy, another cardinalisation is appropriate.

But our worries extend well beyond intercultural comparison. We might imagine two psychometric measures of happiness, A and B. Which is best? Cortisol levels might say A, and Duchene smiles might say B. Which is correct? We might hope that, as we investigate additional validators, they will "break the tie", and indeed they might all overwhelmingly favour one measure or the other, but it is equally possible that the study of further measures might just extend the tie.

Multiple things could be driving the divergence of A and B, and the validating variables they aim to track. One of the simplest possibilities is that happiness is itself multifaceted and incapable of being reduced to a single quantity. A correlates with factor 1 better, and B correlates with factor 2 better. If the split is strong enough, we might take this as disconfirmation of the model which posited subjective well-being as a unitary construct, and confirmation of a rival model in which subjective wellbeing has multiple distinct aspects that need to be quantified separately.

Another possibility is that there are multiple factors, but only one of them legitimately represents happiness, the other represents a biasing factor which happens to be correlated with our validating measures, in this case at least there is hope that further work might unpick the knot.

There is some reason to hope that studies of the factor structure of subjective well-being might reveal these problems “in advance” as it were- with a unitary factor structure giving us reasons for optimism, and a weak common factor giving us reasons for pessimism. However, this need not be the case- it can even be that factor structure in one cultural context doesn’t equal factor structure in another cultural context. For what it’s worth, studies of the factor structure of questionnaires assessing subjective wellbeing have generally revealed a reasonable degree of unity. For example, a study by Linley et al. (2009) found a good fit by a two-factor model- factor 1 roughly representing eudaimonia/flourishing and factor 2 roughly representing both happiness (in the sense of the presence of positive emotions and the absence of negative emotions) and life satisfaction. However, while a two-factor solution fitted better than a one-factor solution, these two factors had a very strong correlation with each other (r=0.76)- notably this also provides further evidence of a claim we will often rely on in this thesis- measurements of different types of wellbeing concept are strongly correlated with each other. They also found: “that these loadings did not vary according to gender, age or ethnicity,” (Linley et al., 2009).

Mach’s problem is beyond this thesis, both because its treatment is necessarily deeply mathematical, and because the premise of this thesis is to use psychometrics to address problems in welfare economics, not to address the philosophical problems of psychometrics. Even fully describing the problem (or perhaps family of problems) requires close and extended treatment. Its resolution would involve not just technical questions- but possibly ethical questions as well. If subjective-wellbeing concepts fragment like Mach thought could happen to temperature concepts, to which portion should attach our ethical concerns? Once again we are reminded of normative validity.

3.7 Why instrumentalism will not work when your ethical concern is about the thing-in-itself

History was not kind to Ernst Mach’s concern, as we learnt more about the temperature it was largely forgotten- no citations are listed on Google scholar for Mach’s *Critique of the Concept of*

Temperature (Concluded) (1903). Temperature is now thought to be uniquely quantifiable in ordinary cases. Yet we can imagine a world where things turned out differently and different temperature scalings were appropriate for different purposes. Perhaps we can even imagine a world in which different temperature *orderings* were appropriate for different purposes. What is the metaphysics of there being, or not being, an "amount" of something in the world? Should we be realists or anti-realists about this? How does our position vis a vis being realists or anti-realists on this question of magnitude change when the magnitudes in question have an ethical value? Is there any real difference between realism and anti-realism about measures of subjective-wellbeing with regard to practical implications? I think so.

There are scientific topics about which we do not just care *extrinsically*- for example, for its worth in predicting matters that we do care about- but *intrinsically*- for their own sake. We care about how a certain drug prevents death *extrinsically-* we do not want people to die, and if the drug does prevent death, we can use it for this purpose- it will not bother us if our understanding of the mechanism is just a useful fiction. However, if a certain study suggests that a particular policy makes people *happier,* we care about whether it is correct in itself- if the idea that they are happier is merely useful fiction for predicting things, but we do not have good reason to think they are, truly and really happier, than the test has lost most of its value in a policy assessment context. There is an interesting question about whether scientific instrumentalism and other forms of scientific anti-realism can explain our willingness to rely on scientific knowledge *where the facts in themselves, not simply prediction matter*. Prima facie, either one must abandon instrumentalism, at least about these kinds of subjects, or reject relying on scientific knowledge when we need to know how things are in themselves, and not merely make predictions.

3.8 What kinds of welfare can psychometrics study?

Numerous ways to define subjective wellbeing have been developed. The four most popular philosophical approaches to welfare are the *hedonic, desire satisfaction, flourishing/eudemonic and objective list theories,* though these last two theories are sometimes wrongly equated – see Fletcher (2016) for a clear conceptualisation which brings out the differences. Psychometric methods of measuring each of these *or at least related constructs* have been developed by social scientists.

It seems to me that there are grounds for thinking that a psychometric approach can give us information about the measurement of welfare almost irrespective of our theory of what welfare is. This will be important if the argument of our thesis for a psychometrically based approach to welfare economics is not to depend upon a particular theory of the good life. There is a long-standing habit of equating the economic approach to measuring well-being with the desire/preference satisfaction theory of wellbeing, and the psychometric approach with mental state accounts of well-being, but most especially the hedonic theory. Arguably, Angner 2013 treats psychometrics as associated with hedonism, and traditional methods in welfare economics as associated with preference satisfaction. Although he does not make exactly this conflation, see also Hausman 2010 who contrasts hedonism- treated psychologically- with preference-based approaches, treated non-psychologically. The debate between the psychometric approach and the traditional approach in economics is thus transformed into a normative debate between the preference satisfaction and mental state accounts of wellbeing. While it is true that the traditional approach in economics to welfare measurement cannot directly capture mental states (or at any rate, many important mental states like pleasure), it is far less clear that the psychometric approach cannot be used to measure preference satisfaction at least as well as methodologies like the revealed preferences approach. This is true even though, as Angner reminds us, preference satisfaction accounts of welfare are not mental state accounts of welfare at all (Angner 2013) and superficially we might expect psychometrics to be concerned with measuring mental states alone. We will advance this argument most extensively in chapter 9 though we will briefly touch on it here.

Now let us consider some theories of welfare, and what the psychometric approach has to say about them.

3.8.1 Special problems of the eudemonic, objective list and related theories

Psychometrics isn't always of the psyche

Many eudemonic theories refer to things that are not clearly "in the head", or that involve an interaction between objective and subjective circumstances- for example a flourishing social existence. Prima facie, we might wonder whether psychometrics is up to the task of measuring things which are not in the head but rather contained in the subject's lifestyle. However, it is not uncommon for psychometrics to quantify such things, for example Harthorn and Griffith (2000) develop a scale measuring friendship. Thus, there is no special difficulty here.

You might also worry that the subject may have imperfect knowledge of their own life in this regard. I discuss the problem of imperfect knowledge further in section 3.8.3 where I argue that it is unlikely other methods can give us better access. Potential limitations in the subject's knowledge will be taken into account by any competent psychometrician designing a questionnaire, and if these limitations make the questionnaire as a whole very inaccurate, this should show up in reduced measures of validity. For example, reduced convergent validity, perhaps especially in relation to third person reports; and perhaps even in reduced reliability- if the subject is inaccurate in an inconsistent way. Additionally, it is important to keep the differences between knowledge of factors internal and external to the subject in context- we do not have perfect knowledge of the 'objective' features of our life, but there is little reason to think that we have perfect knowledge of the subjective features either! Introspection is not a perfect guide to questions like "how frequently am I worried", and there is no a priori reason to think we *must* be more accurate in our estimates of how worried we are, than in our estimates of how many close friends we have.

3.8.1.1 Evaluating the Eudaimonia and virtue of a life

Eudaimonia is a conception of well-being or the good life. Although its exact dimensions are often vague and difficult to demarcate, it typically aims at flourishing, often conceived of as the excellent exercise and development of human capacities. In classical formulations of this view, "virtue" is considered to be a part of flourishing. It can be superficially similar to the objective list view, according to which what makes a human life good is the achievement of a collection of objectively given goals- friendship, learning, goodness, pleasure or happiness, etc. However the additional element that distinguishes it from the objective list view is the suggestion that the relevant goals represent an end of the human species as such.

We might posit that there are two ways to create a eudemonic conception of happiness. The first way might be called the orthodox way. It takes Aristotle at his word about what eudaimonia is. Such an approach would, at least apparently, have to include an evaluation of the subject's *virtue* because for Aristotle there is a conceptual connection between being virtuous and wellbeing. A person who is apparently in a state of well-being, but who is vicious, isn't really happy.

This will create special philosophical difficulties. For many people, the idea that psychology can make evaluations about who is and isn't a good person will be a bridge too far. But these difficulties will be further magnified in the context of welfare economics. If the problem of values in welfare economics is bad now, it will be much worse if we start making evaluations of whether a new highway is likely to make citizens braver, or a more progressive income tax system will improve chastity.

There is a different approach possible, which is to simply ignore virtue, and focus on other elements of flourishing in Aristotle's theory. For example, in Ryff (2014) "Psychological Well-Being Revisited: Advances in the Science and Practice of Eudaimonia", a widely cited review paper on the topic, there is exactly one reference to virtue, which did not seem to be related to the measurements themselves. There were scattered references to specific virtues, but on the whole, virtue was not central to the conception of eudaimonia the authors were concerned with. At one point the authors list "theory-guided dimensions of wellbeing": *Autonomy*, *environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance*. While some, maybe even all, of these aspects are *linked* to virtue (primarily in being consequences of virtue), none of them are explicitly about such. Proctor and Tweed (2016) even refer to the "glaring absence" of virtue and related concepts in the area.

We will concede that a weakness of our psychometric approach is that it will have difficulty coping with any theory of happiness that posits the happy must be virtuous, at least if we are unwilling to judge virtue. But even Aristotle admitted that other things are necessary for happiness, and we could at least be of some use to advocates of this view of wellbeing by measuring the non-normative parts of what they consider happiness to be, which may be worthwhile information in the context of policy formation. There remains an interesting question of how much our differences over virtue matter in reality. Perhaps we could come to consensus measurements of, for example, courage and kindness although certain kinds of virtue (e.g., sexual and romantic) seem inherently political. It may be that normative disagreement is not an impassable barrier to the measurement of some virtues, at least in some respects.

3.8.2 Special problems of the capabilities approach

The capabilities approach could be seen as a eudemonic or objective list theory, or as outside the taxonomy we gave altogether and simply focusing on the preconditions for other kinds of welfare, but we will treat it separately due to its substantial development in the literature.

Robeyns & Byskov (2023) define the capabilities approach, as developed by Sen (1985) and Nussbaum(1992) as:

> a theoretical framework that entails two normative claims: first, the claim that the freedom to achieve well-being is of primary moral importance and, second, that well-being should be understood in terms of people's capabilities and functionings. (Robeyns & Byskov, 2023, para. 1)

Can psychometrics measure capabilities? What about non-mental functionings? The capabilities approach has always included variables that are best measured psychometrically. Consider for example Martha Nussbaum's list (2007) of fundamental capabilities: bodily health; bodily integrity; **senses**, **imagination and thought**; **emotions; practical reason; affiliation**; other species; **play**; **and control over one's environment**

The highlighted terms are those that seem like they would be especially easy to approach from a psychometric perspective.

The second thing to note about our approach is that while not all capabilities are mental, our approach in no way contradicts or excludes measuring capabilities separately or in addition to psychometric measurement to supplement our data.

Even measurement of non-psychometric variables will benefit from a psychometric evaluation, *especially if we are in practice reliant on self-report*. A non-mental feature of a person is not necessarily beyond measurement by the tools of psychometrics as we have already discussed.

3.8.3 Special problems in desire satisfaction accounts of well-being and their psychometric measurement

According to desire satisfaction accounts of well-being, your life is going well *to the extent that you have what you want*. There are many debates about the exact parameters of this theory. For example, do all types of wants count, or only certain types, perhaps about how your life as a whole is going? Is it what you actually want that counts, or is it what some informed version of yourself would want that counts instead? Crisp (2021) makes the interesting claim that the historical reason for the rise of desire theories was a result of the rise of new welfare economics- it is easier to measure the fulfillment of desire than it is to measure pleasure and pain:

> Pleasure and pain are inside people's heads, and also hard to measure—especially when we have to start weighing different people's experiences against one another. So, economists began to see people's well-being as consisting in the satisfaction of preferences or desires, the content of which could be revealed by the choices of their possessors. (Crisp, 2021, Section 4.2)

The theory is not a mental state account of subjective well-being insomuch as your desire can remain unsatisfied even while you think that it is satisfied. To use Nozick's (1974) example, we can imagine a machine which simulates an enjoyable life for a person, full of triumphs and achievements. On mental state views, such a person will have wellbeing, yet on a desire satisfaction account, the person lacks wellbeing, or certainly at least has less than they believe. It may be true that they *believe* that they have climbed Mount Everest, but they have not actually climbed mount Everest, thus, unbeknownst to them, their desire remains unsatisfied, though Chalmers (2022) offers a challenge to this interpretation.

We could create a version of the theory that was a mental state view as follows- what matters is your *belief* that your desire has been satisfied- regardless of whether or not it has been. Such a view would have few charms- it would lose one of the main advantages the desire theory has over the hedonic view, in that being in Nozick's experience machine would be "just as good as the real thing".

If that proviso is too strong, a more popular candidate for an addition to the theory is that you must have a true belief or awareness that your desire has been satisfied in order for it to count towards your wellbeing (Heathwood, 2006). We could call versions of the theory which have this proviso quasi-mental state desire theories.

As we previously mentioned there has been a view in the literature that has associated the desire satisfaction view of welfare with economics, and the hedonistic view of welfare with psychometrics. Debates about psychometrics in welfare economics then become, in part, debates about whether the desire satisfaction or hedonistic view of welfare is better. But this view seems odd, and insufficiently defended. At least in theory, it is hard to see why psychometrics could not assess desire satisfaction.

There is a practical difficulty though- there is no existing direct psychometric measure of desire satisfaction. As best I can tell, no one has got around to making one. Ideally, we would like to be able to say that the psychometric approach can handle all of the popular approaches to defining the good life, but in the absence of a *direct* psychometric measure of desire satisfaction, this is difficult. We will discuss this in greater depth in chapter 9.

Another objection- what about the possibility of the agent being wrong about their desire satisfaction a la the experience machine? Just because the agent *can* be wrong about the satisfaction of their desires, doesn't mean it is likely that they are. People *usually* know whether they have what they want- hence simply asking "are your desires satisfied" will suffice in most cases. Furthermore, if we are in an experience machine, it is not like any method apart from psychometrics is *better* placed to detect this!

There is a popular psychometric construct of life *satisfaction*. Life satisfaction is unlikely to correspond perfectly with the simplest desire satisfaction accounts. However, as Michael Plant (2020b) notes it may have a closer correspondence with the global desire satisfaction theory, where desires about how one's life should go overall are what count. I find Plant's argument persuasive in relation to global desire satisfaction accounts- it seems both intuitively and on the basis of the considerations he develops in the piece, that the degree to which I am overall satisfied with my life, and the degree to which my overall desires about how my life should go are satisfied, should fully or largely coincide. We might further wonder whether life satisfaction approximates the general- non-global- desire theory of wellbeing. It is plausible that people are satisfied with their lives to the degree their wants are satisfied in general, and thus we might find that life satisfaction was tracking desire satisfaction generally. This is an empirical question for which I would commend future research.

Future research could also attempt to create and validate a desire satisfaction questionnaire. Although I am loath to pre-empt empirical results, I see little reason to think this wouldn't work from a psychometric point of view. Why double up when we already have revealed preference approaches to measuring the same? For one thing, as we will discuss in Chapter 5, psychometrics brings with it certain tools for tackling interpersonal comparison. For another, as we will discuss in chapter nine, such questionnaires can get at what people *intrinsically* desire, and not merely what they *extrinsically* desire. People generally only extrinsically desire commodities, money and even time, yet economic measurement of wellbeing is focused on commodities, money and time because these are easiest to get at.

One interesting question then is: which will be a better guide to desire satisfaction? Two answers are possible:

1. People's own beliefs about the degree to which their personal desires are satisfied, or even whether they are satisfied with their life as a whole, or
2. the satisfaction (or not) of a subset of desires, most of them instrumental, that can be measured (at least with any facility) via willingness to pay and revealed preferences based approaches.

We will return to this in chapter 9, where we will argue that even though self-evaluation of desire fulfillment is imperfect, there are reasons to think that it is more accurate than using revealed preferences and similar as a proxy. Effectively in chapter 9, we will argue that if the desire satisfaction theory of well-being is true, our best way to measure well-being is to survey people about their degree of desire satisfaction.

As we have noted, the *externality* of desire satisfaction and the agent's *fallibility* with respect to it doesn't mean psychometrics has nothing to tell us about desire satisfaction. Indeed, if we drop the hidden premise that there is something epistemically special about knowledge gained through introspection, we see that the situation of using psychometrics to evaluate desire satisfaction is not so different to the situation of using psychometrics to measure anything. In all cases, the subject's knowledge is a limiting factor and there is, in principle, no reason to think that the subject knows their own conscientiousness or anger better than the subject knows their desire satisfaction, except one is in some sense partly external, and, barring extraordinary faith in introspection, it is not clear why this should be a decisive difference. Besides- the vast majority of psychometric questionnaires probably require us to go beyond introspection. Even filling out a questionnaire about a seemingly internal trait, like happiness, surely involves more than introspection, as presumably one must ground one's answers through comparison with others.

After all, in other senses, both anger and conscientiousness might be said to be partly 'external' to the agent. We might consider Larry constitutionally less angry than Lucy, even if he got angry more often if that was because his work involved customers constantly screaming at him, and Lucy's didn't. We might consider Lucy constitutionally more conscientious than Larry if she regularly turns down rich bribes, but he never receives them at all. This externality of these concepts has never prevented the psychometric measurement of anger or of conscientiousness.

Finally, I will note in passing that some discussions in the literature treat *life satisfactionism* as its own approach to welfare, wellbeing or happiness. For example, (Rossi & Tappolet, 2022) write:

> *An alternative theory of happiness that has made inroads amongst psychologists and economists over the past forty years, and that has eventually reached philosophers, is life satisfactionism. According to it, happiness consists in an attitude of satisfaction towards one's life.*

From here it would be a short step to make life satisfaction a theory of welfare - simply add the premise that wellbeing consists of happiness. Insofar as some people regard life satisfaction as in itself welfare it follows that life satisfaction, the psychometric construct, is presumably well placed to measure life satisfaction the state- at least if psychometrics can, in general, deliver on its methodological promises.

3.8.4 Special problems of estimating welfare and the "anti-theory" of welfare- Hausman, Agner and others

By the *anti-theory* of welfare I mean an idea, perhaps most explicit in Hausman (2011) and implicit in Adler & Posner (1999) that welfare has not yet been adequately defined, and may never be, but that we can trust that the things we think of as indicators of welfare, joy, fulfillment, a rich social life, physical health and so on are indeed *associated* with welfare, even if we don't know if they constitute welfare. This view could also be called *evidentialism* about welfare- we can know whether or not a person has welfare based on a number of signs, but we cannot know what welfare is. We previously alluded to a similar view in chapter 2 as “Question mark realism”.

I would argue that psychometrics is well placed to capture welfare on this (anti)theory of welfare. If we are unsure what welfare is, we would be well advised to consider *many* different possible indicators of welfare, since it will be difficult to be certain which indicators matter the most. Our project seems well equipped for this.

3.8.5 Correlation between types of subjective-wellbeing: Eudaimonia, hedonic & life satisfaction

Throughout this thesis, an empirical pillar of our work will be that the differences between these theories of welfare are less important than one might think in public policy because different concepts of welfare intercorrelate. There are important ways in which they diverge- for example, life satisfaction is known to be considerably more responsive to income than happiness (Diener et al., 2010) or positive affect (Kahneman & Deaton, 2010). Nonetheless, the intercorrelation of different welfare measures allows us to argue that information gained from measurements of one type of welfare are likely to be relevant to those who believe in another concept of welfare. Even information about how policy affects a different notion of welfare to the one you prefer is better than no information about how policy affects welfare at all.

Strong correlations between happiness and life satisfaction (Nemati & Mehdipour, 2016) are well attested in the literature. Correlations between happiness and life satisfaction on the one hand and eudaimonia on the other are somewhat weaker- Clark and Senik (2011) find excellent correlation between life satisfaction and happiness, but weaker correlations between these variables and various measures of eudaimonia, including questions assessing vitality and flourishing. Huta (2020) finds a clear distinction between momentary hedonia and eudemonia but that: “Eudaimonia and hedonia showed little or subtle distinctness when measured as global experiences, or situational experiences for a one-week time span.” Measures of eudaimonia are not as popular, well studied or standardised as happiness. It is important to remember that correlations at the level of groups are also likely (though not guaranteed) to be stronger than correlations at the level of individuals.

3.8.6 Experience sampling, the foibles of other methods and the need for further research

Experience sampling/Ecological momentary assessment is a methodology in which, at random intervals, the participant is asked to score their experiences, often on an unbounded scale. There is a variety of work that suggests experience sampling-derived well-being averages do not have an easy correspondence with scores derived using a more traditional method.

Some research by (Fredrickson & Kahneman, 1993) suggests that individuals when reporting on the positivity or negativity of their experiences display certain predictable biases. (Fredrickson & Kahneman, 1993) argues that individuals may arrive at their assessment by combining the *peak* or most intense moment with the *last* moments of the experience. Such an estimation suffers from many flaws- for example, it is invariant to the length of the experience. One can imagine that this might bias happiness measures. For example, a happiness estimate might consist in the subject’s experiences right now and/or right before taking the test, combined with their most intense experience recently.
”

If Kahneman and Fredrickson are correct that people’s evaluation of their own subjective well-being does not reflect an approximate average of their experience of well-being at each specific moment, this raises both empirical and axiological questions about the degree to which we can trust these ratings as an evaluation of overall subjective wellbeing.

Kahneman and Fredrickson (1993) examine h*edonic* wellbeing. There are interesting conceptual challenges in thinking about whether a methodology like this could be applied to life satisfaction, as while we expect overall hedonic subjective well-being to equal the average of momentary subjective wellbeing, we have no such expectation of a relationship between satisfaction with moments and satisfaction with life, or satisfaction with your life across all moments of life, and satisfaction with life now. Still, we can imagine something like the same bias affecting life satisfaction. It is an open question to what degree this would bring into question life satisfaction results, an open question which might require attention from psychologists and philosophers working in tandem. Much the same is true, perhaps even more so, of eudaimonia.

If experience sampling methodologies give us knowledge about subjective well-being that more static methodologies cannot, then a priority from the point of view of the project in welfare economics we outline in this thesis is developing cost-effective forms of experience sampling, so knowledge of the effects of economic policy on sampled subjective wellbeing can be expanded. Alternatively, it might be possible to create protocols for estimating how the “true” experience sampling-based results are expected to vary from wellbeing estimated using other methodologies.

Of course, all this assumes that experience sampling-based results represent the ‘true’ results. This assumption has been contested by no less an authority than Kahneman himself:

People do not want to be happy the way I’ve defined the term—what I experience here and now. In my view, it’s much more important for them to be satisfied, to experience life satisfaction, from the perspective of ‘What I remember’, of the story they tell about their lives.

(Kahneman in Haaretz 2018, quoted in (Frijters et al., 2021))

Future research could investigate the income responsiveness of experience sampled subjective well-being, corresponding to the various philosophical theories of welfare. A potential bridge between the intensive methods of experience sampling and more traditional approaches is day reconstruction e.g. (Kahneman et al., 2004) which has been found to have reasonably robust correlation with experience sampling methods (Dockray et al., 2010). Depending on the context the authors found: “The between-person correlations adjusted for attenuation ranged from 0.58 (stress, working day) to 0.90 (happiness, leisure day).” (Dockray et al., 2010)

3.9 Conclusions for chapter three

We have briefly reviewed foundational psychometric concepts such as reliability and the various forms of validity and considered a sample of philosophical, empirical, and methodological concerns and thinking on the topic of the psychometrics of wellbeing. Although we could not, and did not, set out to do justice to all of psychometrics critics and interrogators, we outlined our position that psychometrics as a practice can provide knowledge about the various forms of subjective wellbeing and, most importantly, all the major theories philosophers have put forward about the good life.

4 Subjective Well Being Effect Estimation

In this thesis, I defend the idea *that a desirable alternative to many methodologies in welfare economics, including traditional cost-benefit analysis, would be estimating the effects of policy on subjective-wellbeing- both its distribution and level.* I refer to this as Subjective-Wellbeing effect estimation (SWBEE).

4.1 The aims of SWBEE

The aims of SWBEE are to provide an approach that:

A) Is directly implementable in an applied welfare economics context. SWBEE is very much an approach in *applied* philosophy or, one might even say, philosophical engineering. We aim to make something as far as possible ready to go off the shelf.
B) Allows the welfare-economist to make a *useful* contribution to almost any policy discussion where welfare economics might be thought to have something useful to say.
C) Is not deeply ethically controversial in a way which will limit its usefulness due to ethical sectarianism.

4.2 The principles of SWBEE

The core principles of SWBEE are as follows:

1 The most important role of welfare economics should be to estimate the likely effects of policy on subjective wellbeing, keeping in mind the broad definition of subjective wellbeing we are using in this thesis. We are concerned with subjective well-being as a proxy for, or perhaps measure of, welfare- as opposed to other constructs often used in welfare economics. We believe that subjective well-being offers promising approaches to a variety of traditional problems in welfare economics.
2 Because there are many different kinds of subjective wellbeing construct, many related to popular views of the good life in philosophy (pleasure, life satisfaction, eudaimonia), the ideal would be to measure every type, so as to be maximally non-sectarian. In practice, we may be forced to compromise on this principle, to greater or lesser degrees.
3 “Effect” on wellbeing constructs here is understood not just as an estimate of the mean or aggregate level of a construct in a population, it is to be understood as a full study of all the effects of a policy on all parameters of wellbeing constructs- including not just the mean (and aggregate level, if possible), but also the distributive parameters across both persons in the present and over time.

4 This study should be conducted for the purpose of *describing* the effects of a policy on ethically relevant variables, not prescribing action as the result of some decision rule like utilitarianism. However, while the description does not pretend, in itself, to offer a direct solution to the problem of what should be done, it should be made with the explicit intention of informing discussion around policy choice.

These principles can be distilled into two underlying ideals:

A Our approach should be founded on providing empirical knowledge about subjective-wellbeing, in order to inform discussion about human welfare.
B Our approach should seek not to evaluate the political or moral merits of a policy but seek to provide information that would be of value to the evaluator(s).

One way to express this point is that we are looking for a principle that is as close to uncontroversial as the principle that Pareto improvements are good, but that will allow welfare economists to comment on a much broader range of topics and options than Pareto efficiency alone. We suggest two such principles- *the principle of subjective-wellbeing and welfare* - it's worth knowing things about a person's subjective-wellbeing if you want to know about their welfare, and *the principle of information*: It is a good thing to have information on the welfare relevant effects of policies. Although the principle will not allow the economist to prescribe policies in the same way the Pareto improvement principle does, it gives the economist a meaningful basis on which to investigate policies and feed into normative discussions about them.

4.3 The limited role of this thesis in relation to SWBEE

Although we have codified the basic ideas in different ways to other authors, and with a different emphasis, we did not invent the building blocks of SWBEE. Of course, we make original contributions in relation to it, for example, with regard to interpersonal comparison, and the role of values in science. Most important is how we combine these elements into a total approach.

A sampling of previous authors that feed into this tradition includes Ng (1997), the Leyden approach (van Praag & Frijters, 1999), the UK *Green book* (HM treasury 2022), Alexandrova and Haybron (2011), Layard, Mayraz and Nickell (2008)- *The Marginal Utility of Income* and all of the various efforts to incorporate subjective wellbeing into economic and development statistics ranging from the "Gross National Happiness" of Bhutan to the OECD's manual on the subject. Bronsteen et al.'s (2013) proposed analysis of aggregate levels of happiness- wellbeing analysis (WBA)- as an alternative to cost-benefit analysis. Although Angner (2013) is not a proponent of the view he must be credited as one of its pioneers in arguing that economics has unfairly dismissed psychometrics.

It will be our task in this thesis to argue that SWBEE is appealing as an approach because it offers attractive strategies for dealing with longstanding antinomies in welfare economics, viz:

the problem of interpersonal comparison, the problem of cardinality and the problem of the role of values in welfare economics. Although we do repeatedly argue in this thesis alternative approaches, including more traditional approaches, have difficulty dealing with these problems, our primary concern is only to establish that our method can deal with them. We are not in the business of trying to establish that SWBEE is the only workable approach to welfare economics, merely that it solves significant problems.

4.4 A brief sketch of what an SWBEE approach might look like in practice

There are many ways that the SWBEE methodology could be applied in practice, and these methodological details are not the focus of this thesis- our focus is on the philosophical portions of the machinery. However, for the sake of illustration, here is one model of how a SWBEE of a policy could be worked out.

On the assumption of constant income elasticity of the marginal utility of income, various estimates have been made of that elasticity (eta) (see Latty 2011 for a review). These include the estimates of the elasticity where utility is conceived of as subjective wellbeing (e.g., Layard et al. (2008), Clark et al. (2017)) . Such a value could be used to perform weighted cost-benefit analyses with weighting on the basis of prior income- using an assumption that receiving X for which one is willing to pay Y is equivalent in subjective-wellbeing terms to an increase in one's income equal to Y. Thus, in certain special cases SWBEE turns out to have a great deal in common with weighted cost-benefit analysis, although due to principle 3 there will be less of an emphasis on the total aggregate figure- since the rationale is not an estimation of whether the policy is right according to some ethical standard, but the provisioning of information about effects of a policy on subjective-wellbeing. Eta is also likely to vary somewhat by subjective-wellbeing type, so multiple analyses by form of subjective-wellbeing will be appropriate.

But this assumption could be dropped for parts of the analysis for which there is more specific existing research on subjective wellbeing effects. For example- in an analysis of the welfare effects of transport infrastructure, we could directly use in the model what is known about the effects of a reduced length of commuting times on SWB (in its various forms) rather than willingness to pay for reduced commuting times.

We can also imagine situations where we have some non-specific evidence about the effects of a policy of a certain type and this non-specific evidence might be incorporated as an adjustment to willingness to pay. Suppose, for example, that we had evidence that recreational amenities typically had one and a half times the impact on subjective well-being than would be predicted on the basis of people's (income weighted) willingness to pay alone. We could thus adjust willingness to pay, to better predict subjective wellbeing. Of course, whether effects take exactly this convenient form (multipliers of income weighted willingness to pay set neatly by the kind of intervention) is dubious. However, there is a clear strategy here for getting at the psychological effects of policy- investigate the relationship between various kinds of interventions, willingness to pay and subjective well-being.

4.5 A clarificatory contrast between our approach and a utilitarian SWF approach

Consider, for example, a hedonic utilitarian who accepts as a measure of pleasure some SWB metric. Their preferred approach to social policy might well look superficially similar to ours. Like us, they may use a variant of weighted cost-benefit analysis to estimate total effects, even the weights will be the same. However, their project varies from ours in two ways.

The first is rationale, the utilitarian picks the weights for their weighted cost-benefit analysis in order to estimate effects on happiness, and so might we. However, they do so *because they think a cost-benefit analysis so weighted will point them to the right policy conclusion*. We do so simply because we think it is good to know the effects of a policy on wellbeing. The second distinction is what we intend to happen with the results. For the utilitarian, in the ideal case, a well conducted and wholly accurate weighted cost-benefit analysis, accurately capturing all implications of a policy for their preferred wellbeing metric indicates, of itself, the correct decision. They plan to use the overall sum of (weighted) benefits minus costs as a guide for action. For us the figures generated, both disaggregated and aggregated (weighted) benefits and costs, are simply grist for policy conversation.

The distinction will seem to some very fine- after all, identical methods for estimating happiness are used, but it is the difference between assuming, say, utilitarianism is right in the construction of our weights and methodology, versus merely assuming that the public want to know about how a policy will affect this or that subjective-wellbeing metric.

4.6 A note on terminology: “Cost-benefit Analysis”

In many cases, our preferred approach will have a great deal in common with weighted cost-benefit analysis. It can be applied to willingness to pay and modify it on the basis of existing income and what the money is being spent on. In other cases- for example, where we have direct information about the happiness effects of an intervention, and we thus disregard willingness to pay, the procedure we outlined has little in common with cost-benefit analysis.

In those first cases, where we do use willingness to pay, our SWBEE procedure can be seen as a variant of cost-benefit analysis, or, perhaps it is more accurate to say, it can be seen as a variant of weighted cost-benefit analysis *and more importantly, a different rationale for weighted cost-benefit analysis*- a rationale without a need for a Social-Welfare function.

Despite this convergence between our method and cost-benefit analysis, where we use the phrase "cost-benefit analysis" in this thesis -whether weighted or unweighted-, we generally mean to refer to more traditional modes of cost-benefit analysis rather than SWBEE. Any exceptions to this rule will be clear from the context.

There is another key difference between our method and weighted cost-benefit analysis which we have already mentioned, but which it is important to emphasise again. Because our approach aims to provide knowledge on the *distribution* as well as *average* of wellbeing, there will be less emphasis on the total aggregate figure than there would be in weighted cost-benefit analysis. To treat estimated aggregate wellbeing as the primary output of the analysis would be to come too close to endorsing utilitarianism, distribution matters as well.

4.7 Many subjective well-being constructs, one subjective-wellbeing construct, or many but with varying degrees of fidelity?

As we emphasized earlier, the ideal of our approach is to estimate the effects of policy on many subjective-wellbeing constructs. In practice, this ideal might be difficult or resource-intensive to meet, and despite the compromise to the ethical neutrality our approach is supposed to represent, it may be necessary to home in on a single subjective-wellbeing construct.

It is important not to overestimate the degree to which this compromises our ethical neutrality. As we have already discussed there are correlations between different kinds of subjective wellbeing. Generally speaking (though by no means always!), correlations will be even larger when we correlate the averages of populations, rather than the scores of individuals, and the scores of groups (whether income levels or neighbourhoods) are really what we are most concerned with in this inquiry. This dampens the necessity of contests between the advocates of different approaches to subjective wellbeing. Suppose, for example, that you are a partisan of life satisfaction. You care not one whit for whether people are happy or not, only if they are living the kind of life that they most want to live. Even given this outlook, you should still be glad for the information a model on the hedonic effects a policy provides, because such a model will estimate the life satisfaction effects of a policy *especially if your alternative is a more standard approach from welfare economics, such as a willingness to pay approach.* Even so, we acknowledge that there will be divergences, and these may well be large enough to have practical implications, thus it is important, in avoiding sectarianism, to make a wide variety of estimates across multiple forms of SWB *wherever possible*.

Experimentation will be needed in order to determine the practical difficulty of the kinds of modelling we are proposing, including the additional costs of modelling the effects of policy on not just a single wellbeing variable, but many well-being variables. In this thesis, we have largely helped ourselves to the assumption of plentiful resources, in order to focus our attention on theoretical difficulties, but implementing the method will likely involve exploring various kinds of tradeoffs.

4.8 The problem of ‘open requests’

Sometimes the economist might be asked, simply, to ‘find the best way to do x’ rather than being given a menu of policies to evaluate. SWBEE seems to provide little guidance about what to do in such cases. There are options for the extension of the method we describe to such cases, albeit at the cost of sacrificing some of the virtues of our approach. One could simply seek to maximise a preferred social-welfare function defined over a numeraire of subjective-wellbeing, a hybrid of the psychometric and SWF approaches.

Another possibility is to generate a range of ‘plausible’ value functions (we hesitate to call these value functions social welfare functions, given the commitments that implies in this context, though they would be welfarist) and work out a few well supported proposals for each. Will such an approach achieve what SWBEE sets out to achieve- providing information to a variety of decisionmakers and the general public without undue ethical sectarianism?

This will depend on two questions, both open to empirical and theoretical study:

1.The degree to which the range of value functions chosen by the analysts captures the range of values held by the relevant public and decisionmakers.
2.The degree of sensitivity of the preferred policy to the exact parameters of the value function adopted.

We will discuss this as an area for further research in the final chapter.

4.9 Conclusions for chapter four

We have outlined Subjective-Wellbeing Effect Estimation (SWBEE), an approach to welfare economics which aims to estimate the effects of a proposed policy on the distribution of subjective-wellbeing. Such an approach aims to provide ethically relevant *information* for the public and decisionmakers without engaging in ethical decision-making or privileging one or another set of values. Multiple estimates, made across different forms of subjective-wellbeing, are desirable where resources allow, as not everyone agrees on which forms of subjective-wellbeing provide the most direct evidence about the degree to which people are living good lives. Under some circumstances SWBEE can be evaluated directly through knowledge of the welfare effects of similar past interventions. In other circumstances where less is known it might resemble weighted cost-benefit analysis, weighted so as to try and anticipate the subjective-wellbeing effects of policies, rather than suggesting the correct action from the point of view of a social-welfare function.

5 The problem of interpersonal comparison

> *The major problem on this view of desires as theoretical entities is that, with the possible exception of descriptive decision theories that provide the support for individual utility measurement, there exists very little that would count as any kind of precise, empirically supported theory involving desires. This is not to say, however, that there is nothing that at least approximates such a theory. Much of the work in psychology on motivation could well be taken as preparing the ground for such a theory. Despite the fact that a satisfactory theory has not been developed, there seems no reason to think that it could not be developed on the basis of further research. We may even think of our everyday beliefs concerning desires as giving a rough outline of the kinds of connections between desires and observables that such a theory might justify.*

-Waldner (1972)

5.1 Introduction

This chapter is concerned with the interpersonal comparison problem. In this thesis, the interpersonal comparison problem is distinguished from the interpersonal aggregation problem as follows. The interpersonal aggregation problem is the problem of weighing up benefits and losses running in different directions for different people, varying across a menu of policies, and either saying which policy should be preferred, or at least saying something of substantial interest to the question of which should be preferred. The interpersonal comparison problem is the methodological problem of how to measure a psychological state in one person and compare it to the psychological state of another person- whether that state is happiness, fulfillment, anger, hunger, fatigue or anything else. This way of framing the problem is related to, but distinct from, Drakopoulos' (1989, p. 2) framing, that the problem of the interpersonal comparison of utility is: "The comparison of subjective sensations in general."

Our concept is even still broader because we include other kinds of mental states apart from sensation- for example, motivational states, cognitive states, mixed states like attitudes and even long-term mental states, like dimensions of personality. It will also necessarily include states like *desiring* and *preferring* which presumably form the underlying psychological basis of action and choice.

As stated, the problem of interpersonal comparison is really many separate problems. While generally discussion of interpersonal comparison has been restricted to discussions of welfare and/or utility, there is no reason we might not talk equally of the problem of interpersonal comparison of anger, hope, extroversion, or any other mental state or trait that exists along a continuum. These problems may or may not be closely related- It is very possible that a solution to the problem of how to compare the intensity of wanting will not necessarily entail a solution to the problem of how to compare the intensity of anger, for example. Other subproblems are arguably solved or have plausible candidate solution- partisans of the subjective interpretation of probability have given us an approach for comparing degrees of belief between persons through the willingness to accept different gambling odds- see Hajek's summary (2023). While the ambit of what we are describing may seem impossibly broad, nonetheless we will try to propose a fully general method for such comparisons- or rather show that one has already existed for many decades in the form of psychometrics.

According to many approaches to policy choice, the solution to the interpersonal aggregation problem will depend on the solution to the interpersonal comparison problem, but this does not need to be true. These problems are logically distinct. Simply knowing the intensities of one person's mental states relative to another person's mental states does not, in and of itself, imply the right way to aggregate gains and losses between persons, nor need even be necessary to do so. The easiest way to see this is to consider it in terms of Hume's guillotine- there is no necessary connection between the *is* of the sum of feelings and the *ought* of the normative value of one situation vis a vis another. The thought that these problems nonetheless bear a tight relation is very old, going back at least to classical utilitarianism, and as we will later discuss in chapter 7, no one or almost no one thinks that this information about relative magnitudes of mental states would *hurt* to have in deciding between policies.

There's an ambiguity in the language of interpersonal comparison generally. It's sometimes not clear whether the difficulty is in comparing two people's psychological states (which happen to be valued psychological states), or in comparing the value that accrues to two people's psychological states. In this chapter, our primary concern *unless otherwise specified* is with the comparison of the magnitude of mental states, in particular, mental states of *subjective well-being* such as happiness and satisfaction. We take this to be more like comparing the temperature of two objects than comparing the aesthetic value of two objects. Any implications for valuation are downstream. Of course, mental states can tell us a great deal about welfare on many theories of welfare. Welfare itself is an inextricably value-laden concept, so the information we gather about mental states might inform us about matters of value. *However, this information on mental states is not itself a value judgment*. Let us consider some examples of theories of welfare, and what knowledge of- and the ability to compare- mental states between persons would allow us to know on these theories.

1. The hedonic theory of welfare: You have welfare to the degree that you have a balance of pleasure over pain. In this case, a knowledge of mental states will quite straightforwardly translate to a knowledge of welfare, and the ability to compare mental states will translate into the ability to compare welfare between persons.
2. Desire satisfaction theories of welfare: You have welfare to the degree you have satisfied desires. Although the possibility of a mismatch between what the subject has

and what the subject thinks they have is always possible as with any desire satisfaction theory, it seems that people's feelings on the satisfaction of their desires should be a reasonable approximation, and thus are reasonable proxies for welfare on the desire satisfaction theory (for further discussion see chapter 9).

3. Eudemonic accounts of welfare hold that what it means for a human life to go well is for that life to flourish. There are almost as many ways of spelling out what this means as there are philosophers who endorse this theory. However, most eudemonic theories of welfare include a variety of criteria that either *are* mental or will correlate with certain mental attributes reliably. Consider, for example, (Waterman et al., (2010))'s scale of eudaimonia, which includes questions about self vs other directed motivation ("It is more important that I really enjoy what I do than that other people are impressed by it.") and intensity of concentration on and investment in tasks ("I find I get intensely involved in many of the things I do each day").
4. Objective list theories in some ways can be seen as a residual category. They hold that what it means for a person's life to go well for them is for them to achieve certain outcomes, with the exact nature of those outcomes varying from theorist to theorist. We would need further detail on the relevant objective list theory to comment, but many plausible theories will either include mental states as part of the list or include things very well correlated with mental states. Often pleasure and desire satisfaction are part of the list.
5. Suppose you took what we will call an *anti-theory* view of welfare. Welfare was something of which you might say "I know it when I see it". However, you hold that specific attempts to define it, let alone quantify it, are subject to counterexamples- either necessarily, or at least with respect to all theories we have created so far. This approach seems likely, in practice, to closely track a variety of mental states, happiness, satisfaction, tranquillity, a feeling of companionship and so on- all of these tend to make us think that a person's life is going well, even if we don't have a specific theory of what that means.

In some of the above theories, the ambiguity between comparing values and mental states becomes purely semantic since an answer to one question will be an answer to the other- for example in the hedonic theory. In other theories, there is a gap between "true" fulfillment and mental states- for example, in the case of desire fulfillment theories, a person may think that their desires have been fulfilled, when they have not. In other cases, mental states will only partially constitute value- e.g., in eudemonic theories. Thus, *some sort of* subjective-wellbeing is almost always related to wellbeing on any theory, but the degree and nature of the relation will vary.

In this chapter, we generally take it for granted that welfare-relevant states like desire have an experiential component. I would note that, if this is not true, and even beings without conscious experiences of desire can have desires that matter from the point of view of welfare, then our argumentative task would become easier, not harder.

5.2 A note on the ontological problem of comparison

In this chapter we consider interpersonal comparison of psychological states primarily as an *epistemological problem, the* problem of how we could know that a comparison is true. But while this is one sort of problem you might have about interpersonal comparison; it is not the only possible problem. You might also be worried that interpersonal comparison is a lot like trying to compare a length and a temperature- the quantities are simply incommensurable. Call this the ontological problem of interpersonal comparison.

Exactly who the burden of proof falls on in this case is somewhat unclear. Should we tend to think that comparison is inherently impossible unless we have evidence to the contrary, or the opposite?

This argument might have special relevance to the desire satisfaction account. Yes, desires are mental states, but perhaps there is something tricky about conceiving what it would mean for me to want a lemonade more than you want a cola. If desire is a relationship to the world, perhaps the meaning of comparing my way of relating to the world and yours in quantitative terms is incommensurable in some way.

One of the classical defences of interpersonal comparison is a comparison to intertemporal comparison within a person e.g., (Gibbard, 1986; Sidgwick, 1874; Harsanyi, 1955; Briggs, 2015). This seems like it might be an apt response here. Intrapersonal, intertemporal comparison seems to be possible. But I at one moment in time might be very different from myself in the future. If there is no further fact about "me-ness" that unites me together, just sequences of causation and similarity (denying what Parfit calls the further fact view (Parfit, 1984)) how to explain why I can compare my own experiences, but not those of other people? The simplest answer might be to concede that interpersonal comparison is possible.

There are two potential replies to this argument:

The first is to insist that there are insurmountable disanalogies between intertemporal intrapersonal comparison and interpersonal comparison that make comparisons meaningful in one case, and like comparing temperature to length in the other. We might deny the "no further fact" view about personal identity over time and give some account of why this special continuity gives special access *and* meaning to comparison. I'll leave assessment of the strength of this argument to the reader and to further debate in the literature.

The second is to bite the bullet and accept that intertemporal comparison of an intrapersonal nature is impossible or merely reflects value judgments. We will not discuss this further, except to note that it is a cost, and some reason would have to be given as to why we should be willing to pay it.

We note this problem here, and some of the preliminary lines of argument, for completeness, but will not attempt a full treatment here.

5.3 The psychometric solution to the problem of interpersonal comparison

We will now attempt to show that psychometrics gives us a method to interpersonally compare subjective wellbeing, and thus even welfare, on certain theories of welfare. The structure of our argument is:

Psychometric validation premise: Psychometric validation shows that equivalent scores on validated psychometric instruments tend, on average, to represent similar functional states, in terms of their causes, effects, and biological correlates. If we both score X, then our functional state is, at least in expectation and in relation to the named construct (e.g., happiness, satisfaction), similar. Greater scores represent greater degrees of a specific functional aspect or state.

Weak functionalist premise: It is reasonable to assume similar functional states in relation to a given psychometric construct like happiness or satisfaction represent similar experiential states in relation to that construct, whereas greater functional states in relation to a construct represent more intense experiential states in relation to that construct. Functional states are defined by both their outputs and inputs. Note that this premise is not just a *general* assumption of functionalism, it is an assumption of functionalism about specific psychometric constructs, e.g., sadness. Note also that although the conceptual vocabulary is different, it bears analogy to Harsyani (1955, p. 317): “If two objects or human beings show similar behaviour in all their relevant aspects open to observation, the assumption of some unobservable hidden difference between them must be regarded as a completely gratuitous hypothesis and one contrary to sound scientific method. (This principle may be called the "principle of unwarranted differentiation)”

Weak functionalist supplementary premise: If we know the strength of a functional state, and on that basis we are entitled to assume we know the strength of the experiential state, it is reasonable to hold that we know the strength of the state as a whole- either because the only important features of states are *functional* and/or experiential *or* because the functional and experiential aspects of a mental state are an acceptable basis on which to make assumptions on the other aspects of the mental state.

Conclusion: Therefore, psychometrics provides us with reasonable grounds to engage in interpersonal comparison of welfare, on a certain class of theories about what welfare is.

We do not say that welfare, measured in some way other than the psychometric, cannot be interpersonally compared (or that it can be). Our only purpose is to show that there is a neat, conceptually well-founded way of comparing subjective-wellbeing and similar mental states and that this possibility allows a practical method for meaningful interpersonal comparison in welfare economics.

Chapter 3 gives the broad reasons one might think the first premise is true. We defend the second premise in what follows. I first explain functionalism in the philosophy of mind, then lay out why the weak functionalist premise is more modest, and hence appealing than functionalism tout court. We then show that we can weaken the weak functionalist premise even further, to an assumption only about group averages, without losing the conclusion, with reference to Lerner's equal ignorance argument.

The main predecessor of our approach to interpersonal comparison is Waldner (1972). For Waldner, talk of desires is talk of theoretical entities which are used to explain a wealth of behaviours- not just market transactions, but everything from facial expressions to daydreaming to choice latency. Waldner's discussion is complex, contains many caveats, and is framed in the language of an approach to the philosophy of science that is now largely extinct. A reasonable gloss is that Waldner suggests that work on the psychology of motivation, examining choice in the light of a range of variables from probability to latency, may lay the groundwork for a future quantitative framework for understanding concepts like desire strength. For Waldner, this would enable interpersonal comparison. Waldner acknowledges that a holdout might still try to insist that certain individuals have a special "sensitivity" that is not captured by the model (say, all their desires may be twice as intense, thus cancelling out), but argues that rejecting the existence of this special "sensitivity" is not a mere value judgment, but simply a rejection of special pleading in considering what entities to posit for explanatory premises and after all *Pluralitas non est ponenda sine neccesitate*. We should reject the excessive complexity as easily as we would reject the idea that Neptune might be twice as strongly affected by gravity as everything else, yet a variety of special factors ally to prevent us from noticing. Thus, Waldner rejects the view that interpersonal comparisons of desire cannot be empirically grounded. Against a critic who responded that while such a quantitative and precise science of motivation could in principle exist, but does not yet exist, Waldner (1972, p. 103)could argue:

> *Even without the establishment of such a precise theory our common beliefs about the connections of strengths of desires to other factors may be thought of as forming the vague outlines of such a theory and as having at least some empirical support. Insofar as this is true, we may be able to use such common beliefs to make interpersonal comparisons of at least a rough nature, and such rough comparisons may suffice for many of the applications for which they are needed.*

In some ways, the approach we outline in this chapter can be seen as an update on Waldner after half a century's empirical and conceptual progress, bringing in ideas that were in a relatively early stage of development in Waldner's day such as functionalism in the philosophy of mind and the psychometrics of subjective wellbeing in psychology.

Let us now turn to the explication of the two key premises one by one:

1 *Psychometric validation premise:* Psychometric validation shows that equivalent scores on validated psychometric instruments tend, on average, to represent similar functional states, in terms of their causes, effects, and biological correlates. If we both score X, then our functional state is, at least in expectation and in relation to the named construct (e.g., happiness, satisfaction), similar. Greater scores represent greater degrees of a specific functional aspect or state.

To give an example, your seven and my seven, on the happiness ladder, will, so long as the measure is well validated, on average, represent similar functional states. A functional state is a complex of situation, behaviours, behavioural tendencies, and biological correlates of a sort that might align with a construct- whether that be happiness, desire, anger, belief, or whatever else.

In chapter 3 we reviewed both the methodological basis and empirical basis, for the claim that psychometrics allows us to measure constructs- behaviourally defined dispositions of the subject- using their scores on various measures. The best way to make sense of this is if a similar score generally implies a similar functional state. If you thought the fact that we both scored seven on the anger questionnaire implied no similarity between my (functionally defined) anger and yours, you would be in a difficult situation from an explanatory point of view. You would have to explain how these measures -filled with questions about anger- can predict constellations of things that seem conceptually related to functioning like an angry person. Not a handful of predictions either, but an array of biological, situational and behavioural predictions.

So, assuming the empirical and methodological bases we talked about in chapter 3 are correct, psychometrics can provide us with evidence about the degree to which people possess certain functionally defined states and where they fit on spectra of functionally defined mental attributes. But why should we think that similarity of function implies similarity in underlying mental state?

2 *Weak functionalist premise: Weak functionalist premise:* It is reasonable to assume similar functional states in relation to a given psychometric construct like happiness or satisfaction represent similar experiential states in relation to that construct, whereas greater functional states in relation to a construct represent more intense experiential states in relation to that construct. Functional states are defined by both their outputs and inputs. Note that this premise is not just a *general* assumption of functionalism, it is an assumption of functionalism about specific psychometric constructs, e.g., sadness.

In chapter 3 we sketched out the conceptual grounds, and gave some examples of the empirical evidence, for the claim that psychometrics can create adequate measures of subjective wellbeing *as a functional concept,* involving complex interrelations between behaviours, environments and even biology*.* But there is a long history of dualism in western thought about the mind, and so the fear arises, what if, notwithstanding our behavioural similarity, your joys are twice as high and your pains are twice as deep as mine, in a mental realm not accessible to techniques of behavioural or biological analysis?

I want to deal with this concern at the level of methodology by putting forward a premise I call weak (or if you prefer "merely epistemic") functionalism, but first we must outline functionalism.

5.4 An introduction to functionalism

There are many different varieties of functionalism, but when we talk about functionalism without further clarification in this thesis, we have in mind David Lewis's Analytic Functionalism (e.g., Lewis (1966)) as updated by Braddon-Mitchell and Jackson (1999) where it would make a difference. For our purposes, the exact doctrinal features will matter little, beyond what is already spelt out in our weak functionalist premise, but for the interested reader, we give Phelan and Buckwalter's (2012) excellent brief of Analytic Functionalism: "[...]people ordinarily define mental states relative to the causal roles these states occupy in relation to environmental impingements, external behaviours, and other mental states."

Functionalism, in this sense, is a philosophical theory of mind according to which a mental state is whatever plays the functional roles that categorize that state. For example, suppose that the firing of C-Fibers was caused by damage to the body, tended to make people yelp, and led people to avoid the circumstances that had caused their firing in the future. On a simplified functionalist account, this would suggest that the firing of C-Fibers is pain because the firing of C-fibres does what pain is supposed to do. A more sophisticated functionalist account would hold instead that pain is a functional state of the body and nervous system as a whole.

Functionalism vs behaviourism

Functionalism is importantly different from behaviourism. Behaviourism identifies mental states with actual behaviours, and with dispositions to behaviours. This has a number of odd consequences- for example, it implies that being angry does not *cause* us to yell, rather our being angry is constituted, in part, by our tendency to yell. Functionalism keep's behaviourism's tight link between behaviour and mental states but allows us to preserve intuitions such as the causation of behaviour by mental states. Our tendency to yell is not anger, rather (approximately) whatever state-of-affairs within our central nervous system that causes us to yell is our anger.

Functionalism vs identity theory

Functionalism can also be contrasted with the identity theory of mental states, another physicalist theory. The identity theory e.g. (Lewis, 1966; Smart, 1959) claims that mental states are identical to specific brain states. Functionalism agrees, and depending on subtleties of philosophical terminology, can be seen more as a particular version of the identity theory rather than a rival. What basic identity theory doesn't include is a response to the problem of multiple realisability. Surely, we can imagine that, for example, an alien could, in principle, be very psychologically similar to us, even while its underlying physiology was very different.

Functionalism has an answer to this- many vastly different arrangements can fulfil the same function. It is not so clear that other forms of the identity theory have an answer to this question- they may have the resources to find one, but this is a matter for the philosophy of mind literature. It is important to note that the exact border between functionalism and the identity theory of mind is subject to both philosophical and classificatory debate, and what we have given here is only a brief overview of one way of looking at the distinction.

It may be useful to consider functionalism as a kind of synthesis, or midpoint, between the classical identity theory and the behaviourist theory, that aims to preserve many of the advantages of both. For example:

1. The tight conceptual link between behaviour and mental states that behaviourism provides.
2. The causation of behaviour by mental events that the identity theory provides.

Functionalism and the impossibility of P-Zombies, inverted spectra etc.

A philosophical zombie (Kirk, 2023; Kirk & Squires, 1974) is a benighted creature. Though its behaviour is exactly that of a regular human, it has no internal experiences. It perceives, reasons, and wants in a sense, but it never encounters the redness of red.

The brother of the philosophical zombie is the inverted spectra person (Byrne, 2020; Shoemaker, 1982). This person experiences red as green, blue as yellow and so on. Because they have experienced the world this way since childhood, they have no frame of reference to realize that their experience is different from the experiences of others. Because their behaviour is absolutely identical, other people have no way to realize it either.

Both of these cases are given as arguments against functionalism- and physicalism generally. Whether or not such people actually exist it seems that they *could* and, since they could exist (whether they do or don't) this undercuts the conceptual identification between functional states and mental states.

At base, the worry that your desires and emotions could be vastly stronger than mine despite our behaviour being identical is founded on a similar premise to the inverted spectra and philosophical zombie arguments. The argument that your desires might be vastly more intense than mine, despite all your behaviour being identical, is the economist's equivalent of these arguments in the context of welfare economics. The same animating idea is there- *that the quality of an experience- whether a phenomenology of colour or the intensity of a craving, can be divorced from actual behaviour, and even from counterfactual behavioural dispositions or functionalism.* What we think of as mental state is really composed of two separate bits that can split-phenomenological state and functional state. Functionalism contradicts this idea, and functionalists generally hold that zombies and people with inverted spectra are impossible, although there are debates about whether that impossibility is contingent or necessary.

Functionalism and Harsyani's principle of no unwarranted differentiation

One way to integrate our discussion of functionalism into the previous literature on interpersonal comparison is as follows. Harsanyi (1955) develops what he calls the principle of no unwarranted differentiation, described earlier in this chapter, which has often been used to defend interpersonal comparison.

Functionalism entails the metaphysical claim that this is not merely a plausible or even a probable principle as regards human behaviour, rather it is impossible for it to be any other way, perhaps it is even conceptually necessary (c.f. *Analytical functionalism* or the view that functionalism is a conceptual truth. e.g., Phelan & Buckwalter 2012. If functionalism is a conceptual truth, then depending on one's preferred account of concept and necessity, this may well make it a necessary truth).

The weak functionalist premise is weaker than functionalism proper

With respect to *weak functionalism*, it seems to me reasonable to infer mental state from functional state in the absence of any contrary evidence, if we accept physicalism about the mind as most do and given the breadth of 'functionality'.

As the chapter proceeds, we will try to weaken the weak functionalist assumption further. However, it is always possible that the reader will find no persuasive force in this assumption, and for such a reader we must retreat to Harsyani's principle of no unwarranted differentiation as a bare claim. Even for such a reader though, there is still value in this chapter, as we show how Harsyani's principle interacts with psychometrics to allow practical interpersonal comparison of a variety of welfare related constructs.

Our premise is that: "It is reasonable to assume similar functional states represent similar experiential states." This premise is weaker than functionalism generally, so even if you do not accept functionalism, you may have reasons to accept it. Hence, we call it *weak functionalism.* I suspect that almost all philosophers in this sense are weak functionalists- even if they think that p-Zombies and inverted spectra could exist in principle, they do not believe them very likely, thus, at least in practice they hold it is reasonable to infer from functional similarity to phenomenological similarity.

Weak functionalism is epistemic rather than ontic

The first reason our theory is weaker than normal functionalism (and hence called weak functionalism) is that it is an epistemic rather than an ontic theory. It does not assert that identical functional states equals identical mental states because functional states just are mental states, rather it merely asserts that *it is reasonable to believe that experiential states are similar where functional states are identical*.

This has several helpful consequences. For example, whereas the ordinary functionalist (probably) needs us to believe that P-Zombies and inverted spectra are impossible, we simply require our interlocutor to accept that they are not actual, or rather that they are *likely* not actual.

Weak functionalism depends on similarity rather than identity

Relatedly, the weak functionalist premise does not assert that the same functioning gives us grounds to believe that mental states are *identical* but merely that they are *similar.* If you wanted to hold that there were likely to be slight phenomenological differences between persons with functionally similar behaviour that would not hurt the project of broad interpersonal comparability we are aiming at, so we have no need to rule it out.

It allows the incorporation of biological measures, which would be irrelevant to many forms of functionalism

Suppose as we do, we had good reason to think that cortisol is a stress hormone. Suppose that Bob and Peter had identical levels of cortisol, that would be evidence of phenomenological similarity on our account, and more importantly, some degree of evidence that a measure which correlated with cortisol was a valid indicator of phenomenological similarity.

In our concept of functional similarity, we have included similarity in biological factors. This is a departure from ordinary functionalism, insomuch as ordinary functionalism is a theory of *how biology comes to constitute mental states* biology won't play this evidential role*.* This is a second way in which our conception of functionalism is more expansive, and less epistemically demanding in this context than other forms of functionalism, insomuch as we permit this further additional source of evidence, although it may be that there are variant kinds of functionalism proper which would allow for this role.

5.4.1 The argument extended- the principle of ignorance

Consider Lerner's theorem. Let us suppose that we know the following things about a population- that their utility in some good is concave and strictly positive, that the variance in utility under any distribution is finite, and nothing else. From this Lerner (1944) has shown that the best way to maximise utility in expectation is an equal distribution.

We might say that considerations *along the lines of* Lerner's theorem give us an additional reason to accept in practice epistemic functionalism, or perhaps they are simply a more explicit way of spelling out the reasons we already had to accept it. The justification is simply that, in the absence of information to the contrary, with nothing to break the symmetry, A is as likely as B to have the more intense desires. We might see this as an expression of the principle of indifference in probability theory. The principle of indifference in probability theory holds that when you have no information distinguishing several options, you should treat them all as equally likely (Eva, 2019).

Suppose there were two boxes before you, both contained sand, and you knew that one contained more sand than the other. Beyond that you had absolutely no information whatsoever about which contained more sand. If you were to decide, on a purely arbitrary basis, that you thought the left box contained more sand, many would think you irrational.

There are paradoxes with respect to the principle of indifference as a general principle involving ambiguities with respect to the exact variable we are indifferent to (e.g. Bertrand's paradox of the cube (Eva 2019). These debates are ongoing and will likely remain so for some time, However, regardless of whether the principle of indifference is a guide *in every case*, in this *particular* case it seems like a reasonable principle not to assume that one individual's feelings are more intense than another's without reason.

As an aside, Lerner's appeal to concavity perhaps gives us a further, pragmatic reason to favour the principle of no unwarranted differentiation- insomuch as it is likely to imply unequal treatment, in expectation, it will decrease utility- for further discussion see (Bennett, 1981; Pivato, 2013; A. K. Sen, 1973).

Can we apply the principle of no unwarranted differentiation/equal ignorance to psychometrics? Yes. Since we have no information indicating that your feelings if you score a certain way are greater than mine if I score the same way, or vice-versa, they are both equally likely, and hence the principle applies.

5.4.2 Equal Ignorance and psychometrics

Our method can plug a hole in Harsanyi's unwarranted differentiation approach. One of the major weaknesses in reasoning from unwarranted differentiation is that there are different subgroups in the population that plausibly have different intensities of feeling towards money, thus, feasibly, there could be many *warranted* forms of differentiation. There are many reasons some people become richer than others, but it is surely not implausible that a greater love of money contributes in some part. Or, if we are talking about subjective wellbeing, we may think that certain subgroups feel happiness more intensely than others, in a way which is not necessarily reflected in their ratings e.g., if they called a feeling that would be ecstatic happiness for anyone else a 7, because it is just ordinary for them, or if they used smaller numbers to avoid an impression of boastfulness, as it is sometimes suggested may be common in collectivist cultures (e.g. Tov & Nai (2018), but see also our discussion on cultural differences in the section on psychometrics). Given these possibilities, even though it may be fine, in theory, to talk about not making unwarranted differentiations, actual comparisons may be hampered by fear of the many warranted differentiations that are possible.

This is where an argument from functional equivalence- demonstrated through psychometrics- and an argument from equal ignorance can intersect and mutually support each other. Psychometrics allows us to investigate whether there are any functional differences. We can study whether equivalent answers are associated with different criterion measure levels in different populations. For example, are equivalent responses associated with different cortisol levels, or different frequencies of Duchenne smiles, among the poor and rich?

Having established that there are no detectable functional differences between subgroups - or found the differences and adequately controlled for them- we can then apply the principle of no unwarranted differentiation (or, we might just as well say, the principle of equal ignorance). The appeal to equal ignorance is thereby bolstered. Not only do we not have any reason to suspect your ratings reflect a stronger phenomenology of desire or affect than mine, but we have also been unable to detect behavioural differences empirically between subgroups (or we have detected them and then adjusted for them). So even if we thought we had some a priori reason to suspect that the love of money or the happiness of rich people is more intense, we've either done the empirical investigation to show that this is not the case, or done the empirical investigation, quantified it and controlled for it based on functional differences.

What I have outlined here- checking that scales mean similar things to different groups as a prelude to applying the equal ignorance argument over phenomenological states might seem like a distant methodological possibility. Yet empirical work examining similarities and differences in how various demographics use scales already exists. Existing work has examined scale usage, often focusing on culture as a source of differences. For example a large literature has investigated inter-cultural variation is attitudes towards reporting high SWB values, with evidence suggesting that is some cultures, notably collectivist cultures in East Asia, this is considered improperly boastful, and in others, as associated with personal success and virtue (Eckersley, 2013; Kitayama et al., 1997, 2006; Lau et al., 2005; Lu & Gilmour, 2004, 2006; Y.-K. Ng, 2002; Oishi, 2002; Safdar et al., 2009; Suh et al., 1998; Uchida et al., 2004). There has also been work investigating procedures via which these differences in scale use could be corrected (Angelini et al., 2014; Harris et al., 2020; He et al., 2017; J. W. Lee et al., 2002; Roster et al., 2006). Work has also touched on gender- for example Batz-Barbarich et al. (2018) briefly discuss the idea that it is more acceptable for women than men to express extreme emotions, and how this may bias results.

An excellent further step in the program we outline in this thesis would be to review existing literature on differences in scale use and create a manual to inform the policymaker of which groupings require adjustments if we are to best estimate the effects of policy on their subjective-wellbeing, how to perform these adjustments, risks inherent in the practice of adjustment, and how large distortions in unadjusted scores are likely to be.

The hypothesis that we should assume there are “extra-functional” differences that this method cannot find or account for- that we must then factor into weighting, placing, for example, twice the weight on changes in the measured mood of blue eyed people- is even more speculative than normal attempts to dodge the principle of equal ignorance because we have given evidence that there are no behavioural differences. It may be plausible that the rich people love money more, but it is far less plausible that they love money more in a way which is undetectable in their behaviour (or physiology).

Here is another way to make the same argument:

Earlier we spoke of the weak functionalist postulate. Suppose you rejected that postulate because you found it objectionably strong. We can propose a new ultra-weak functionalist postulate:

A person's functional state is an unbiased predictor of phenomenological state across demographic groups and phenomenological states have finite variance

The premise is weaker insomuch as it allows individuals to vary drastically, so long as for economically relevant subgroups- like income groups- on average the same functional states equal the same total mental and thus phenomenological states. *Or at least that it is reasonable to assume such, in practices like welfare economics*.If this is true, the functional state is still our best guess as to the phenomenological state.

We include the assumption of finite variance in imitation of Lerner. It is possible that it is not necessary, but consideration of the infinite case is beyond this thesis, both because of the kind of treatment it would require and because of our practical concerns.

5.5 The shades of progressively weaker arguments you could use to reach our conclusion

Let us consider a progression of ever weaker claims we can use to support the idea that psychometrics allows interpersonal comparison.

We first argue that if functionalism is true about a mental state, psychometrics provides us with a means to interpersonal comparison of that mental state.

We can then weaken the argument further (though we skip this step in the preceding discussion) to:

> *As a matter of contingent fact, functions are well correlated with phenomenological states*

Call this moderate functionalism. If psychometrics enables interpersonal comparison given functionalism, psychometrics also enables interpersonal comparison given moderate functionalism.

From here we can weaken our functionalism still further, to weak functionalism:

> *It is reasonable to believe, in the absence of contrary evidence, that functions are well correlated with phenomenological states*

From here we can weaken the functionalism still further to what we call the ultra-weak functionalist postulate:

> *A person's functional state is an unbiased predictor of phenomenological state across demographic groups and phenomenological states have finite variance*

And finally, we can give what we might call the ne plus ultra weak functionalist postulate:

> *A person's functional state is a predictor of phenomenological state with no known biases in any particular direction, and equally likely to be biased in one direction as another. Additionally, phenomenological states have finite variance.*

5.6 Conclusions for chapter five

We have argued that if we accept the accuracy of psychometrics, plus an assumption that it is reasonable to believe that alike functional states correspond to similar experiential states, the problem of interpersonal comparison is solved. Then we have gone on to show how our approach may complement the ideas of Harsyani and Lerner on interpersonal comparison, and how our assumptions can be further weakened.

6 The problem of cardinality

6.1 Introduction

Economists have often worried about cardinality in welfare economics, with classical concerns about the topic going back to foundational authors in new welfare economics, especially Pareto (2018/1906) and Robbins (1932). The problem of cardinality is deeply associated in the literature with the problem of interpersonal comparison, perhaps due to Robbins (1932) work which tackled both.

Many have argued that the problem is soluble within classical welfare economics, c.f. Andreas (2011) for a thesis length defence. Others are sceptical, arguing that while an analysis of preferences under risk or delayed utility might give a cardinality of a kind, such a cardinality is not to be equated with a true cardinal utility function. See for example the "The Cardinal utility which is ordinal" by Baumol (1958). One notable problem in this regard is that as we have already discussed there are multiple methods for cardinalizing welfare, and there is no guarantee they will lead to similar results.

Our interest in this thesis is in whether the problem is soluble in the context of psychometric studies of subjective wellbeing because the same problem arises there as well. Consider, for example, Jenny scores a 7/10 on her happiness test. Bob scores an 8/10 on his. Xiao scores a 9/10. Is the gap in underlying happiness between Jenny and Bob's score the same as the gap between Bob and Xiao's score? Bigger? Smaller? By how much or how little? It is possible, for example, that as one approaches either end of the scale, the gap between each answer gets larger and larger, or conversely smaller and smaller. Kristoffersen (2015a) suggests that the three main possibilities are that the relationship between happiness scores and true underlying happiness could be logistic, linear or logit, but in principle any shape imaginable is possible.

Unless we know the answers to these questions there are a number of things we cannot do. For example, we cannot take the mean of Bob, Jenny and Xiao's score, and if we can't take means, many types of overall assessment of the effects of a policy are impossible. These include not just utilitarianism but approaches that rely on more complex Social-Welfare functions, like prioritarianism.

This should not be seen as a problem for studies of subjective wellbeing alone. As Plant (2020a) notes, it also applies to: "health, hotel quality, employee satisfaction, etc." Just as we worry about the relationship between a 7, 8 and 9 in the context of wellbeing, so we could just as easily worry about it in any of these contexts. However, the problem may be less deadly in these areas, because the theoretician there has a manoeuvre that is not available to us- pragmatism. Consider a respondent's ratings about their level of anger. So long as these ratings had predictive power when treated as cardinal, allowing the scientist, say, to predict violent behaviour, all might be well. In other words, the researcher can treat cardinalisation *pragmatically*.

Such sanguinity is not available to the researcher who wants to study happiness *at least if they want to rest any ethical weight on their findings*. Predictive power alone is not enough, and that is all a pragmatic interpretation would afford. We must have reasons to think that our results tell us something about the real distribution of happiness (as we already discussed in chapter 3 this gives us an interesting vantage point in relation to questions of scientific realism- a reason to think that a certain kind of reliance on science in this particular area only makes sense if we presuppose realism.)

Psychology has long been fairly open to an assumption of cardinality, whereas economics has generally preferred to treat scales as ordinal. Why such a disciplinary mismatch? We might speculate that it has a lot to do with the history of ordinalism as a movement in welfare economics, a movement which saw interpersonal comparison and cardinal utility as a bundle package and rejected both with enthusiasm. Most notably Robbins (1932) project is clearly invested in both rejections. We might wonder if part of the disparity doesn't arise from the different approaches to mathematics in the two disciplines.

Part of the reason for the mismatch may be that theories in psychology, as opposed to econometrics, often generate predictions about the direction of effect only. Subjective wellbeing variables tend to occur in the context of these theories which only aim to predict the direction of effect, which may account for why there has been little need to establish "true" cardinality.

6.1.1 Introducing the epistemic/Bayesian approach to cardinality

I concur with the new Bayesian epistemic approach to cardinality championed by Philippi (2021) which sees debates over cardinality - for example, debates over the cardinality of Likert scales- as debates over uncertainty in the Bayesian sense. Assuming the absence of "metaphysical ordinality", where there really is not a fact of the matter about the distance of things on the scale (and we will briefly discuss our reasons for rejecting this later) we posit that there is some underlying "real" gap between each number on the scale. However, our estimate of what the gap is between each number, and the degree and shape of our uncertainty over that estimate, can vary. As our estimates of the distance between each number get more and more similar to each other, and as our uncertainty over those estimates shrinks, the scale becomes more and more "interval". A scale will be acceptably interval relative to a certain context if our estimates of the distance between each item on the scale are very similar, and the uncertainty on those estimates is low enough- that for some purpose we feel confident enough to perform certain kinds of mathematical operations on it and trust the results of those operations enough to use them. If the interval is one end of spectrum, total uncertainty of anything except the order is the end of the spectrum, although as Philippi (2021) notes, on closer examination, it becomes clear that it is hard to give a coherent account of what is usually conceived of as ordinality within this Bayesian framework.

Ultimately, all that will matter if our concern is extraction of something like averages will be the degree of uncertainty in our estimates, not the equality of the size of our estimated gaps. Suppose we were very certain of the size of the gaps, but they were unequal, we could simply produce modified scores which reflected what the scales would look like were the gaps equal. E.g., suppose we had a scale that went 1, 2, 3, 4 and we were convinced the gap between 2 & 3 were twice the size of the other gaps, we would simply make the scale 1, 2, 4, 5 before performing arithmetic operations

6.1.2 Our goals in this section

Issues of history and disciplinary methodology aside, in this discussion, I wish to convince the reader chiefly of three things:

1. There are reasons to think that subjective-wellbeing scales may already have an approximately linear relationship with underlying happiness, and so it may already be appropriate to treat data gained from these scales as cardinal.
2. Sensitivity analysis reveals that this uncertainty over the relationship between the scale and the construct may matter less than one might think, for purposes such as determining and comparing the average happiness of groups.
3. There are promising directions for further research on this subject to resolve cardinality.

6.1.3 Four kinds of objections to the notion that happiness scales are cardinal

We might organize objections to treating SWB data as cardinal, on a scale from most stringent to most permissive like so:

1. SWB isn't something that can be ordered by degree- even ordinal.
2. SWB isn't interval, the underlying thing in the world is itself ordinal.
3. Respondents don't have any kind of first-person access to interval strength information about how much SWB they have.
4. Respondents do have such access, but current questionnaires won't draw it out.

There is a 5th kind of objection, related to what we called Mach's problem in the chapter on psychometrics. According to this objection happiness is cardinal *but it is cardinal in multiple different ways*. For one purpose or construction, the gap between 6 and 7 is twice as large as the gap between 8 and 9, for another purpose, it is the other way round, and there is no fact of the matter as to which is correct- both cardinalisations are true in different contexts- just as Mach thought was true of temperature. We will not be discussing this further in this thesis, but commend it as a problem for further research.

6.1.4 An assessing of objections to the cardinality of subjective wellbeing scales

With respect to objection 1, any evidence that measures of happiness have predictive power, with changes in the results of measurement associated with changes in other variables is evidence that happiness scales capture at least ordinal data. Thus, the whole literature of validation studies on happiness instruments- from health, to income, to cortisol- is evidence against the first objection.

With respect to objection 2, The question of whether a mental state like happiness could practicably be inherently ordinal is interesting in and of itself, and many of the solutions we discuss to the cardinality problem do not have much to say on this matter. Five points are worth raising against objection 2, viz:

A) Running the brain in this way, with an indefinitely large number of possible, ordered, but not interval states of happiness seems like it would be far more complex than the alternative. This perhaps gives us some reason to doubt it.

B) There have been prior- and prima facie successful- attempts to empirically create laws quantitatively and apparently cardinally interrelating certain mental states, including expectation, degree of desire and degree of feeling measured using visual analogue scores (Price et al., (1985)) and one very natural explanation of the success of these experiments would be that the underlying states are indeed quantifiable, though naturally these mental states are somewhat different to subjective-wellbeing, they seem closely related. Further research is needed.

C) Participants are often asked to give their subjective wellbeing integrated over a period of time- often “in general”- presumably over the lifespan, or some recent portion thereof, or perhaps an indeterminately longish period. Participants are able to answer such questions with little trouble, low non-response rate and short response times (OECD, 2013). Assuming that what they are doing is not arbitrary, and assuming that they are not simply picking median or mode happiness, or similar, the possibility of this procedure indicates that the underlying values at different moments might be more than ordinal (though c.f. Redelmeier, Katz and Kahneman’s 2003 Colonoscopy experiment).

D) In this thesis we have been primarily discussing subjective wellbeing, however let us turn our attention to the closely related topic of motivational utilities for a moment, which we might think plausibly will give us some indication of the structure of subjective wellbeing as well. Thinking through real world cases, it is difficult to imagine how an underlying ordinal utility function could be behaviour guiding. Von Neumann and Morgenstern (1944) already showed that a cardinal utility appears to be necessary for decisions under risk, but if motivation has a cardinal structure, surely this gives us at least some reason to think that so might feeling, given their close connections.

E) Ng (2015) presents evidence using the example of being thrown into a bath of boiling water that people at least *believe* in cardinality:

> *Consider the following three simple alternatives faced by a person: A: Her current situation. B: Her current situation plus being bitten by an ant (non-poisonous one) once. C: Her current situation plus being thrown bodily into a pool of boiling water. Obviously, she prefers A to B and B to C. If preference/utility is purely ordinal, this is all she can say. However, even you, not being her, know that the intensity of her preference of B over C is at least many thousand times larger than that of A over B. Moreover, you may also be confident that the intensity of her preference of B over C is at least many thousand times larger than that of your preference of A over B (interpreting A and B as applied to you).*

With the exception of the brief discussion above, we will mostly take it for granted that emotions are in reality cardinal, and that issues with our measurements are *epistemological* problems with capturing this cardinality correctly rather than *ontological problems* of trying to apply a kind of quantification which simply does not exist in nature.

With respect to problems 3 & 4. Problem 3 is obviously a more fundamental problem than problem 4. Some of our solutions are potentially solutions to both problems, while others only address problem 4.

Note that we do not need participants to have perfect access to the interval position of their happiness, approximate access will do.

6.2 Solutions to the problems of cardinality

6.2.1 Solutions to the fourth problem of cardinality: constructing cardinal measures of subjective wellbeing

We will now consider solutions to the fourth objection to cardinality: *Respondents do have such access, but current questionnaires won't draw it out.*

6.2.1.1 Transforming scores into utilities

One option, which I have developed in personal conversation with Kieran Latty (2020) is to transform scales into utilities using a method similar to Von Neumann-Morgenstern (1944). Asking respondents which trade-offs they would accept of various happiness scores at various rates of risk. Would you prefer, for example:

1. A 100% chance of a level of happiness that you would rate as a 7,
2. Or a 50% chance of a level of happiness you would rate as a 6, and a 50% chance of a level of happiness you would rate as an 8?

If there are concerns about risk, the cardinalisation can be done in other ways, for example, by observation of behaviour in Nash bargaining games over happiness levels, or by time tradeoffs, e.g. would you prefer to feel like a 7 for one week and a 9 for one week, or like an eight for both weeks.

We could also use extended sympathy to have participants make evaluations about, e.g., hypothetical grandchildren. “Would you rather one grandchild feel like a 9 most of the time, and the other a 6, or one grandchild feel like a 7 most of the time and the other an 8?” It is possible also to use a society wide model and have participants pick a society with different SWB profiles from behind a veil of ignorance, on the understanding that they were equally likely to have the SWB profile of any person in the society they selected from behind that veil. Unfortunately, these proposals involving extended sympathy are unlikely to work, due to the complicating effects of inequality aversion.

A major problem here is that, as experience will attest, there is considerable evidence that human decision making is not always rational. Further, if it is rational, people may not be trying to maximize what we would expect them to. For example, people might pick the profile that offers the greatest variance, on the understanding that what they want from life is not the highest overall level of wellbeing, but variability- to experience the highs and lows. Likely, cardinalized decision utilities over subjective-wellbeing cannot be equated with subjective-wellbeing, although perhaps they provide evidence about it. One could even argue that, for ethical reasons, *we should* cardinalise the wellbeing scale according to people's preferences, for this is what has ethical weight.

Thus, we come to a difficult impasse regarding cardinalisation using decision utilities. Not everyone will care for utilities over subjective-wellbeing states. Either for scientific reasons, or for ethical reasons people may think that it is important to know true underlying happiness (or life satisfaction, or whatever else). Only if respondents are at least approximately utilitarian (regarding their own welfare) will utilities over subjective-wellbeing and levels of subjective-wellbeing coincide. Nonetheless, A) their level of utilitarianism may be close enough B) We may be able to estimate their level of personal-utilitarianism separately C) We may think their decision utilities over wellbeing have their own value or matter intrinsically.

One exotic option would be to map out decision utilities controlling for the ethical preferences of respondents. Then, relying on the decision utilities of utilitarians, on the basis that their decision utilities will reflect a desire to maximize happiness because they are utilitarian, work out a mapping between true underlying wellbeing and response options on a scale. As conceptually tidy as this strategy is, it would probably involve too many steps and inferences to be trusted, and would rely on the assumption that utilitarians didn't use the scale in a different, perhaps more linear, way. We would also have to discriminate between different types of utilitarians on the basis of which wellbeing concept they aim to maximise.

6.2.1.2 Experience sampling

Kahneman (e.g. (Kahneman, 1984/2000)) using a methodology of *experience sampling* first developed by Prescott and Csikszentmihalyi (1981) asks respondents at different moments how they feel, with zero being neutral, and possible answers being unbounded in both directions. They are told to think in cardinal terms- e.g., a 40 should represent an experience twice as good as a 20. Since the moments taken are random, the average score produced can be seen as a person's average level of utility. These scales are *unbounded* making them more likely to be linear.

If you have no worries whatsoever about the extent to which people have self-knowledge of cardinal subjective wellbeing, but you are worried that existing scales may distort the expression of this knowledge through scale compression, experience sampling provides a natural solution.

By mapping average utility levels produced by this method with responses on a standard test of subjective wellbeing, we can make the test of subjective well-being into a cardinal measure. The only barrier is that we need to put each respondent's experience sampling answers onto the same scale as each other respondents. There are several ways this could be done e.g., seeing what scores multiple people assign to comparable experiences, perhaps averaged across multiple such experiences. We could also use our old methodological ally, construct validation to assist along this process, correlating behavioural and biological measures etc., as we do with the task of putting Likert scales in interpersonal alignment.

6.2.1.3 Investigation of the use of existing scales to reveal their implicit cardinalisation.

Ng (1996) has an admirably direct proposal- simply ask respondents what they think the size of the gap between the various responses is. The question is surprisingly difficult to formulate and comes out much longer than most questions of the sort we typically see in questionnaires, although Ng does state that "Despite its length and complication, in my experience, respondents had no difficulty comprehending and answering it." Here are part of the instructions that respondents were given while being asked about their experiences:

> *Now estimate the amount of your positive happiness and negative happiness (i.e. unhappiness) over an average week over the last 12 months in terms of the number of times this happiness and unhappiness is equivalent to the happiness you derive from eating or drinking (one serve of) the something you mentioned in Question 6 (called "that item" in the table below). In estimating the happiness and unhappiness of an average*

week, take account of unusually happy/unhappy days as well as a normal week, i.e., the average is the average figure over the last 12 months."

Niko Tiliopoulos (personal communication, 2021) has impressed on me that an immense, almost insurmountable sort of difficulty in this regard would be persuading respondents that they are not being tricked, that they should not overthink the question but should try to understand it etc.

Another potentially promising approach owes to Galenter and Plinter (see below) who analogise the difference of income to differences in subjective loudness (which is thought to be logarithmic in real volume). They had participants compare the loudness of sounds to the utility changes involved in different gains and losses of money. It should be possible to perform similar experiments, but for differences in happiness scales, as opposed to differences in the value of money. While the existence of this paper and its authorship by Galenter and Plinter was relayed to me in conversation, I was not able to find it despite a long search and so cannot verify its existence. However, it would certainly be possible to try its methods. We might also try asking participants to place the items of a visual analogue scale.

6.2.1.4 Reasons to think respondents are already thinking in linear terms

An argument can be made that the numerical scale itself implies at least an interval level of measurement, so if I say "rate yourself on this scale from 0 to 10, where zero is the most unhappy you can imagine being and 10 is the happiest you can imagine being" we should assume that the distance between each number is equal for our participants *because that's what the setup on the question implies*. 8 is the same distance from 9 as 6 is from 7, mathematically speaking, so in order to be truthful to that scale, participants must interpret the distances linearly.

One approach here might be to "lock this in" by building it very specifically into the wording of the questions.

Kristoffersen (2015) points to useful research on perceptions of cardinality.

> *As mentioned, numeric measurement scales convey some intention of cardinality, and research into the perceptions of these survey instruments for the measurement of subjective wellbeing have revealed that people interpret them as cardinal and intend to provide responses that reflect this as accurately as possible (Van Praag 1991; Parducci 1995; Schwartz 1995).*

Plant (2020) makes the point that it is not even clear how participants would understand the scale if they saw it as purely ordinal- as an 8 not meaning anything specific except more than a 7 and so on, so the participants must envisage some kind of specific graduation between the points. Plant suggests that, in the absence of more specific information, the Schelling point (natural point adopted by participants who cannot communicate) is to treat 0 as the lowest practically possible level of happiness, 10 as the highest practically possible level of happiness, and each gap between the numbers between these extremes as equal. Unfortunately, Plant's speculations rely on- admittedly plausible- conjectures about how participants think. "Think out loud" protocols could be a way to access more specific information regarding this.

One of Plant's other suggestions is that individuals might interpret the scale with 10 as the happiest they have ever been, and 0 as the unhappiest they have ever been. Plant says that this would make scales cardinal but non-interpersonally comparable, but given some plausible assumptions, comparability might remain even on this interpretation *at the aggregate* level. However, there could be factors causing this to break down- e.g., differences in recreational drug use and differences in the rate of traumatic events at a population level.

Plant notes a number of studies that support this view of how scales are interpreted. Oswald (2008) finds that participants' descriptions of their own height on a 1 to 10 scale correlate 0.8 with their actual height- even more encouragingly, the correlation was effectively linear. Van Praag (1991) explicitly asks participants to place various evaluative verbal labels on a line with a cardinal scale between 1 and 1000 and finds that participants tend to impose approximate equidistance between descriptions. Another finding by Krueger and Schkade (2008) is that the change in happiness from week to week is homoscedastic, meaning that differences are about the same regardless of where one is on the happiness scale. Kruger and Schake argue that we would expect happiness itself to vary about the same amount wherever one was on the scale, at least approximately, if the scale is linear, therefore we have evidence of at least approximate linearity. Finally, a study by Prati & Senik (2020) finds a remarkable match between participants' description of their own life trajectory of happiness in the present, and data taken at past times, suggesting, as Plant argues, that participants most likely use the same scale over time.

Plant gives a series of conditions which, if met, will result in a properly cardinalized scale. But we could take the proposal in another direction- rather than trying to establish that participants are already following Plant's requirements, why not simply instruct them to do so? For example, we could label 10 "The happiest a person can feasibly be on an ongoing basis" and 1 "the saddest a person can feasibly be on an ongoing basis". We can say "The gaps between each number are the same size". After the participants have filled out the questionnaire, we can give them questions to test their understanding. We can then compare the answered questionnaires to other, similar questionnaires without such elaborate instructions. and see if the distribution is significantly altered. Any differences could be used to revalue tests given without such elaborate instructions.

Like all the solutions we have considered thus far, this approach is limited in that it only grapples with objection four. Now we will review some solutions that potentially grapple with both three and four.

6.2.2 Potential solutions also to the third problems of cardinality: the problem of first-person access to interval scale information

The third objection is: *Respondents don't have any kind of first-person access to interval strength information about how much SWB they have.*

6.2.2.1 Just noticeable differences

A just-noticeable difference is the smallest change in stimuli that an individual can notice.

Ng, (Y.-K. Ng, 1975, 1984a, 1985, 1996) following Edgeworth (1881) and Armstrong (1951) makes another proposal- that we may assume the magnitude represented by a just-noticeable difference is the same for each person and at each point on the scale, and from this, derive an analysis of the cardinality of the happiness scale (and also interpersonal comparison). The primary practical difficulty is that Ng's approach will require an empirical effort to map how many increments there are between any two points on the scale.

We previously discussed Ng (1996) in relation to simply asking participants about their understanding of the underlying scale, but his proposal goes much further than that. Ng suggests tying just noticeable differences in subjective wellbeing (in particular, happiness) to differences in a stimulus. Ng imagines, for example, two ice-creams, alike in nutritional qualities but different in flavours, with one superior. Through graduations in the proportions of the recipe we uncover how many steps there are from one ice-cream to another. Ng's (1996) next step is quite ingenious, although we may have concerns about how psychologically feasible it is:

> *Several methods of such indirect measurement may be possible. One is using magnitude measurement mentioned above. For example, if one would like to know how happy or unhappy a person is on a particular day or on an average day, he may be asked the following question: Write down a number representing the ratio of the amount of your happiness on that day to the amount of extra happiness you have in eating one serve (precise quantity as used in the measurement mentioned in the preceding paragraph) of Ice-cream A over Ice-cream B under condition X. Since we already know the amount of this extra amount of happiness in terms of just noticeable units of pleasure, we may then calculate his happiness on that day in the same terms by simple multiplication.* (Ng, 1996, p. 8)

Not only does this deliver us cardinality and interpersonal comparison, but it also even delivers us a ratio scale!

There are, however, grounds for concern about how plausible the assumption that each just noticeable difference is the same magnitude. It seems quite possible that the minimum jump in the intensity of an experience necessary to make you notice that jump will change, as the intensity of the experience itself changes. But this raises interesting conceptual questions. Is it even meaningful to say that someone is experiencing a greater amount of happiness at one moment than another, if they can't tell the difference, or is there something inherent to the phenomenology of a subjective state like happiness that *requires us to notice*? Or be able to notice, at least in principle?

Ng goes on to present the results of a very complex survey of happiness involving university students. Many of the questions are inherently complex, and many build on preceding questions. We will not discuss these results in detail, except to note that in order for the methodologies Ng describes here to receive widespread acceptance, either the questions asked would need to be greatly reduced in complexity, or very substantial work would need to be done proving that the respondents fully understood the questions.

In sum, while Ng's method based on just-noticeable differences is not fully developed, it represents at least a promising avenue for future research.

6.2.2.2 Does sensitivity analysis show us the problem is overstated

Sensitivity analysis has tended to suggest that it doesn't really matter if we assume ordinality or cardinality. Kristoffersen (2015) notes:

“So far, the most salient such justification appears to be that estimates of models for ordered discrete and continuous data tend to be highly consistent.”

One very simple approach is to explore variations from the assumption that there is a linear correlation between happiness and happiness score and see to what degree this distorts apparent results. Several empirical papers have found similar results with both ordinal regression and OLS (Ferrer-i-Carbonell & Frijters, 2004; Kingdon & Knight, 2007; Knight et al., 2009), In some cases the same results are found despite different cardinalisations (Ferrer-i-Carbonell & Frijters, 2004; Layard et al., 2008; Vendrik & Woltjer, 2007).

6.2.2.3 linear relationships with independent variables

A paper by Kaiser and Oswald (2022) titled “The Scientific Value of Numerical Measures of Human Feelings” found evidence that Likert scales essentially measuring domain satisfaction in areas like career correlate with behavioural measures- specifically behavioural measures related to exit behaviours. This paper sparked a heated reaction online because many thought that this was already settled science and wondered why it had been published in the Proceedings of the National Academy of Sciences.

One of the findings of the paper was that Likert-type scale answers or, as the paper called them, “feelings-integers” correlated linearly with exit behaviours. Johannes Haushofer (2022, communication on Twitter), a professor of economics mounted a defence of the paper on the grounds that its real value was showing linearity. He argued demonstrating the linearity of correlations with behavioural outcomes was vital to showing the linearity of the scale.

Such an argument has been made before, and the typical response is that it is possible for Likert scales to have a linear relationship with dependent variables, but for happiness not to be linearly related to its Likert scale measurement, if the non-linearity of the relationship between happiness and its measurement, and happiness and the dependent variable, cancel each other out in the right way. But would Haushofer be wrong to reject this as special pleading? Isn’t the simplest explanation for a broadly linear relationships between the scale and things we would expect to be correlated with subjective-wellbeing that the scale is linear, rather than fussing about with deviations from linearity at both ends that cancel each other out?

But this paper by Kaiser and Oswald (2022) was certainly not the only paper to establish linear relationships, or approximate linear relationships of variables dependent on SWB with Likert scale measurements of satisfaction, happiness, and other welfare related attitudes. See for example (Kristoffersen, 2015b) in discussion of the relationship between mental health variables and SWB variables. Review of research showing a *linear* relationship between happiness and related variables might then offer at least tentative evidence for the linearity of the SWB scale.

It is difficult to assess how much evidence for the linearity hypothesis linear relationships between dependent variables and Likert scales amount to. A formal framework for tackling this problem might be possible using Bayesian, maximum likelihood and/or Akaike information criterion statistics.

6.3 A brief note on Bond & Lang and the degree of variance at the ends of the scale

Bond and Lang (2014) in “The Sad Truth About Happiness Scales” have an objection to attempts to rescue cardinality- they are concerned about the top and bottom end of scales. Scores have a maximum and a minimum at both ends of the scale. We have no way of knowing how much variance exists at these ends of the scale and given assumptions of sufficiently large amounts of variance it is possible to reverse many important results from the study of happiness. My view is that in order to get this result, Bond and Lang have to make unfeasibly large assumptions about how big scale-end variance might plausibly get.

A quote to illustrate this: “In Appendix A we further show that nearly every result can be reversed by a lognormal transformation that is no more skewed than the wealth distribution of the United States”- as if the United States wealth skew were a moderate thing! I think that it’s pretty implausible that an evolved aspect of human psychology- presumably grounded in human biology and evolutionary need- is as skewed as wealth distribution in the United States- it’s hard to see what purpose this would play, given that happiness is, after all, a mechanism for regulating behaviour towards fitness (Cabanac, 1992; Y.-K. Ng, 1995; Rolls, 2007). As of the time of writing, Jeff Bezos is nearly one and a half million times wealthier than the median American. Discounting results because we cannot rule out the possibility that happiness is this right-skewed seems misguided.

Similarly, where they write: “To be clear, we are not proposing that satisfying this minimal criterion would make a result convincingly robust. It is plausible that happiness is more skewed than wealth is. And it is certainly not self-evident that happiness must be normal or lognormal. Happiness could be left-skewed for men and right-skewed for women, and their distributions might come from different families.” We must pause and ask for a modicum of psychological plausibility. A simple assumption that there is not more variance at the ends of the scale than in the whole in between will do away with Bond and Lang’s argument.

Perhaps their concern seems more plausible to them, in part, because the United States General Social Survey, upon which they base much of their analysis, uses a scale with only three possible answers- which makes it entirely possible that most of the variance is really at the extremes! Fortunately, this is not standard practice with most happiness questionnaires, the OECD (2013) recommends 11-point scales.

Despite my reservations about the strength of Bond and Lang’s conclusions, I recommend further research to establish the degree of variance at the scale ends.

6.4 A Positive objection to the claim that there is a linear relationship between subjective-well-being scales and real subjective wellbeing

Wodak (2019) gives an argument that we should not expect happiness scales to be linear on the basis of analogy with loudness scales. There is a certain circularity to Wodak's argument about loudness. He argues as follows:

1. When participants report loudness in numbers, the numbers they give have a logarithmic relationship with the actual, physical loudness.
2. Therefore, why not expect reported happiness to have a logarithmic relationship with actual happiness?

But just because physical loudness has a nonlinear relationship with reported loudness, doesn't mean that subjective loudness has a non-linear relationship with reported loudness. In effect, Wodak engages in circular reasoning and thereby begs the question by assuming that the nonlinearity shows up in the step between the experienced loudness and the reported loudness, rather than in the step between the physical loudness and the experienced loudness.

Arguably if this argument worked, it would tend to show that the relationship between happiness scales and happiness is likely to be logarithmic. If this is the case then A) we can easily adjust for this in computing averages and B) even if we do not know this, the logarithmic form of the relationship alone may not disrupt the ranking of groups by happiness very much.

6.5. A note on the informational bases of social choice

[refer back to table 1]

Our approach, as we will discuss in the next chapter, is to try to be "all things to all men" and provide information to everyone, regardless of their ethical outlook.

However, most ethical views have stronger informational requirements than that supplied by interval information about welfare. Yet in this chapter, we have only discussed establishing *interval* dinality. Information of only interval strength creates severe limitations with respect to various ethical views. Aggregate utilitarianism will break down with only interval information in situations with the potential to increase or decrease the population. Prioritarianism, where extra importance is placed on the worst off, typically requires a ratio scale (Adler, 2012; Brown, 2007). Sufficientarianism, a theory according to which additional weight is to be given to those below a critical value (Brown, 2005; Crisp, 2003a, 2003b; Frankfurt, 1987) will create a ratio scale of a sort, through the selection of that critical value (Adler, 2012).

Thus, even if we are successful in this chapter at showing that subjective-wellbeing scales can be safely treated as interval scales, we will not thereby have made our method fully applicable to *all* ethical frameworks. This is a shame because it reduces the degree to which subjective-wellbeing effect estimation can be called a fully non-sectarian method.

An acute and practical consequence of having merely interval information on welfare is that it becomes very difficult to value lives, and the risk of death, in a common framework with information about the quality of life. This makes welfare analysis of decisions which affect safety liable to be incomplete.

Although it is not ideal to lack ratio information, there is clear significance to establishing interval measurement. The information provided through SWBEE is likely to be of relevance, even to those whose ethical frameworks ultimately require ratio information for a full analysis.

Fortunately, there are at least two options for establishing a zero point through further research. One would be comparing participants momentary welfare level to unconsciousness by asking whether they prefer their current state to unconsciousness. Levels not preferred to unconsciousness would be taken as negative, and levels preferred to unconsciousness taken as positive. Another option would be to consider longer term wellbeing to establish at which point on the scale the average participant wishes they no longer existed. If the values derived by these two methods contradicted each other significantly, ethical and perhaps empirical inquiry would be required to select which was primary.

6.6 Conclusions for Chapter six

We have reviewed the question of the cardinality of psychometric implements. Having classified objections to the idea that these instruments are cardinal, we have argued that there are multiple reasons to think such measures may already be cardinal, that even if they are not cardinal, there are likely ways to cardinalize them, and that the evidence suggests that, perhaps surprisingly, any difference the scales have from cardinality may matter relatively little. We have concluded by noting that, in addition to achieving cardinality, it is essential for many purposes to give the scale a zero point and reviewed a few methods by which this could be done.

7 The problem of the role of values in welfare economics

7.1 Introduction: A political problem

Arguments over the roles of values in science come in many forms. Moreover, there is special argument about the role of values in the social sciences, stretching through foundational authors on social scientific methodology including Weber (Weber, 1949, 1922/1949). Without trying to enter into this classical debate, I would suggest that a way to usefully divide concerns one might have about values in sciences is as follows.

1 One might fear that the utilization and appeal to values in science will “contaminate” scientific reasoning in some respect, diminishing or removing the attributes which make science a uniquely successful enterprise in the quest for knowledge. (See Reiss and Sprenger (2014) for an overview of this hoary debate)
2 You might be worried about how you are going to include specific value commitments in science. In a democratic society with diverse values the adoption of any one set of values might make your research irrelevant to a broad swathe of academics, the public and policy makers. This is perhaps nowhere truer than around questions of economic advantage and economic justice. I term this the problem of ethical sectarianism. For a classical statement see Bright (2018) on du Bois. or Betz (2013) who argues science should be kept value free as: “political decisions are informed by scientific findings, the value free ideal ensures—in a democratic society—that collective goals are determined by democratically legitimized institutions, and not by a handful of experts”.

This way of dividing the problem is not novel. For example, Intemann (2015) mentions both the fear that:

“Even if they can play positive roles, political and economic values have the potential to lead to bias” and the fear that “the extent that scientists make value judgments[...] their values will be undemocratically privileged over those of other potential stakeholders.” (Intemann, 2015, pp. 217–218)

In this chapter I will be concerned with the second sort of question, not the first, and as will become apparent, we will narrow our focus even further. We will not attempt to resolve this first problem of bias induced by values in this thesis, both because it is too big a task, and because it’s unclear that there’s any aspect of this problem, which is peculiar to, or unusually difficult in, the context of welfare economics. The first sort of question has been fielded well in many places (Alexandrova, 2012; Colander, 2018) and even if one does not accept that it is a resolved question for *theoretical science* it might plausibly be said to be solved for applied sciences, insomuch as action-guiding enterprises are necessarily, in part, value-driven exercises. Values enter at the very least, to the degree that if we are to give recommendations for action, these recommendations must be specified with respect to a goal, i.e., they must take the form of hypothetical imperatives (Harsanyi, 1958). We do not choose the goals that we wish to study or how to achieve them at random (Machlup 1969) and so there will necessarily be a degree of value 'contamination' in the selection of goals. This cannot be restricted to the outset of the process of science- in a kind of "set and forget" for values- because- again as Machlup and others (Harrod, 1938; Streeten, 1953; Hall, 1959; Klappholz & Agassi, 1959; Klappholz, 1964; Machlup, 1969; Kaysen, 1969; Lerner, 1969; Lowe, 1969; Nelson, 1987; Y.-K. Ng, 1972; Nelson, 1987; Weston, 1994) argue, goals are always underspecified before contact with reality. Thus, the scientist will have to keep referring back to values throughout the process.

We now turn our attention to the second sort of problem regarding values in science. In the context of this thesis that problem is how welfare economics can proceed while remaining defensible in a society with extreme divergence on questions of distributive justice. The claim of this chapter is that it is feasible to meet the second challenge regarding values in science whilst meeting the needs welfare economics is supposed to meet using the approach to welfare economics I have championed in other chapters. To recapitulate- that approach is to measure and estimate the effects of economic policy on subjective-wellbeing- an approach we call *subjective wellbeing effect estimation* (SWBEE).

7.1.1 Making the political problem of value commitments "bite" for welfare economics

Previously some philosophers have attempted to address this second sort of problem and argue that there is nothing improper about sciences taking strong value stances in a liberal democracy (Kitcher, 2011). Alternatively, some philosophers have attempted to argue the opposite, suggesting that democracy gives us important reasons to comply with the value free ideal (e.g. (Bright, 2018), not so much presenting his own views but sympathetically explicating the views of W. E. Du Bois)). We do not propose to review all existing solutions here, because we intend to narrow the problem still further into something specific to welfare economics.

The second problem is worse for welfare economics than for many other fields for a number of reasons. Firstly, the role of values is as direct as can be imagined. The welfare economist is called to evaluate menus of policies, and some for example the presidential candidate Bob Dole in relation to cost-benefit analysis (Nyborg & Spangen, 2000) actually called for welfare economics to be the sole decider as to whether projects and/or regulatory changes go ahead, or at least act as a veto point. Secondly, values are rarely so starkly divided as in the question of *who gets what*. Both communism- the view that all things should be held in common- and forms of libertarianism like Nozickeanism which hold that government redistribution is violence- have adherents in the public.

Furthermore, it's not even clear what is on the menu. There is no simple correspondence between political philosophies and decision mechanisms in applied welfare economics. Copp (Copp, 1987) raises an interesting point in relation to cost-benefit analysis, viz that there is often a scale confusion. Many values frameworks in political philosophy are not well described or given in terms of decisions on localized issues. For example, Copp reminds us that Rawl's political philosophy is meant to apply to choosing the basic constitutional framework of society- not, for example, particular choices about road building. Even the utilitarian might prefer something different to a utilitarian social welfare function when it comes to making particular decisions, for example to preserve political stability. Harsyani (1976) argues against strongly prioritarian (speaking to the 'Rawlsian' case) SWFs because implementing them may cause civil war. A utilitarian might have similar concerns about directly implementing utilitarianism through a SWF if they thought it was likely to be strongly redistributive in a way they feared could lead to political conflagration or a disaster of some sort. Stigler (1943) makes a similar but broader argument that welfare economics should aim at wide ascent for the methods it chooses. Alternatively, perhaps the utilitarian would prefer not to use a utilitarian social welfare function in the context of cost-benefit analysis, and instead conduct redistribution through tax and transfer. Thus, even once we have laid out the enormous range of philosophical views on justice in distribution, still more diversity might remain about how to implement these values frameworks. There is no simple equation between a social welfare function in the sense of a rule that is intended to structure social welfare tradeoffs at the level of a society, and a social welfare function intended to weight cost-benefit analysis.

We are approaching the second problem not as a purely theoretical problem that could be solved by showing something general about the relationship between science and values. We are considering the problem as an actual barrier to the public, bureaucratic and academic acceptance of policy and research work, a barrier that has led to condemnation of welfare economics both inside and outside economics (Bromley, 1990; Volokh, 2011).Current approaches have not, in practice, won anything close to universal assent, and if the criticism they have faced has been relatively moderate, we may attribute as the main cause widespread ignorance of their operation. Thus, they need to be amended to win assent even if only, at the least, in their presentation. In this chapter we will outline why we think our preferred paradigm in welfare economics is neither value sectarian nor irrelevant to the tasks of welfare economics.

To sum up, *we are interested in the contextually specific features of the problem of democracy and values in welfare economics- specific features that have aggravated the problem beyond its normal bounds in the rest of science and sometimes made progress in the field difficult*. We are not so much looking for a total solution to the problem as a way to proceed that will render the problem no worse here than it is in other fields of applied science.

I do want to address a few strands of an already existing argument here about the second problem in a general sense, because of its direct applicability to welfare economics. The case is sometimes made that the problem of democratic legitimacy in science could be addressed through democratic consultation to decide an "official" values framework for this or that science. For example Lusk (2021) argues that we can overcome the problems of values in science through a process of deliberative democracy: "Since proper deliberation secures political legitimacy, then selecting the values to use in research in this way would also be legitimate".

Lusk does not discuss welfare economics directly (except a brief discussion of time discounting in climate mitigation). To some extent the question of whether such an approach would work in welfare economics is an empirical question- we would need to actually attempt the procedure with a large and representative deliberative body and see if they are capable of coming to a collective decision, and then see if the public accepts that decision. Lusk is entitled to be optimistic about the probability that a deliberative democratic process could deliver a clear set of values which would resolve the problem of the interpersonal aggregation of costs and benefits. However, until such a process is carried out (a task which would require among other things, funding and a great deal of time), the welfare economist has to face the challenge of saying *something* about choosing between different policies by evaluating the aggregation of harms and benefits. In other words, our response to Lusk is to plead that even if such a process might work in principle, in practice it has not yet been tried.

What about everyone simply declaring their values? You might imagine that a way out of the second problem is as follows. Every researcher simply states *their* values very clearly. Baujard (2021) calls this approach the transparency requirement. Such an approach might work for theoretical welfare economics and has been championed in by voices as diverse as Robinson (1964), Myrdal (1970, 1953), Machlup (1969), and it is a feature of the classical methodological tripartism of both N. Keynes (1890) and Walras (Mosini, 2012), where explicit normative content is permitted and encouraged within the sub-field of normative economics, an approach later defended by Robbins (1938, 1981) and widely discussed (Backhouse & Medema, 2009; Colander, 2009; Hands, 2009; Masini, 2009; Scarantino, 2009).

A similar position to this is implicit in the view that welfare economics is a form of normative economics, connected to moral philosophy, in which the theorist engages in ethical argument to defend the approach they advocate (Hawtrey, 1946; Boulding, 1969; A. K. Sen, 1987; F. M. Fisher, 1956; Weston, 1994; Little, 2002b; Atkinson, 2001, 2009, 2011; Broome, 2008; Beckerman, 2010; Mosini, 2012; Hausman, 1992; Hausman & McPherson, 1996; Hausman et al., 2016).

The major problem with this approach in applied welfare economics is that in applied welfare economics, research is often commissioned for assistance in making specific government decisions. Thus, there is often a sense, in the public, government and even among those working on the project, that any value framework adopted should be representative of the values of the public or values that the public has endorsed.

Broader aspects of the argument aside, even if it could work for theoretical welfare economics, in the realm of applied welfare economics (applied in the strong sense of officially sanctioned work on concrete problems in a society) I am pessimistic that “laying our cards on the table” will solve problems. The notion that we can improve the situation via government bureaucrats announcing at the start of their report that they are looking at the data from a Rawlsian or utilitarian perspective is bold because a public servant is not, in general, paid to give their own ethical views on a subject.

But what if the policy were set at a government-wide level rather than by individual bureaucrats? What if the government of the day simply threw the problem of value sectarianism aside, arguing that its democratic mandate to govern *a fortiori* gave it a democratic mandate to set the value function by which the consequences of policy were assessed?

This would require the party in power to distil their complex and changeable ideas- their own and the public's- about the common good- into a single "official" value set. Perhaps this might be a formal social welfare function, or perhaps it might be something more complex, like a social-welfare function combined with a flowchart of deontic side-constraints. Such an official values set could be deployed and redeployed for different questions about government policy. In principle, I suppose, there is nothing stopping governments from creating such a value set. There would be a number of barriers though, most importantly:

A) It would be a large political target. If the government does not formally announce a social-welfare function, every voter can imagine that the government uses roughly their own values in making decisions. If they do announce a social welfare function, the extent of their ethical divergence will become clear.

B) It is not clear that the complex (Colander, 2018) and varied values of a governing coalition can be distilled into a single social welfare function or more complex ethical structure.

C) Many governments at least like to create the impression of consultation with the public on ethical questions and concern about public ethical views on the questions of the day. It is unclear how adopting a specific value function as canonical would be compatible with this.

7.1.2 Applied weighted CBA- the case of the HM Treasury Green Book

Thus, the practical feasibility of this approach is, at least, dubious. Still, as previously discussed in chapter 2, we must acknowledge that, in the UK, the Green Book has since 2003 (HM Treasury, 2003) set out guidelines for CBA which recommends weighted cost-benefit analysis where policy notably affects the income distribution, suggesting that such an approach is not impossible.

The rules governing when weighting is to be applied seem to be somewhat flexible and the application itself somewhat elastic- this is far from a total Social-Welfare Function. Nonetheless, we note that it is clearly not *practically impossible* for a government to insist on specific weights. Still, the rarity of the approach internationally, and the elasticity with which the UK seems to have approached it, may back up our contentions regarding the difficulty of an "official" value set.

7.2 How SWBEE navigates the dilemma

7.2.1 SWBEE as a solution to the problem of values in welfare economics

In chapter four we introduced the idea of SWBEE, now we will try to show that SWBEE works as a resolution to the problem of values in welfare economics using cost-benefit analysis as our case study. Our argument for SWBEE here is best framed in terms of the navigation of a dilemma. How can we articulate an alternative approach to traditional CBA which does not shy away from engaging with values “honestly”, but also does not presume for the economist to usurp the role of either the democratic public or policymaker? A good, explicitly value informed approach has to avoid two things- the Scylla of presumptuousness- usurping the democratic public and/or decision makers and the Charybdis of irrelevance, telling us things which are not “strong enough” to be useful in making decisions about whether or not to proceed with projects, regulatory changes etc.

In order to show how SWBEE navigates this dilemma, I want to introduce a distinction between two different ways that applied science can comment on matters of ethical concern. *The prescriptive mode* advises us to act in a certain way. On the other hand, the *indicative mode* simply makes observations about the presence, absence and degree of various ethically relevant variables (or variables that some hold to be ethically relevant) under different scenarios. For an example of economic science adopting the prescriptive mode, consider the social welfare function approach as typically understood, discussed in section 2.7, and its application- e.g. through weighted cost-benefit analysis (Adler, 2016)- an approach which claims to tell us what the ‘right’ course of action relative to some Social-Welfare Function is. A quantification of levels of wellbeing, on the other hand, proceeds in the indicative approach.

We shall introduce another distinction- one closely related but separate. The difference between a method which quantifies facts, and a method which quantifies value judgements. One feature of traditional SWF based evaluations of policies is that they are often direct claims about *values* rather than claims about *facts.* For example in the stipulationist case, SWF may be based on a direct ethical intuition over the goodness of the income distribution. In some variants of welfarism, for example prioritarian and sufficientist variants (Frankfurt, 1987; Crisp, 2003a, 2003b; Brown, 2005, 2007) the measure of social welfare is fully relevant only from the perspective of some particular ethical system which is encoded in the SWF. One noteworthy exception here is the utilitarian case where the SWF, while also dependent on some ethical system, nevertheless is also equivalent to some measure that can be naturally interpreted as a factual indicator, for example the mean or sum of welfare, where welfare is some empirically measured variable, for example happiness or life satisfaction.

We might think of this in terms of directly and indirectly ethical variables. An indirectly ethical variable is a statement about things in the world, e.g., “positive emotions” that many people happen to attribute value to. A directly ethical variable is, rather, a variable that represents a pure ethical judgment about a situation, or at least a partial judgment over a feature of a situation, for example, an assessment that the social utility of the marginal dollar is inversely proportional to income based on a pure value judgment about distributive justice. The corresponding measure of social welfare in this case would be a directly ethical variable.

There is a certain consonance between what I call the prescriptive mode and directly ethical statements about value. To say that “such and such a state of affairs is good” - to make statements about value-is to come very close to saying “we should pursue x state of affairs”, to make a prescriptive statement. Nonetheless, the value statement/fact statement - prescriptive mode/indicative mode distinction are not quite the same thing. To see this, consider that the statement “policy x maximises the good”, and the statement “we should adopt policy x” do not imply each other, at least without further ethical assumptions such as consequentialism. The possibilities are outlined in table 2.

Table 2 A typology of positive and normative measures

	Prescriptive mode	Indicative mode
Statement about value	1	2
Statement about facts	3	4

It seems to me that, in order to avoid making presumptions about the correct ethical framework, a methodology must fall into quadrant four.

Traditional unweighted cost-benefit analysis when used for policy formulation falls into quadrant three. It deals with a positive, not normative variable- willingness to pay- and on that basis delivers a judgment about whether the project should proceed. Traditional weighted cost-benefit analysis when similarly applied falls into quadrant 1- it starts with a positive variable (willingness to pay) then performs a normative transformation on that variable- then on the basis of the overall sum, delivers a prescription on the project. Note, however, that as we have discussed weighted cost-benefit analysis here, we are excluding interpretations similar to ours that apply weights to estimate some potentially positive variable, such as ‘aggregate life-satisfaction’, where, as discussed above, yhis would be equivalent to the sum of some empirically measured welfare indicator.

An outstanding issue which may have already occurred to the reader is whether we can reconstrue unweighted cost-benefit analysis as purely indicative, in which case unweighted cost-benefit analysis would fall into quadrant four- a purely positive measure of the sum of monetary valuations for and against. The problem, as we will discuss later, is that it is unclear what relevant information is being given and what value, if any, it has.

SWBEE falls into quadrant four. It is an indicative method- it provides information rather than directly making a policy prescription. It also makes statements of facts, rather than value judgements. SWBEE aims to tell us how policies will affect changes in the distribution of subjective-wellbeing according to a variety of different definitions of SWB.

This is not a statement about values though it may well inform value judgements. Nor is it intended as a prescription or claim that a particular policy that optimizes effects on subjective wellbeing is thereby the correct policy- i.e. it is in the indicative mode rather than the prescriptive mode. In this way it avoids the Scylla of presumptuousness.

But such an approach can also dodge the Charybdis of irrelevance as well because it gives us useful information. It tells us how psychometrically measured wellbeing is likely to be altered under various policy scenarios. On most theories of welfare, it will thus tell us something about how policies will affect welfare, as most theories of welfare imply that it is at least partially psychometrically measurable, or likely to be closely correlated with factors that are. Thus, for almost anyone who cares about welfare in policymaking- that is almost everyone who cares about policymaking simplicter- SWBEE gives something of value.

Of course, the claim that it is worth carrying out SWBEE is a value judgment, in the sense of an implicit and limited endorsement of ethical systems which place value on the knowledge uncovered by SWBEE. If the credibility of such ethical systems was disavowed, then there would be little case for SWBEE analysis. This value ladeness of research stemming from prioritisation is hard to avoid (Klappholz, 1964) and our hope is that it is like the kind of value judgment involved in Pareto efficiency- a value judgment that relatively few will object to- even as it escapes certain limitations of the Pareto approach we discussed in chapter 2. The motivating value judgment is simply that it is worth knowing something about the effects of policies on subjective wellbeing of various sorts, before proceeding with those policies. We might hope even those who object to that value judgment may not object to it very strongly- very likely not as strongly as others might to welfare economic frameworks such as Social-Welfare Functions that make direct judgements over the good and even the right.

One might also worry that there will be value judgments necessarily involved in predicting the effects of policies on subjective wellbeing measures because predicting social outcomes is part of social science, and all social science, or perhaps all science generally, is value laden. My response is that, if this is true, it is an achievement to have returned us from the kind of value ladenness traditionally involved in welfare economics, to the ordinary sort of value ladenness that- on this theory- attaches to social science. Ordinary empirical and quantitative social science, if it is necessarily value laden, is value laden in a way which does not forestall research and measurement at least not as completely as differences about values in welfare economics which have often derailed it.

But what about those who aren't interested in information about how policy affects subjective wellbeing? Perhaps some deontologists, such as Nozickians would fall into this category (see our discussion in chapter two of non-welfarism). Yet even for these people, information about subjective-wellbeing effects might be useful as a tiebreaker between equally permissible or impermissible policies. We might imagine that they could want the information for other reasons as well- perhaps simply because other citizens care about this sort of thing in their decision making, and so it will be important if only to follow and engage discursively with their arguments. Still, I concede that the information will be more relevant to some than to others, but this is not, I think, enough to show that the method is ethically presumptuous or sectarian- at least not to anything like the same degree declaring a singular social welfare function is.

In other words, we don’t claim that proceeding in the indicative mode can ever be entirely free of independent ethical judgment, and hence some degree of presumption. As Machlup (1969) among others (Harrod, 1938; Streeten, 1953; Kaysen, 1969; Lerner, 1969; Lowe, 1969; Klappholz & Agassi, 1959; Klappholz, 1964; Weston, 1994) note, the practice of the policy economist almost always requires ethical judgment, because directives from policy makers are rarely completely specified. Escaping the role of values in general is not our goal, we will be content to lower the temperature.

One relevant critic of CBA who gives us an opportunity to clarify our approach is Colander (2018). Colander argues that the SWF approach fails because it is not feasible to condense all of society’s ethical and political concerns into a single metric. For Colander, what traditional social welfare functions aim to achieve is- a summary of something -the pros and cons of a policy or action- that is both more complex, and less unitary than can be captured by a single function.

It is important to clarify our approach to show we do not fall prey to something like Colander’s criticism. We are making no claim that the information provided is *comprehensive* in the sense of providing all the information needed for ethical judgment, rather, only a much more modest claim is being made, namely that the results generated by our method are often important enough to justify the costs of gathering them.

Colander’s perspective is compatible with our own. His argument against the SWF approach mirrors our own argument against the prescriptive approach to cost-benefit analysis. It is too much to expect a quantifiable metric to give us the answer. Both due to the complexity of ethical analysis and due to the diversity of democratic society, metrics should only give us more information to debate and discuss, i.e., they should proceed in the indicative mode.

7.2.2 Our solution does not exclude the existence of other valid approaches

In this thesis, we argue that subjective-wellbeing estimation of the effect of policies is a valid approach to welfare economics, insomuch as it does many of the things that scholars have traditionally wanted welfare economics to do and has strong options for dealing with many of the traditional limitations of welfare economics. However, as we have noted elsewhere, this does not exclude the possibility that other kinds of analysis of policy is rightly conducted under the banner of welfare economics and does not suffer from these flaws.

The relevance to this section is that we also do not deny that there might be other approaches that move between our Charybdis and Scylla of policy irrelevance and ethical presumptuousness. We are simply arguing that our approach navigates these difficulties, while in chapter eight we will argue at more length that another popular approach (traditional CBA, weighted and unweighted) does not pass this threshold. We acknowledge there could be other solutions that successfully navigate the value problem- whether similar or dissimilar to our own.

7.2.3 Supporting ethical frameworks which are not adequately represented by psychometric measures of wellbeing or which are outside scope

There are many different kinds of subjective wellbeing. Many partially or wholly correspond with what certain theories of the good life say constitute human welfare. If decision makers are split or uncertain over which theory of the good life to favour, one natural response will be to give decision makers estimates of how policy may affect different kinds of subjective wellbeing as we discussed in chapter four.

However, one problem with trying to include estimates of the effect of policy on many different kinds of subjective-wellbeing without privileging any- is that it will be difficult to know where to draw the line as to which measures of subjective-wellbeing researchers should use, and which can be excluded. Resources for such modelling are, after all, finite.

Suppose, hypothetically, that a study on the likely effects of a policy on subjective wellbeing included life satisfaction, many kinds of eudemonic wellbeing and hedonic wellbeing but did not include, to the chagrin of a supporter, some other measure, such as one philosopher's particular conception of eudaimonia. Maybe the study excluded that form of eudaimonia because of resource limitations. Regardless of the cause, If the study aims to avoid ethical sectarianism, surely it has failed at least vis a vis the concerns of this particular philosopher, their supporters and their concept of eudaimonia.

Part of the solution here, as we have stated so many times, must surely be to argue that different measures of wellbeing are intercorrelated with each other, making the problem of selecting which concepts of the good to include less acute, since concepts of the good that we have not included in our study will at least correlate with those that we have included.

If these correlations are thought to be large enough that, say, life satisfaction outdoes traditional welfare economic methodologies, such as unweighted cost-benefit analysis, in predicting the effects of a policy on eudaimonia, then this will give reason for the eudaemonist to prefer the measurement and estimation of policy effects on life-satisfaction to the use of unweighted cost-benefit analysis. The same might well be true even of weighted CBA for any given weighting scheme- see chapter nine for a discussion of the limitations of any revealed preferences framework set against a self-report framework.

Another way to deal with the problem of which measures of wellbeing to include- given there are a potentially infinite number- is through democratizing the process. Alexandrova (2017) documents the use of a public consultation process to decide on a theory of welfare for official government purposes. A similar process of public consultation could be used to decide on the kinds of subjective wellbeing which the democratic public is most interested in estimating the effects of policy on. Deciding on a list of important welfare concepts seems likely to be much easier than deciding on an official, publicly endorsed, set of values for welfare economics.

There are other ways to democratize the process apart from a public consultation- for example, looking at surveys of what people value. In the European Social Survey 2016 dataset (European Social Survey Core Scientific Team, 2016) ~69% of the sample said that it was "at least somewhat like them to think it is important to have a good time". Approximately 66% of the sample said that it was "at least somewhat like them to think that having fun and seeking pleasure are important". What this tells us is that a measure of whether people are enjoying life is going to capture something of importance to a majority of people (66%-69%+ a figure I must admit is lower than I would have anticipated). Results that show most people value hedonic wellbeing is of cumulative importance with results that suggest hedonic wellbeing correlates with other measures of subjective wellbeing. Both sets of results give us confidence in thinking that, whatever true welfare is for the ordinary person, it is probably correlated with hedonic wellbeing.

Much the same argument could be repeated for life satisfaction. Arguably it is definitional that most people value life satisfaction *at least in their own case*. How could one desire to be dissatisfied overall with one's life, since having life satisfaction is having what one wants for one's life, and by definition one wants for one's life what one wants for one's life? Similarly, it seems likely that most people are trying to flourish, to achieve fundamental capabilities for themselves and so on.

Although I do not wish to place too much faith in philosophy as a profession, it also seems likely that, if there was a popular concept of the good life among the public, far distant from any of these, some philosopher would have popularised it in the literature by now.

The combined result of these considerations, applied to any given welfare metric, is to give us confidence that these metrics likely captures something of importance to a majority of the public, even if only through correlation. Thus, for example, we can rely on evidence that people value hedonic happiness as a kind of democratic assent for measuring happiness in policy assessment. Further work on the public's views of the good life is vital, but we have at least a basis for proceeding on the assumption that many existing concepts of wellbeing matter to the public or are strongly correlated with things that do.

There is a kind of awkwardness that could arise in the context of pursuing our approach to welfare economics. What if an assessment of the policy effects on subjective-wellbeing gives utterly different, even contradictory results for different forms of subjective-wellbeing? One area in which this is perhaps especially likely is a contradiction between *life satisfaction* and *happiness* on the effects of money on wellbeing. As we have previously mentioned, income has been shown consistently to be more closely related to life satisfaction than to happiness (Diener et al., 2010; Kahneman & Deaton, 2010; Killingsworth, 2021).

Hersch (2015) in a rejoinder to Hausman's (2011, 2012a) evidentialism argues that it fails because different sources of evidence about welfare can diverge greatly. Hersch gives the example of a colonoscopy study by Redelmeier and Kahneman (1996) which we discussed in chapter 3. A moment-by-moment assessment of discomfort, and a global assessment of discomfort, were given. Naturally, the length of the procedure had a great influence on the overall quantity of the discomfort experienced during the procedure- more moments to be discomforted leads to a larger overall sum of discomfort. However, the global assessment of discomfort had virtually no correlation with the moment-by-moment assessment. What good is 'evidence' about welfare if that evidence can point in wholly different directions?

Because the approach we describe is meant to give purely indicative information, the answer to this problem is simply to report the contradictory results. We will then at least have much more specific information than we did previously. Even if you regard such a contradiction as troubling, that's no reason to let the possibility of such an impasse prevent us from even gathering the data. It's quite possible that measures will be unanimous on questions like "which policy most increases mean welfare" and "which policy most reduces welfare inequality" more often than not. If they are not unanimous, we will be left in a difficult situation, but it would be better to face that situation with more data rather than less. On the specific issue of integrating experience sampling and estimating probable experience sampling results from other forms of data, more research is needed as we discussed in chapter 3.

7.2.4 The problem of selecting plausible measures of subjective wellbeing inequality

As we have defined it, part of subjective wellbeing effect analysis is studying the effects of policy on the distribution of happiness, and not just the aggregate or average level of happiness. This includes, for example, measures of the inequality of happiness in the population. We can study welfare inequality simply by inspecting distributions directly, but it is often superior to provide summary statistics of inequality.

There are many different measures of inequality- for example, the Gini, Theil, generalised entropy, and Atkinson indexes. The proposal I have developed is to act as an "honest information broker", providing information on the level and distribution of ethically relevant effects of policy (changes in welfare and in the inequality of welfare) without ethical bias, which measure(s) of inequality, then, should we choose?

Inequality statistics themselves can be construed as corresponding to Social-Welfare functions (Aigner & Heins, 1967; A. K. Sen, 1982; Osmani, 1978; Sheshinski, 1972; Atkinson, 1970; Atkinson & Brandolini, 2015; Blackorby & Donaldson, 1978; Dagum, 1990) so their selection is ethically fraught. This is no welcome news for an approach that is trying to avoid ethical controversy!

Part of the solution is that unlike the problem of selecting welfare measures, we can be responsive to further requests by decision-makers after the experiments are conducted, as we can simply recalculate inequality using a new metric as required. The best approach is likely to illustrate the distributional effects of policy on subjective wellbeing in a variety of ways, using inequality metrics, graphics etc. Yet there are limits to the analyst's time, as discussed by Machlup:

> *It is not feasible to propose a sufficiently large number of alternative value systems to do justice to the existing variety of tastes and preferences. It is not feasible to tell those in the seats of government that they have a choice amongst millions of different value systems, and that, corresponding to each, there may be a different answer to their specific questions. The welfare economist, if he is very conscientious, will at best specify a small sample of alternative welfare functions and, in limiting the open choices in this way, will have engaged in normative economics. (Machlup, 1969, pp. 123–124)*

Will any attempt to provide a *satisfactory* range of subjective-wellbeing inequality statistics then inevitably fail due to ethical divergence? There are two possibilities:

i. People's ethical views regarding welfare, subjective wellbeing and its distribution are so multifarious that there is no *plausible* number of summary statistics and models regarding subjective-wellbeing inequality that would even approximately satisfy everyone.
ii. A relatively small number of inequality statistics will capture the main views regarding optimal SWB distribution.

My sense is that the latter, rather than the former is true, but this is a question for future empirical and conceptual study.

Even if SWB inequality statistics are hopelessly mired in sectarianism, and no practicable number could feasibly satisfy everyone, we do have alternatives, such as simply presenting the contrasting expected distributions following different policies. Perhaps a computer program could be devised to let anyone- government decisionmaker or curious citizen, specify the parameters of their own normative measure of inequality and then run it over the expected distribution.

In the case of some subjective-wellbeing metrics with a fairly limited number of scores (e.g., Cantrill's ladder with an eleven-point scale) we can very easily present the projected distribution on that scale before and after among the population affected by a proposed policy. This data will presumably be much easier to consider visually than an equivalent graph of the income distribution.

7.3 Will presenting welfare information be a source of disagreement or consensus?

We have tried to square the circle- to provide a method through which economics can engage in matters of ethical importance, in a way that matters to just about anyone with ethical concerns, without becoming ethically sectarian.

I have taken it for granted that although the welfare economist can gather data on the effects of policies that everyone will want to review, it is unavoidable that there will be enormous disagreement on the right way to respond to this data, along the lines of ethical disagreement. While the role of welfare economics in an ethically sectarian world is thus assured, nonetheless the tragic reality is that the gathering of this data is only a prelude to strife and disagreement about how to evaluate it. There is, in this picture, little reason to think welfare economics will bring us closer to a consensus on policy.

But I would like to end this chapter on a more hopeful note, first with a conjecture:

> *In most cases, when presented with data about the effects of a policy on various measures of subjective wellbeing, most people, conditional on them trusting these estimates, will agree on whether that data morally supports or opposes a particular policy. This does not necessarily mean that they will thereby all come to a consensus on the policy, as they may disagree about it for other reasons or distrust the data. Nonetheless, the data itself will drive agreement, not disagreement, even if it is outweighed by other factors.*

I base this on a sense that similarities between most people's values in relation to welfare are much greater than the differences in practice. It is the various deontic side constraints and other ethical complications they add onto these that drive them apart- as well as empirical differences. I offer it purely as a conjecture for future research. If it turns out to be false it will not undermine the argument that we have made in this chapter.

Something that perhaps gives us reason to hope that the above conjecture is true, is unpublished research by Latty (Latty, 2015a, 2015c). Latty (2015c) alters the Atkinson index to support prioritarianism, adding a parameter p which sets the strength of prioritarian concern (in isoelastic terms). The paper calculates how much the index increases for stylised income distributions as p increases. Even when prioritarian concern is increased to $(p = 1.2)$ the Atkinson index only increases by 1 to 5%.

Similar results are shown for logistic prioritarianism. In logistic prioritarianism a parameter K specifies the ratio of the welfare weight of the maximally worse off to the welfare weight of the maximally best off. Inequality aversion is modestly raised, even for large values of K. Latty then calculates social welfare losses resulting from "getting the SWF wrong", or more specifically, optimising the income distribution using a value for K that differs from the value used to assess the distribution:

> *consider that a utilitarian has optimised the income distribution by setting $K = 1$, but the [prioritarian] observer supports a higher value . . . modest losses result from disagreement between parameters in the optimising and assessing SWF. For example, when a utilitarian SWF is utilised to optimise the income distribution, social welfare losses under a strongly prioritarian SWF ($K = 256$) amount to 12.3 % of income. Conversely, when the income distribution is optimised under strong priority ($K = 256$) but is assessed under a utilitarian SWF losses amount to 9.8%.* (Latty, 2015a, pp. 16–17)

In other words, surprisingly little difference is made to the selection of optimal policy as a result of the differences between welfarists which Latty has considered. This may give us prima facie reason to believe that welfarism, despite its seemingly endless capacity for variation through different social-welfare functions, tends to converge in its conclusions about a wide variety of cases, and across a wide range of social welfare functions.

7.4 A note on WELLBY'S

Frijters and Krekel (2021) propose an approach that quantifies wellbeing changes as a kind of cost-benefit analysis. They quantify increments in welfare equivalent to a one-point gain on a ten-point scale, and conclude that six such increments are worth one typical life year. We will refer to this as the WELLBY approach, after the unit of measurement its authors introduce—a one-point happiness increment on a 0–10 scale for one year.

Naturally, if we can quantify either the value of a life year or the value of an increment, this will allow us to quantify the other, given their inter-convertibility. Since we can now convert happiness points, life years, and (government-held) dollars into a common scale, we have a natural and powerful mode of cost-benefit analysis.

Frijters and Krekel derive, in my opinion, an unrealistic and dangerously low valuation of one life year—£15 000—using as a revealed preference the amount it costs the NHS to save a life at the margins (the NHS is notoriously likely below the welfare-optimal level of funding); in principle, though, any other figure can be substituted. This, of course, raises a wealth of worries typical of cost-benefit analysis—namely, many free points of methodological choice leading to very different results. Still, the consequences can be mitigated somewhat by sensitivity testing. The difference from our approach is the implicit assumption of a utilitarian social welfare function. A happiness point is a happiness point, and a life-year at a fixed happiness level is a life-year at a fixed happiness level—it's a sophisticated method for utilitarian accounting. If we assume that the value of a WELLBY should be set according to how many dollars would have to be taken out of the budget to, in expectation, cause the loss of one WELLBY, we have something more or less wholly equivalent to a utilitarian approach, including distributional insensitivity, additivity, and insensitivity to concerns not part of the welfare measurement unless they are specified separately.

As a tool in our arsenal, this is quite compatible with the SWBEE approach—an accounting in WELLBYs can be reported along with other metrics. Taken on its own, though, it runs into all the standard difficulties of the utilitarian approach—and has all the standard strengths of that approach.

7.5 Conclusions for chapter 7

We have outlined how SWBEE avoids the problems of value sectarianism in welfare economics- by conceptualising itself as providing *information* relevant to the public and decisionmakers, rather than adopting a specific social welfare function. Finally, we have ended with an expression of hope- that presenting information regarding the SWB effects of policy may tend to drive consensus, rather than disagreement.

8 Cost Benefit Analysis and values in Welfare Economics

8.1 Can the ‘just giving information defence’ also save unweighted CBA? The problem of transfer costs- economic and political

One objection to the picture that we have outlined might be as follows. We have defended a psychometrically grounded, indicative approach and argued that it can meet the challenge of being neither presumptuous nor irrelevant. Couldn’t traditional unweighted CBA also pass this challenge, if we treated it as an information provider rather than a decisionmaker? My first reply is that perhaps sometimes it is wise to include an estimate of unweighted aggregate costs and benefits to the decision maker, as part of a broader package of information, but that alone will not save CBA as a broadly applicable and general methodology. My second reply though is to raise scepticism about how often even that much is true. Exactly how, and how often, does unweighted cost-benefit analysis provide useful information for the decisionmaker and public?

If unweighted CBA provides useful information, it is presumably useful information about whether or not the economic pie has been expanded. The usual way of defining the size of the economic pie, and part of the primary theoretical rationale for cost-benefit analysis, is to ask whether a policy represents a Kaldor-Hicks improvement.

Situation A is a Kaldor-Hicks improvement on situation B if and only if those who would win from a transition from situation A to situation B could collectively compensate those who would lose out from a transition from situation A to situation B. The rationale is that this makes situation A, *a potential Pareto improvement* over situation B- if only that compensation were actually paid, no one would have any basis on which to complain about the transition.

We are not here concerned with some of the more technical objections that have been mounted against the Kaldor-Hicks principle (Scitovsky, 1941b; Silberberg, 1972; Boadway, 1974; Boadway & Bruce, 1984; Blackorby & Donaldson, 1990; Jones, 2002). Instead, we are concerned with a curious and sometimes prominent idea that unweighted CBA and Kaldor-Hicks as a decision criterion or more generally the new welfare economics is “value-free” or even merely “value-lite”, a position rejected both by Robbins (1981) and the Bergson-Samuelson (Bergson, 1938, 1954; Samuelson, 1947; Tintner, 1946) tradition in welfare economics (Streeten, 1953) among others (Little, 1950; A. K. Sen, 1979b; Streeten, 1981). Chipman and Moore (1978, p. 548) give this verdict:

> judged in relation to its basic objective of enabling economists to make welfare prescriptions without having to make value judgments and, in particular,

interpersonal comparisons of utility, the New Welfare Economics must be considered a failure.

What do Kaldor-Hicks improvements represent? Many have argued not much, because it is not guaranteed that the government could actually make everyone better off after transfer costs (deadweight losses of taxation) are factored in. Let us modify an example from Copp (1987) to make this point more clearly- although as I alter it, it is not a very realistic example rather it is intended to show failures in the *logic* of Kaldor-Hicks.

Jane owns a car which she uses for her job as a pizza delivery driver. Now Bob is a rich man and would like Jane's car because it is a classic car and would be willing to pay $50,000 dollars for it. The most Jane can afford to pay to hold onto her car is $40,000 dollars (coincidentally, this is the minimum she would be willing to accept for the car too). The government transfer's Jane's car to Bob, because it is a Kaldor-Hicks improvement. The marginal deadweight loss of taxation is such that Jane can't be compensated in a way that would make the arrangement a Pareto improvement, no compensation is paid, Jane loses her job, Bob gets a vintage car.

In what practically relevant sense is this an expansion of the economic pie? If anything, on the ordinary understanding, the pie is shrunk somewhat because a productive good is transferred to a non-productive use. Now the reader might object that this is a somewhat slavish application of the principle, because any sane government would leave the parties to make or not make the transaction on their own, but more complex examples with the same underlying logic in which the parties independently transacting could not solve the issue are possible. For example, Bob wants the right to hunt pheasants on public land. This will mean the community cannot enjoy recreation on the land while he is conducting his hunts. If Bob gets his way because he is rich, but no compensation is paid (or can be with the enterprise still a net gain, given transfer costs), in what possible construal is the economic pie enlarged by this?

There is a possible modification to unweighted CBA which would deal with this issue. Define a *super-Kaldor-Hicks* improvement as a change in a situation in which:

A At least some people are made better off.
B The government could compensate any losers in the situation by taking money produced by the change and redistributing it to the losers (and the winners would still come out better) even accounting for the deadweight loss of taxation needed to achieve this.

This is a more robust concept of expanding the economic pie, insomuch as it gives us some more assurance that a policy is truly a potential Pareto improvement, and we are fully prepared to concede that it may have some role to play in informing government decision making. Of course, even the metric we have given is not perfect. It assumes that there is no cheaper way for the government to compensate the losers than paying them money (e.g., no in-kind payment that may be cheaper for the government than the loser's WTP) and it assumes that an individual's willingness to pay represents true welfare value (or economic value, whatever that may mean) to the losers and winners- i.e., that people are good judges of their own gains and losses.

By far the most important problem even for this super-Kaldor-Hicks approach is the existence of political costs. As Fenell & McAdams (2016) note, proposals based on the logic of potential Pareto improvements and the logic of efficiency often ignore the reality that there are political costs and difficulties in transferring money. We will revisit this point later in chapter 10 in the section *Political feasibility and Political Costs*.

We also note in passing, as a methodology interesting from the point of view of the task of adjusting for transfer costs, Hendren & Sprung-Keyser's (2020) approach: "We divide the willingness to pay by the net cost to the government to form each policy's Marginal Value of Public Funds, or its "MVPF". It may be possible to determine a critical value for MVPF, above which the required transfer costs are affordable even accounting for deadweight losses.

Although there have been theoretical investigations into what we might term actually compensable Kaldor-Hicks criteria (Dixit & Norman, 1980, 1986) there is a dearth of practical work on Kaldor-Hicks methods, with guarantees that the government can create a real Pareto improvement.

8.1.1 Value neutrality through non-weighting?

One way out of the problem of values sectarianism would be if the Kaldor-Hicks principle could be shown to be value-free, or perhaps in some sense "value light" –for example (Mack, 1952) takes something a little like this view. The least plausible version of these arguments is that the Kaldor-Hicks unweighted cost-benefit approach is more neutral because deciding on a particular system of weights would represent the adoption of a particular set of values. To go beyond "a dollar is a dollar" would be to leave the realm of value neutrality and make controversial value judgements. This has no credibility as an argument. Most of us recognise that it is, all other things equal, better to give a dollar to a poor person than a rich person because we recognise, as Copp (1987) puts it, that the benefit of a dollar to the poor is greater than the benefit of a dollar to the rich, and that to not recognise this in economic analysis is to treat different benefits similarly, and similar benefits differently, *a value judgment* of an inegalitarian sort.

But even If we do not accept this, at the very least we must accept that judging every dollar equal is itself a value judgment. From the point of view of values and policy decision making, there is no such thing as "unweighted" cost benefit analysis, treating a dollar as a dollar is just one way of weighing things. Furthermore, it is, prima facie, a bizarre method of weighing things. A method of weighting things which, as Brad DeLong (2003) once argued in a piece on the implicit utility function of the market, effectively amounts to weighting each dollar in a way that is inversely proportional to its marginal utility. Another way to put this is that standard cost-benefit analysis adopts *disequity weights* (phrase taken from (Copp, 1987))

The naive argument for Kaldor-Hicks value-freedom faces another difficulty- why should an applied science even want to be value neutral? In medicine (Reiss & Ankeny, 2022) public health (Harvard et al., 2020; N. Y. Ng & Ruger, 2011) and engineering (Davis, 2012; Diekmann & Peterson, 2013) and conservation biology (Barry & Oelschlaeger, 1996; Baumgaertner & Holthuijzen, 2017; Charpentier, 2015; Soulé, 1985) it is considered normal or desirable practice for an applied science to deploy value judgements. Health and public health in particular often involves controversial value judgements, for example Ne'eman et al. (2022) which frames its own values basis as right, and the public's as evidence of bias- even where some of the judgments preferred by the public, for example prioritising younger people for scarce health resources, are considered live options in the bioethics literature (Antiel et al., 2020).

At first blush this complaint might seem paradoxical or hypocritical. We are concerned in this chapter with the question of how to avoid controversial value judgements for the sake of the broad acceptability of welfare economics by the public, and now we have the temerity to complain about attempts to avoid value judgements altogether? Yet these complaints are compatible. Just as there might be a legitimate public expectation that a science should avoid being ethically sectarian, there may well also be a legitimate public expectation that an applied science intended to guide policy shouldn't avoid the *right* kind of engagement with ethical questions. Perhaps these demands will turn out to be incompatible, in which case one or more will have to be revised, but it is at least a desideratum to meet both.

8.1.2 What if a dollar is equally valuable for everyone

Karine Nyborg (2022, personal communication) has bought to my attention a possible argument in favour of the claim that an additional dollar for everyone has the same value:

1. Everyone has (approximately) what they deserve under the current income distribution.
2. Plausibly it follows from 1 that an additional dollar for anyone is equally (un)deserved.
3. Thus, everyone has the same (social) marginal utility of income.

Although she does not believe it herself, she has pointed out that the existence of this argument probably negates the claim that no one believes an extra dollar to everyone is equally valuable. But this argument requires controversial value judgements, and thus crashes into the sectarian horn of the dilemma.

We will consider a 'democratic' variant of this argument- that the public has sanctified the existing distribution which makes it 'value light' in some sense- later in this chapter.

8.1.3 The approach of Adler and Posner: weighted and unweighted CBA as providing a subset of the information required for policy appraisal.

In a lengthy discussion, Adler and Posner (1999) offer a defence of cost-benefit analysis that sounds conceptually very similar to our defence of SWBEE- viz:

1. The aim of cost-benefit analysis is really to estimate effects on another, harder to quantify variable- welfare.
2. Cost-benefit analysis should be understood not as an overall decision algorithm, but as an informational input- specifically as a loose indicator of whether a project will, on the whole, increase or decrease welfare.
3. In light of this understanding of what CBA is really trying to do, certain modifications of cost-benefit analysis may be desirable- for example, moving towards cost-benefit analysis weighted on income, where, for example, those for or against a project vary greatly in their means.

This approach mirrors our own in several ways. In both approaches, CBA (or in our case, something often like it) is reimagined as a way of trying to get at welfare. Both approaches aim to avoid the difficulties inherent in giving a definitive answer to the question of "what is welfare". Most relevant to this chapter, one of the difficulties of the social welfare function approach- defining an explicit, final decision rule- with all the ethical partisanship involved- is avoided. It is also avoided by the same method we use- making the process *informative* rather than *prescriptive.* Though in parts Adler and Posner do speak of the procedure as if it were prescriptive, in other parts they make it clear it should be seen as only one input into the decision-making process:

> *We do not conceptualize CBA as the exclusive choice procedure for government, but as one part of the overall set of procedures and institutions by which projects are ultimately approved, rejected, or amended.* (Adler & Posner, 1999, p. 245)

However, it varies from our approach in a few key ways. The primary difference is this- Adler & Posner do not posit an intermediate target between willingness to pay, and the great unmeasurable- welfare. True, both our approaches aim to approximate "welfare" whatever welfare may be. However, without an intermediate target like subjective wellbeing, Adler and Posner's approach will have difficulty dealing with issues like the choice of the inequality aversion parameter(s) such as ε in weighted CBA. I would argue that the most natural approach would be to investigate empirical differences in how humans respond to different amounts of money and resource provision, and such an approach guides us towards psychometrics for reasons I outline elsewhere.

8.2 Democratic defences of CBA

Some twists on the argument from value freedom for unweighted cost benefit analysis have more plausibility than others. In particular, let us consider an argument from democracy. This argument links it to the ideal of not presupposing the results of or undercutting the democratic process. What they propose is not, strictly speaking, that CBA is value-neutral, but rather that CBA only enables, or assists, the fulfillment of the democratic public's values in aggregate, whatever those values may be. In this regard these arguments are a development of what Baujard (2021) calls the value confinement strategy- attempts to base welfare economics not on no values, but only on values that there is scarcely any disagreement about, an approach arguably adopted by the New Welfare economics (see the discussion in Streeten (1953, 1981)) among others (Mack, 1952; Mishan, 1957) – and also subjected to criticism (Archibald, 1959).

An example of an author taking on something of the spirit of this line of argument is Weitzman (2007, p. 712) objecting to the Stern report on climate change and arguing that using the behaviourally derived values of cost-benefit analysis is, in spirit, more democratic:

> *Furthermore, Cline and Stern are soulmates in their cri de Coeur justifying $\delta \approx 0$ by relying mostly on a priori philosopher-king ethical judgements about the immorality of treating future generations differently from the current generation—instead of trying to back out what possibly more representative members of society than either Cline or Stern might be revealing from their behaviour is their implicit rate of pure time preference. An enormously important part of the "discipline" of economics is supposed to be that economists understand the difference between their own personal preferences for apples over oranges and the preferences of others for apples over oranges. Inferring society's revealed preference value of δ is not an easy task in any event (here for purposes of long-term discounting, no less), but at least a good-faith effort at such an inference might have gone some way towards convincing the public that the economists doing the studies are not drawing conclusions primarily from imposing their own value judgements on the rest of the world.*

(I am indebted to (Broome, 2012) for Weitzman as an example).

The argument goes as follows. CBA uses market prices and willingness to pay. As such it represents a "democracy of the dollar", as purchasing choices can be seen as a sort of voting process for setting prices. Cost-benefit analysis is really a kind of direct democratic algorithm. On the other hand, using a specific social welfare function is almost dictatorial- it places the values of one individual above all others.

This argument fails for two main reasons.

The first reason is that an unweighted CBA gives some people far more 'votes' than others due to economic inequality. Thus, it is not a democratic voting system so much as an oligarchic one. One could try to argue that the rich *deserve* their greater allocation of votes, but we have now entered the realm of exceedingly controversial value judgements.

In a period of special concern about the injustices done to those of marginalised races, genders, and sexualities, as well as disability status, it may be of note that giving the rich more votes also amounts to giving white able-bodied men more votes. Unweighted cost-benefit analysis may therefore violate norms and rules requiring equal treatment for protected categories like race and gender. We will come back to this problem of oligarchy in CBA later.

A second objection is that while CBA often depends on market prices, people may have different values than the values realized in aggregate in their consumption choices (Broome 2012 also makes this point). It is plausible, for example, that the same people who buy things which hurt rainforests or only give a relatively small amount of their money to charity might prefer government policies which protect rainforests and the poor. Why so? Because solidaristic communal action to protect and advance things and matters of value is more attractive than individual sacrifices, which can feel like taking a cup from the sea.

Even when CBA is based on surveys of willingness to pay for items of public expenditure through tax increases, this despair about individual action might nonetheless spill through. Questions are often framed individually as *are you willing to pay x much more tax to y* and not at the level of the collective.

Respondent's answers that they are not willing to pay more tax may also reflect a sense not that a project isn't worth a certain number of tax dollars, but instead that the government should already be capable of dealing with these matters by prioritizing existing expenditure. It may be that if they accepted the premise- that raising taxes is the best feasible way to achieve this- they would agree taxes should be raised.

Thus, if one were to ask people how much they are willing to pay to reduce global warming, one frames the problem in an individualistic way. This might make respondents leery of nominating themselves (even in a hypothetical) to make large sacrifices "on a limb" (Broome, 2012) and be exploited (even in a hypothetical) by free riders. On the other hand communal, not-individualistic provision may be seen as imposing a fair sacrifice on everyone in order to purchase public goods, maximize positive externalities and minimise negative externalities. For many reasons then, people may have different values regarding how they think the government should act on behalf of society and impose the required burdens of that action on society, and how they think they should act as individuals.

There is another issue regarding collective action problems here, and revealed values, specifically in relation to redistribution. Mill (1848/1909) argued that taxation of luxuries would be less injurious than other forms of taxation, because luxury taxes diminish the expenditures on status goods by social competitors:

> a great portion of the expenses of the higher and middle classes in most countries ... is not incurred for the sake of the pleasure afforded by the things on which the money is spent, but from regard to opinion, and an idea that certain expenses are expected from them, as an appendage of station . . . expenditure of this sort is a most desirable subject

> of taxation. If taxation discourages it, some good is done, and if not, no harm; for in so far as taxes are levied on things which are desired and possessed from motives of this description, nobody is the worse for them. When a thing is bought not for its use but for its costliness, cheapness is no recommendation. (Mill, 1848/1909, Book 5, chap. 6, pt. 7)

A similar argument was later made by Pigou:

> a larger proportion of the satisfaction yielded by the incomes of rich people comes from their relative, rather than from their absolute, amount. This part of it will not be destroyed if the incomes of all rich people are diminished together. The loss of economic welfare suffered by the rich when command over resources is transferred from them to the poor will, therefore, be substantially smaller relatively to the gain of economic welfare to the poor than a consideration of the law of diminishing utility taken by itself suggests. (Pigou, 1912)

Analogously, we could imagine, for example, a community in which many people want to redistribute income, but do not do so on an individual basis, for doing so would affect their social ranking (based for example on consumption in comparison to some social standard) as they fall behind near peers they are in intense social competition with. Government action for redistribution which affected their near peers and then maintained the extant social ranking might be more welcome. Thus the income distribution can be seen as a public good, as in Thurow (1971).

A more general way to put this point is that it is an error to take government, intended to solve collective action problems, and 'load it' with individual's preferences over individual action. Government, as a solver of collective action should, if it is to be democratic, take on our preferences over collective action. Of course, we could modify the wording and setup of contingent valuation studies to be more in line with seeing things like a society and less with seeing things as an individual. I would support such experiments with truly democratic cost-benefit analysis, and will discuss them further in section 8.2.2, but such a methodology would have to be worked out in detail, and while we do not dismiss the possibility it could help address the problem of values sectarianism in welfare economics in a different and interesting way to the method we describe, this is for other work.

On the subject of democracy in cost-benefit analysis, and individualistic versus collective preferences, it would be interesting to see what surveys reveal about the public's beliefs concerning the social pure time discount rate when they are asked about it as a question of their preferred government policy.

8.2.1 Are extant income distributions democratically supported?

In the above, we objected to the democratic defence of cost-benefit analysis on the grounds that it is oligarchical. One defence of this kind of oligarchy (Dobes & Bennett, 2009) is to suggest that the existing income distribution represents, or approximates, society's aggregate preferences on distribution:"[t]he current distribution of income in a democratic society reflects (albeit imperfectly) existing social preferences. To introduce any other set of weights risks the adoption of a paternalistic or authoritarian approach by the individual analyst or decision-maker." (Dobes & Bennett 2009 pp 23).

Any inequality in voting power in cost benefit analysis, therefore, is sanctioned by the social choice of our existing income distribution through democratic mechanisms. I have many objections against this line of argument, but let us focus on four:

1 It is not clear that democracy is *effective* with regard to implementing the preferred income distribution of the public (Bartels, 2004, 2005a, 2005b; Bonica et al., 2013; Flavin, 2012; Flavin & Franko, 2016; Gilens, 2014, 2005, 2009; Gilens & Page, 2014; Page & Gilens, 2017; Karabarbounis, 2011; C.-S. Lee, 2005; Lloren & Wuest, 2013; Torija, 2013; Rowbottom, 2010; Rosset, 2016; Schlozman et al., 2013) because there is evidence that the public would prefer a more egalitarian distribution of income and wealth (e.g. (Horowitz et al., 2020) who found 61% of Americans think the government should do more to reduce poverty and a poll by The Australia Institute 2018- with the remarkable, finding 79% of Australians think the government should do more to reduce inequality). I can only quote with admiration Bromley's (1990) witticism: "the rationale for this position is as follows: the current distribution of income must be the appropriate one for otherwise the politicians would change it. This rather surprising declaration of faith in politicians is the only time that an economist will admit to any confidence in the outcome of the political process. "

2 It is not clear that our existing democracy is procedurally fair in relation to discussions over the distribution of wealth. Thus, even if it could be established that the actually existing public preferred the existing distribution of wealth, it would not be clear from this that their preferences had full normative weight. They may have been formed under the coercion of an unfair conversation especially in terms of who has access to mass communication, etc. For a critique of the legitimacy of actually existing democracy, see Skinner (1973) and numerous others.

3 Even were existing democracy effective with regards to implementing the preferred income distribution of the public, and even if this were to establish that the existing economic distribution is legitimate, it is not clear that public support for the current income distribution implies public support for policies implied by unweighted cost-benefit analysis- e.g., building vastly more parks, hospitals, schools etc. for the wealthy than the poor. Public consent for the income distribution does not necessarily imply public consent for oligarchical decision making in the distribution of public goods.

4 Cost benefit analysis *is* part of government and thus part of the democratic process. There is something curious about saying "we must not change the democratically sanctioned distribution, that is for the elected government to decide" *when one is talking about decision procedures sanctioned by the elected government*. If we assume that democracy regulates the income distribution than cost-benefit analysis is part of how

> "the current distribution of income in a democratic society reflects (albeit imperfectly) existing social preferences" (Dobes & Bennett 2009 pp 23). If cost-benefit analysis were to avoid changing the income distribution in order to better conform to public opinion, it would paradoxically prevent the maintenance of that reflection of the public's views by that income distribution. Perhaps the presupposition is that redistribution is most properly done through the tax and transfer system, but it's not clear the public agrees- see section 10.2.4. for discussion.

Democratic sanction of the income distribution seems dubious as a premise in an argument that unweighted cost-benefit analysis avoids ethical sectarianism.

8.2.2 Could there be a truly democratic form of Cost-Benefit Analysis?

There is an interesting question, as to whether a democratic form of CBA- using a weighting that gives everyone "voting power" which is in some sense "equal" could be imagined. Such a form of CBA would approach the value-sectarianism-in-science problem in a wholly different way to the approach we have described in this chapter. Whether it would solve the problem is debatable. At a first pass the way to give everyone equal voting power might seem obvious- a form of cost-benefit analysis which is based on the *proportion* of a person's income that they are willing to pay. This will give everyone an equal number of votes, although there would be problems in the case of people with zero income, and in the case of people with little income but much wealth, a resolution could doubtless be found.

Yet it is not clear that this approach succeeds in giving everyone effectively equal voting power. The poor might greatly value education or entertainment, for example, yet because a large portion of their income is tied up in rent and basic necessities, this large value they place on education or entertainment won't show through. The natural solution is to weight based on something like the inverse of *discretionary* income, perhaps using a Stone-Geary (Geary, 1950; Stone, 1954) utility function but the exact definitions will be contestable.

Could such an approach also deal with the issue of private versus public priorities and the relation of the individual to the collective we raised earlier? Perhaps this could be managed by explicitly asking questions about the size of government and public expenditure priorities? The aim here is to turn CBA into a kind of direct economic democracy ballot. Would that be desirable? Would a conceptualisation of cost-benefit analysis as a direct democratic endeavour nonetheless be ethically sectarian, insomuch as many would not approve of direct democracy *at least if run in this way*?

I do not propose to deal with these matters here. As previously stated, my contention is not that there are no other paths to a welfare economics that addresses the ethical sectarianism problem. Instead, we are just arguing that: A) Our preferred approach is a solution to the problem and B) Some common traditional approaches do not work as well. We flag this project to create a “democratic CBA” as an interesting possible alternative route around ethical sectarianism.

8.2.3 Potential Pareto improvements and democracy

Another argument in relation to democracy and unweighted cost-benefit analysis is the following, related to considerations we have already examined around the Kaldor-Hicks criteria. Every Kaldor-Hicks improvement is a potential Pareto improvement. Pareto improvements are *unobjectionable* in the sense that everyone- or near enough- would endorse them. Where a decision results in a potential Pareto improvement, it must therefore be unobjectionable to democratic decision makers because they have the choice of subsequently providing compensation or not. If they implement the compensation, good and well. On the other hand, if they do not implement the compensation, it must be because they have judged the new distribution better.

There are a few problems with this argument. Two of the strongest counterarguments, both already mentioned in slightly different contexts, are:

1. As we have already discussed in relation to Kaldor-Hicks criteria, when we consider deadweight loss and transfer costs that would be involved in compensating the losers, it is not true that each Kaldor-Hicks improvement is really a potential Pareto improvement from the point of view of the government.
2. There may be political costs involved in redistribution which mean that the government could have the option to enact a policy which creates a potential Pareto improvement, but not have the political capacity to enact that improvement (Fennell & McAdams, 2016).

One way to think of the second point is in terms of political coalitions. Consider a Kaldor-Hicks improvement that harms A, but helps B. The kind of political coalition likely to support this probably doesn’t contain A, and probably does contain B. Meanwhile the kind of coalition most likely to support the compensation payment afterwards probably does contain A and probably doesn’t contain B. Thus, it may be particularly difficult to enact Kaldor-Hicks improvements and ensure that compensation is paid because the political balance of forces has to be just so in order to win A their improvement AND B their compensation. Even if some sort of agreement to proceed can be reached, the one side will be trying to prevent the whole change from happening and the other to prevent the full compensation from being paid.

Moreover, this argument depends on strong assumptions about the robustness, scope and legitimacy of actually existing democracies, some of which we raised concerns about in section 8.2.1.

8.3 Weighted CBA where weightings are direct ethical judgements?

I have made many objections to unweighted CBA, but only a minority of them have an obvious application to weighted CBA. So, what about weighted CBA, can it avoid both ethical sectarianism and ethical irrelevance?

The simplest form of weighted CBA adopts some weighting function where the weights are equal to the marginal value of a marginal income increase for a given individual, and then willingness to pay for a good is treated as equivalent to an income gain. See section 2.7 for discussion. Typically, there is some isoelastic weighting function of the form discussed in 2.7.

The problem is *what determines the weighting function, or in the isoelastic case, ε.* Is it a direct value judgment over how much we care about Janet getting an extra dollar versus John (Chapman, 1913; Robbins, 1938; Boulding, 1952; Myrdal, 1953; Reder, 1958)? If so, then that seems like an extremely controversial value judgment, crashing headlong into the Scylla and failing the injunction that welfare economics should not be ethically sectarian.

But the most telling criticism against this approach is due to Nyborg & Spangen (2000). Ideal democracy is generally thought to be a process, in which the public is meant to be intimately involved in decision making at each step. An algorithm set by policymakers to evaluate policy shuts out the deliberative element of democracy, which is meant to include not just ongoing deliberation over *facts* but ongoing deliberation over *values* and the complex interrelation of facts and values as well.

So, a value judgement by the decisionmaker, canonised, is unlikely to play the role. What if the weighting is an empirically justified estimate of effects on some sort of wellbeing metric? If it is an empirically justified estimate of effects on a measure of wellbeing, we will have to specify what we mean by welfare. If we just mean a particular type of subjective wellbeing, the method converges a little with ours though unlike ours is focused on a single kind of welfare. If it then takes on our methodological changes to standard cost-benefit analysis- (e.g., using direct estimates of the effects of a policy on happiness, where possible, rather than mediating through weighted willingness to pay) it will converge with our method still further.

If it is still intended as a decision algorithm on the basis of aggregate welfare, it will collide with all those who do not agree with utilitarianism, and all those who do not agree that our chosen wellbeing metric is equal to welfare in the ethically relevant sense. If, on the other hand, it is only intended as informational, it will converge once again with our method and starts to look much like it.

8.4 Other approaches to the value problem in applied welfare economics

8.4.1 Nyborg and Spangen on democratic theory and cost-benefit analysis

Nyborg & Spangen (2000) provide a democratic critique of cost-benefit analysis understood through the light of Dahl's (1989) theory of democracy. Their critique extends to both weighted and unweighted cost-benefit analysis, and to cost benefit analysis understood as decision making criterion, as well as cost-benefit analysis understood merely as informational. In this section I will be arguing that while what they outline is a powerful critique of standard forms of CBA, it does not tell against SWBEE.

Nyborg and Spangen consider a number of ideal features of a democracy owing to Dahl (1989):

1. *Effective participation*
2. *Voting equality at the decisive stage*
3. *Enlightened understanding*
4. *Control of the agenda*.

Roughly speaking, effective participation means the ability of citizens to participate in discussion equally. Voting equality at the decisive stage means that, in the final instance of decision making, each citizen's vote is counted equally. Enlightened understanding means that each citizen understands the issues and positions under debate. Control of the agenda means that it is the demos or citizenry that determines which issues and potential decisions are debated. Nyborg and Spangen find that CBA used as a decision-making criterion does not succeed in relation to any of these criteria, although it perhaps does not *fail* enlightened understanding per se. Even if cost-benefit analysis is intended purely as information, it still fails in relation to the *effective participation* and *enlightened understanding criteria.*

I give the following table, based on extrapolation and interpolation of Nyborg and Spangen's work. It is neither quite an interpretation of their work, nor something original to me, but rather intermediate between the two.

Table 3: Problems associated with CBA corresponding to their different type and method of use

Cost-benefit analysis	If the CBA is unweighted	If the CBA is weighted

If the CBA is Intended to make decisions	Unweighted cost-benefit analysis used as a decision rule represents a highly controversial value function.	Weighted cost-benefit analysis used as a decision-making rule neglects the plurality of value systems in a democracy by instituting a single SWF to make policy decisions.
If the CBA is intended to be purely informational	Unweighted cost-benefit analysis used as a source of information suffers from a problem- it is unclear what the information *means.* Such information may be legitimate to gather but will be of somewhat limited relevance to the decision-making process. The existence of a surplus in the cost-benefit analysis does not necessarily imply that the policy is a *real* potential Pareto improvement, due to the deadweight costs of taxation, political barriers etc.	If we use weighted cost-benefit analysis purely as a source of information, it is very unclear what the information we are given is? What is the aggregated weighted willingness to pay representing? Is it simply telling us how some hypothetical person with a particular value function would feel about a policy? Certainly, this is interesting, but why privilege that hypothetical person and their value function in our analysis?

One of Nyborg and Spangen's (2000) most decisive criticisms is against the univocality of social welfare function-based approaches. Such approaches often talk about **the** values of **the** social planner. In a democracy though, decision making involves collectives at every level. Even a representative government is not an elected dictatorship internally. The government as a whole is also not merely meant to represent its own values, but also the values of the citizens overall.

Would Nyborg and Spangen's critique still apply to our approach? After all, SWBEE can take on the form of cost-benefit analysis at times, as we have discussed earlier in the thesis. It seems to me no. To see why, consider the fourth box, which our approach is most like, weighted cost-benefit analysis intended to provide information rather than for direct use in decision making. We say that the fourth box is most similar to our approach because as we have discussed, weighted cost-benefit analysis can be a way to approximate an estimate of psychometric subjective-wellbeing if it is built with the correct assumptions. The critique of the option represented by box four that we give in our table- that any weightings will represent only parochial ethical judgements, does not apply. Rather the weightings represent an attempt to estimate a real quantity, of interest to both the public and to elected and appointed decision makers- degrees of subjective wellbeing. effective participation, voting equality at the decisive stage, enlightened understanding and control of the agenda are no more threatened than they would be by a project that aimed to map the old sewer system before a decision was made about a new sanitation project or calculate how many watts of power are needed and where they will be consumed.

8.4.2 Gul & Pesendorfer's plea for positive welfare economics

One way to avoid the values problem in welfare economics is to insist that welfare economics should simply not be concerned with problems of values or policy decision making at all. Gul & Pesendorfer (2008) argue that welfare economics is a positive science that considers:

A) The degree and method through which existing institutions aggregate preferences.
B) Their persistence (or non-persistence) in light of their performance as preference aggregators.

As opposed to an approach which considers the psychological consequences of institutions and policy choices on wellbeing, or a normative appraisal. This is in the context of a paper more broadly criticising the psychological turn in economics.

If I were to debate Gul & Pesendorfer's first order claim that this is what welfare economics consists in, we would face an interesting question: what *defines a subfield?* Especially when there is such radical divergence as here- with one side seeing welfare economics as an applied discipline, and the other seeing it as theoretical. Perhaps the social goals it is most commonly used for define it? Perhaps the goals of most people engaged in the subfield? Perhaps what the subfield would be under some hypothetical "rational" optimisation of the field as a whole, or even of the sciences as a whole? Perhaps the field's history? Certainly, Gul and Pesendorfer's confident declarations about what welfare economics is would have seemed strange to historic figures as diverse as Robbins (1932), Pigou (Pigou, 1903, 1912, 1920/1932, 1951) and Little (1950).

A wiser approach though is to dodge the argument. There is no need to get into a semantic debate about what "welfare economics" really means, nor is there a need to get into an axiological argument about which form of study would be more valuable. Gul and Pesendorfer's understanding of the field, and my understanding of the standard approach to the field are quite different from each other, and neither approach excludes the possibility of a different body of scholarship exploring the other approach. Gul and Pesendorfer's welfare economics as science of aggregation and non-aggregation of preferences by institutions can continue as before, as can welfare economics in my sense, and within it, SWBEE.

8.5 Conclusions for chapter eight

In this chapter we have argued that cost-benefit analysis (excluding SWBEE), weighted or unweighted, cannot, at least not as it is normally conceptualised, escape the problem of value sectarianism in economics. Attempts to conceptualise CBA as just information provision fails for several reasons. Most notably *it is unclear what the value of the information provided is* because Kaldor-Hicks improvements need not represent real potential Pareto improvements given the deadweight loss of taxation and political barriers. Attempts to defend CBA through value confinement strategies- the conceptualisation of CBA as only involving modest value judgements that almost everyone would agree with due to democratic legitimation- likewise fail.

9 Psychometrics vs self-report in measuring desire satisfaction

9.1 Introduction

In this thesis we have defended the use of psychometric techniques to measure wellbeing, as a source of information about the welfare effects of policies. Naturally, this prompts a comparison with the strengths and weaknesses of alternative methods of trying to get at welfare, such as revealed preferences. In this chapter, I will consider arguments concerning the best way to gather information about welfare (specifically desire satisfaction) in an economic context: self-report of desire satisfaction versus revealed and stated preferences defined in monetary terms. Our conclusion is that in the main, self-report is the stronger paradigm for getting at welfare even given the assumption of a desire-satisfaction account of welfare.

There are a number of traditional critiques of revealed preferences as a means of determining an agent's degree of preference satisfaction (and the conceptually twin rational choice model of behaviour), as well as debates and puzzles which have been raised within the paradigm. These include, but are not limited to: triviality and unfalsifiability (Hodgson, 2012)**,** and lack of attention to irrationality and constraints on human rationality (Tipoe et al., 2022)**.** Our purpose is not to critique the revealed preference paradigm in economics *per se*, which has its own particular uses. Rather our aim is to assess whether, on a particular theory of welfare, revealed preferences and stated preferences, or psychometrics is the better method of getting at welfare.

Does this have anything to do with the interpersonal aggregation problem that we have spent the last three chapters considering? Yes. If we want to work out whether a change will be good for a population on the whole, and that involves summing between advantages to some people and disadvantages to others, we had better be sure that we have the best possible method for working out what is to each individual's advantage and to what degree. If we were considering changes that prima facie helped everyone- then, a relatively simple model like "more money is better" might suffice, but for adding up complex and interpersonally varied combinations of gains and losses, if we accept preferences as the baseline of welfare, we must be sure that we understand individual preferences and how they vary.

By Revealed and Stated Preferences (RSPs) we mean an inference about how much an agent value something from a real or hypothetical buying or selling behaviour. This groups different methods into one overarching category of behaviour- real and hypothetical- defined in financial terms. RSPs particularly mean the kind of revealed and stated preferences that are suitable to be fed into a CBA.

In assessing whether RSPs or psychometric measures are the best indicator of welfare, we face a barrier. We do not wish to take a position on what welfare is. Our solution here is to grant the best possible case to our competitor so we assume the simple desire satisfaction theory is true. Welfare is getting what you want, specifically getting what you want for its own sake, and not just as a means to an end. We will be arguing that, on the whole, psychometrics is a better measure of whether people have what they want than revealed (and/or stated) preferences approaches. Where we refer to welfare without modification in this chapter, we mean the degree to which the agent's desires are satisfied (unless context indicates otherwise). Where we refer to subjective wellbeing in this chapter without modification, we mean the degree to which the agent regards their own desires as satisfied (unless context indicates otherwise).

9.2 Our Reasons for preferring self-report to revealed and stated preferences techniques

9.2.1 The absence of Psychometric measures of preference satisfaction

There is no available psychometric measure of preference satisfaction *per se*. The closest that exist are measures of life satisfaction and of domain satisfaction.

As I have previously stated I see no reason, in principle, that a self-report measure of desire satisfaction couldn't be created. I suspect, strongly, that a question like "to what extent are the desires you care about most satisfied" - a measure of desire satisfaction- would be correlated very strongly with a question like "to what extent are you satisfied with your life", a measure of life satisfaction. To be sure, there is, as yet no such measure. However, the viability of life satisfaction measures, and of domain satisfaction measures gives us reason to be hopeful.

In this section, I will try to argue this psychometric measure of total desire satisfaction- at present yet to be created- would be a better measure than RSPs approaches to desire satisfaction wellbeing. Obviously, to argue that an as yet non-existent method is better than an established method, one must bear a weighty burden. I believe, though, that we will discharge this burden. At the least we strongly motivate the creation of psychometric measures of desire satisfaction.

9.2.2 The stability problem for behaviourally defined welfare: Multiple ways to define willingness to pay give different results

Consider the ambiguity and multiplicity of behaviour in relation to willingness to pay studies. We could use multiple methodologies to value, say, clean air. It would be nice to think of all of these as an approximation of an underlying value "how much people are willing to pay for clean air", but there are multiple contexts in which people make such valuations, and multiple ways the question of how much one is willing to pay can be framed.

The chapter on water valuation in Cost-Benefit Analysis by the Asian Development Bank (2013) captures this nicely. A sample question is given as:

> *Suppose your household receives 24-hour water supply service (reliability), with water that is safe to drink from the tap (quality), accurate billing of the water with reliable and responsive customer service (customer service quality) by a private service provider (service provider). But your household would have to pay higher water bills (payment mechanism, cost).*

With the stated ideal being that one should elicit valuation of water only after having provided accurate information on its costs and benefits.

A question as complex as this is presumably especially vulnerable to framing effects. Is this question truly intended to uncover one underlying singular degree of preference, pre-existing the question, the true willingness to pay of the subject? If so, does such a value exist? Why prefer this value to a value extracted some other way? For example- the value the subject would give after they have been thoroughly educated on the advantages of clean water. Is the approach normative- in the sense that what truly matters is the agents informed preferences, so we should try to find out what they are? If so, there will be endless difficulties in trying to operationalise such an informed preference in the real world-. One problem with such an idealization process is that it can be hard to know what the natural place to stop is- to fully understand the benefits of clean water would take weeks.

We can frame this in two ways- a more psychological approach, and a more conceptual critique. At a more psychological level, there is a literature (Sippel, 1997) that suggests that people’s conception of how much things are worth paying for can be quite unstable- how reliable then are willingness to pay figures?

In relation to contingent evaluation, the literature identifies a problem of scale insensitivity (Baron & Greene, 1996; Carson & Mitchell, 1995; Desser et al., 2013; W. Desvousges et al., 2016; W. H. Desvousges et al., 2010; Diamond & Hausman, 1994; Fetherstonhaugh et al., 1997; Gyrd-Hansen et al., 2012; Kahneman, 1986; Kahneman et al., 1999; McFadden & Leonard, 1993; Norinder et al., 2001; Olsen et al., 2004; Søgaard et al., 2012; Whitehead, 2016), for a review see Rakotonarivo et al. (2016).

Fetherstonhaugh et al. (1997) for example find that willingness to pay to save lives is roughly proportional to the log of the lives saved. This implies inconsistent and/or bizarre behaviour, for example, the willingness to pay to save two groups of X people may greatly exceed the willingness to pay to save a group of 2X people. The “warm glow effect” for positive moral actions, which may not be linear in good done, might well explain this result (Baron & Greene, 1996; Kahneman & Knetsch, 1992; Whitehead, 2016).

With regard to the more conceptual critique, we often want to explore willingness to pay for public goods which are not directly bought by individuals. Hence there is an attempt to find proxies. To work out how much people are willing to pay for clean water for example, we might look at how much they pay for bottles of water, or we might look at how much they are willing to pay for home filtration systems, or we might look at how much they say they would be willing to pay in extra taxes (I believe this example was first given to me in personal communication with a colleague, but I do not remember which). Which, then, is the ‘true’ valuation? In all cases, we have the same two worries 1. is there actually a “real” value which is the agent’s willingness to pay for X- lurking behind the different methodologies which give different results? Or is this a fiction? 2. If there is such a real value, shouldn’t we be applying the full range of psychometric methods- e.g., validation etc. to try and dig it out? Hence, we return to a very similar point to one we made earlier in the thesis- either there is a real valuation underlying appearances in which case we should be applying psychometric techniques to help get at it, or else there are just different figures with no ‘underlying reality’- in which case, on what basis are we to choose between them?

In terms of psychometric style investigations of what the willingness to pay framework “really” captures there is an existing literature, although not as comprehensive as one might like. Most investigations into this area treat actual willingness to pay as a baseline and examine the degree to which hypothetical and actual willingness to pay line up. Kanya et al. (2019) found that hypothetical willingness to pay diverged very significantly from actual willingness to pay, and, among other things, the relation between hypothetical and actual willingness to pay was good specific . Kahneman et al. (1993) find that willingness to pay is “better described as expressions of attitudes than as indications of economic value “.

There is a history of research in behavioural economics into how seemingly extraneous factors can affect people’s perceptions of price and value. For example, willingness to pay does not equal willingness to accept, though in theory these values should be identical for homo economicus (Ginsburgh, 2017; Gopavaram et al., 2021; Martín-Fernández et al., 2021). Anchoring effects influence both willingness to pay and willingness to accept (Simonson et al., 2004)

It is worth mentioning the work of Frey et al. (2017) who studied many self-report and behavioural measures of risk and found that different behavioural methods of measuring risk preference had little correlation with each other, compared with self-report measures and low test/retest reliability compared with self-report measures. Frey’s comment on this matter is revealing:

> Because behavioural measures differ widely in their choice architectures (in terms of their presentation format, specific instructions, or framing), they may be designed, perhaps unwittingly, to elicit different queries and cues and thus trigger diverging constructed preferences. Moreover, it has previously been observed that specific task manipulations can override traits. (R. Frey et al., 2017, p. 8)

In summing up there is a fundamental problem with RSP measures- there are multiple ways to take a measure of RSPs, and there are multiple ways to extend them to allow for things like interpersonal comparison. In each case we can pose the following dilemma, either these definitions are stipulative, in which case it is hard to see how we can choose between them, or each definition defines a different way of trying to estimate some “true value” of an agent’s desire for X. But this “true value model” prompts two separate questions, viz:

1. Does such a true value even exist?
2. If such a true value does exist, we should presumably bring to bear the techniques of psychometrics in trying to estimate it, especially validation. Since validation has already been conducted for many psychometric approaches, but has not been conducted for revealed preferences approaches, doesn’t this give us a reason to prefer psychometric measures, at least in the interim- albeit noting the caveats we have mentioned around the measurement of the general desire satisfaction?

What about a psychometric approach? Using our example of water there are several ways to estimate the effect on desire satisfaction wellbeing of introducing clean water. The most direct method would be to observe changes in overall desire satisfaction via our hypothetical desire-satisfaction psychometric instrument as a quasi-experiment, after clean water is introduced across multiple areas.

9.2.3 Advantages of self-report: anti-paternalism and self-appraisal

We’ve reviewed some arguments about the weaknesses of behavioural measures for assessing preference satisfaction, now let us consider an advantage of self-report. Self-report allows the agent to assess their own overall level of desire satisfaction (or whatever our preferred welfare metric is). This has value for a number of reasons. It might have ethical value for some because it can be regarded as granting greater autonomy to the agent to determine for themselves their own degree of wellbeing. Of course, it might be objected that the RSP account also allows in a sense for the agent to determine their own well-being, but to see the limitations of this approach, consider the unhappy Croesus:

The Unhappy Croesus: The unhappy Croesus has superficially everything he has aimed for in life that can be quantified or measured in terms of external success, yet he still feels both deeply unhappy and deeply discontented and unfulfilled. A RSP approach- construed as a way of measuring preference satisfaction welfare- will assign him a high welfare, yet he believes himself to be doing poorly, however one chooses to measure it. There seems to be something *paternalistic* about insisting that he has a high level of desire satisfaction or welfare in any sense.

Hence, we have turned one of the strongest arguments for RSP back around on itself- that it allows the agent to decide for themselves what is good for them, by countering that it does not allow the agent to decide whether things, in general, are going well for them. Of course, this argument will only work for a certain type of psychometric assessment- those that involve the participant making direct statements about how happy, satisfied etc. they are - but many forms of the assessment of wellbeing in this field of psychometrics are just this- especially the one item scales often used for economic work. Even multi-item scales without transparent scoring allow participants a degree of autonomy in describing their own wellbeing.

Beyond ethical value, allowing the agent to assess their own wellbeing might also have greater empirical value for a variety of reasons including: 1. the respondent may have access to a wider range of information than would be used in a cost-benefit assessment of their situation. And 2. Subjective wellbeing/subjectively appraised desire satisfaction may better capture costs and benefits that the respondent feels reluctant to put into monetary terms

9.2.4 Advantages of self-report: Self-report measures what we care about intrinsically; RSPs often only measures what we care about extrinsically

One argument in favour of RSP as a measure of desire satisfaction might go as follows:

1. Desire satisfaction is not in the head. An agent can be mistaken about the extent to which their desires are satisfied.
2. On the other hand, through revealed and stated preferences measures and the observation of buying, selling and tradeoffs, we can directly measure the degree to which we get what we want.
3. Ergo RSP measures are superior.

In other words, RSP methods of determining welfare are not a purely mental phenomena, and this is a good thing because welfare is not a purely mental phenomenon, as shown, for example, by Nozick's experience machine argument.

The problem with this argument is that there is an ambiguity in the concept of preference. We have both intrinsic and extrinsic preferences, presumably it is intrinsic preferences whose fulfilment carries weight on the desire satisfaction theory of wellbeing. Existing theories often do not spell out the matter precisely but seem to suggest the locus of concern is intrinsic desire (Roger Crisp personal communication 2023, Fletcher (2016, p. 28)). RSP measures (with some exceptions) only directly measure the satisfaction of extrinsic preferences, this is because, for the most part, things that can conceivably be bought and sold or otherwise measured by RSP techniques are things we only value instrumentally.

This point complicates the defence of the RSP approach. An agent's happiness, the argument goes, might be based on a misconception, whereas their RSP computed welfare depends on objective states in the world- did they get it and how much were they willing to pay for it? However, if I am right about RSPs- and they usually do not relate to what is intrinsically valued- RSP measures of welfare are just as mediated through (potentially mistaken) beliefs as subjective-wellbeing assessments of one's own desire satisfaction. It's just that, in the case of revealed preferences, these potentially incorrect mediating beliefs are beliefs about what will fulfil my intrinsic desires.

Consider that almost all market behaviours are motivated by *extrinsic desires*. One does not desire to own a certain car, rather one desires the comfort and convenience a car provides. Even if one wants to have a certain car for the sake of showing off, one does not so much desire to *own* the car as to be *known* as the owner of the car. A way to make a similar point more familiar to economists would be Lancaster's (1966) theory of consumer behaviour. Somewhat loosely, Lancaster's theory of consumption activity implies that people value attributes of commodities and activities one can perform with commodities rather than commodities in and of themselves. People do not wish, at least for the most part, to be the formal legal owner of a diamond for the sake of being the formal legal owner of a diamond. They value it for the properties it has, and the activities they could use it for- admiring it, showing it off and so on. But when one buys something, one will always be uncertain of the activities one will have an opportunity to perform with it (to take the most extreme case, could die of a heart-attack at any moment) and there will often even be a degree of uncertainty as to the properties of objects.

Measures of desire-satisfaction subjective wellbeing, as opposed to revealed preferences, may well be more closely related to these intrinsically desired outcomes than factors like willingness to pay for a given commodity. Suppose Chan badly wants a partner, and as a result he spends a lot of money on expensive fashion in the hope that it will make him more appealing. On a revealed preferences analysis, his coming into possession of a discount card giving a lifetime 50% discount on Gucci products might register as quite a boon to him, yet the truer story about whether this has helped Chan achieve his intrinsic desires will not be told by such instruments, yet simply asking to what degree he has what he wants, or asking about specific domains like romance and companionship, will reveal the effect.

Yew-Kwang Ng (personal communication 2023) gives an argument similar to ours but focused on happiness. Ng notes that, as an empirical matter, most people value happiness very highly. He fashions this into a critique of standard welfare economics. We should be sceptical of measuring what people want only instrumentally, rather than measuring their ultimate desires. It is doubtful that people very often value the kinds of things they pay for at all *except instrumentally*, thus willingness to pay is a second order measurement. Ng's point does not have to be tied to a case for basing welfare economics around happiness. Some might argue that what people really want is happiness- a state that is obtained in part through material things, but only indirectly, in a mediated way. Others might argue that human beings are deeply committed to projects- friendships, families, careers, etc., and they want to achieve what they see as success in these areas in a way which is at least somewhat autonomous from a desire for happiness (this might be more compatible with a eudemonistic, or alternatively, a desire satisfaction, theory of happiness). In either case, willingness to pay is a distant estimate, ex ante, of how much real value will be extracted from something. Measurements of subjective wellbeing are much closer to a direct measurement of what people desire ex-post.

Ng's point ties into our earlier scepticism about whether there is really any such thing as an agent's true, context free, willingness to pay. If what agents really want is happiness, or the fulfillment of various complex desires which are mediated by many factors beyond the purchase of objects, further doubt is thrown on the idea that agents have a singular real willingness to pay for anything. Willingness to pay reflects strategic appraisals of how much a thing in a context will help us get what we want- and such appraisals will vary for endless contextual and perceptual reasons, rational and irrational. What we are truly trying to achieve may be many, many leagues downstream of our purchasing choices or other preferences that can be easily read off behaviour.

Research on the exact list of things people happen to intrinsically desire for their own sake proved surprisingly difficult to find. I would propose though the following list of intrinsic ends as a start for inquiry, as things which art, literature and conversation would lead one to believe are common intrinsic ends, with the proviso that individuals can and frequently do vary on their intrinsic ends and no judgement is intended on the desirability of any of these:

Being happy

Pleasure and specific pleasures

Freedom from mental and physical suffering, or its minimisation to a level thought appropriate

Having a long life

Forming meaningful relationships, loving and being loved

- *Including family relationships*

- *Including friendships*

- *Including romantic relationships*

Appreciating beauty

Social esteem

- *Respect*

- *Fame/popularity*

Acquiring knowledge and skills

Having children

Raising children

Upholding one's ethical obligations

- *Including to family*

-Including to friends

-To the world at large

Upholding one's religious obligations

Conformity to certain valued social roles valued by the agent (e.g., gender, social class etc.)

Personal achievements

-Career

-Financial

-Aesthetic

-Physical

-Moral

-Social

-Religious

9.3 Considerations for future research

Most obviously, our work here gives reasons for the creation and validation of a psychometric measure of desire-satisfaction.

Beyond this, we can present ethical and philosophical arguments here for the superiority of self-report to revealed preferences measures, but there is an urgent need for an empirical assessment of these questions. The most natural way to compare revealed preferences to self-report measures empirically would be to determine which set of measures is better at predicting *validation measures*- things that we think, for conceptual reasons, should be related to wellbeing. For self-rated desire satisfaction, important benchmarks of comparison would include:

1 Happiness- we would expect desire satisfaction to correlate with measures of happiness.
2 Life satisfaction- Satisfaction with life and overall satisfaction of desires should be closely related.
3 Other rated desire satisfaction- others assessment of the degree to which one has what one wants in life.
4 Various forms of Domain satisfaction.

It would then be a simple matter to see whether changes in revealed preferences measures, or changes in a self-rated measure of desire satisfaction, correlates with these outcomes better. For example, a new bridge is built near Isabella's house, meaning that she can now visit town and better enjoys dwelling in the suburb in which she lives. Does the change in the amount of money she would need in order to be willing to accept moving away from her current suburb, or the change in her self-rated desire satisfaction, better predict the above outcome measures?

A review of the literature has not found attempts to measure desire satisfaction psychometrically, let alone such a comparative study.

There remains a question- we've argued in this piece that a hypothetical approach of measuring desire satisfaction via self-report would be superior to revealed preferences as a measure of desire satisfaction. However, in the absence of such a direct method in the literature, are *already existing* measures such as measures of life satisfaction, measures of domain satisfaction and even measures of happiness superior to revealed preferences measures? What you think the answer is will depend, among other things, on the weight you place on the arguments we've made in this chapter, and how closely you think that general desire satisfaction is correlated with being satisfied with your life.

Finally, on a philosophical note, I would like to express a certain, tempered scepticism about the desire satisfaction theory on the basis of considerations we have discussed. It seems to me that defenders of the desire satisfaction theory face a challenge: desire may be more contextually unstable than is ordinarily realised. There may be no such thing as your pre-existing strength of desire for X, outside of a context of presentation. Of course, the desire satisfaction theorist has strategies available to deal with this challenge- for example, a degree of idealisation. The desire theorist might move discussion to what the agent would state as their desires under ideal conditions- many desire theorists already perform a degree of idealisation for other reasons anyway. Also, it may be the features of context specificity we have discussed can be limited in some way, but this seems like a topic well worth the desire theorists time to grapple with.

9.4 Conclusions for chapter nine

We have argued that even on the terrain of the desire satisfaction theory of wellbeing, and even though there is not currently a psychometric instrument for measuring desire satisfaction, it is likely that a psychometric measure would be superior in measuring desire satisfaction to revealed and stated preferences measures. Revealed and stated preferences are contextually unstable, and it is unclear what seeking the 'true' value across multiple different valuations amounts to. Psychometric methods allow the subject to define their own sense of overall wellbeing and desire satisfaction. Revealed and stated preferences measures mostly provide information about *instrumental* preferences, whereas a psychometric measure can more easily track *non-instrumental* preference fulfilment.

10 Conclusions, future research and selected topics

10.1 What about wireheading?: Could the pursuit of happiness lead to ethically abominable results?

Wireheading is a process whereby a person's reward and/or pleasure centre is directly stimulated. In a more metaphorical sense, the term is sometimes used to refer to any process that directly creates positive feelings and pleasure technologically yet gives its experiencer only a limited kind of life. In this sense, for example, if the government gave certain people a kind of drug that worked indefinitely without side effects, leaving the subject in a nearly comatose 'blissed out' state, that would be wireheading. Although some people might support wireheading if it were possible without causing social collapse, the term is usually taken to have a negative connotation. Sometimes it is argued that utilitarianism of this or that sort (and other welfarisms such as prioritarianism)- endorses wireheading, or would if it were practically feasible, and that this is a blow against it. For example, Baujard (2013) alluding to a discussion of Layard (2005) suggests a worry that the measurement of happiness, as an indicator of social good, would give decisionmakers a reason to support the mass distribution of tranquilizers- an allusion to the wireheading problem.

In the context of normative economics, this worry can be seen as an application of an old worry that a lack of concern for choice will lead to an intolerable paternalism. Because it is a kind of synecdoche or extreme version of a larger worry common among economists about choice and paternalism it's well worth engaging with, even if only briefly, in a thesis on applied welfare economics.

I take this objection seriously -despite our approach not being explicitly committed to any theory of the good life- because it seems to me an objection against all philosophies that are in some sense descended from utilitarianism, particularly practical forms of quasi-utilitarianism intended to act as a doctrine for operating government – a form of utilitarianism sometimes called "government house utilitarianism", to distinguish it from a more philosophical utilitarianism with thick ethical commitments.

The approach we have championed in this thesis is certainly not committed to utilitarianism, but it shares with utilitarianism a desire for governance (partly) based on data about what is good for people, thus it is in some sense related to the broad family of quasi-utilitarian philosophies. Baujard's (2013) concern about wireheading is partly a concern about the consequences of making happiness a metric. One is reminded of Goodhart's law: "When a measure becomes a target, it ceases to be a good measure" (Strathern, 1997), and the approach we have outlined in this thesis definitely has, at least prima facie, the potential to encourage the use of happiness as a target in cases where this would be inappropriate.

My most fundamental reply is that our project is founded on the study of *subjective wellbeing in its diversity of forms* and to the extent that our method implicitly values anything, it is all those many forms. It may be true that wireheading would increase hedonic wellbeing- and perhaps also life satisfaction and/or desire satisfaction, but we support a plurality of different measures of wellbeing. Even if you take it for granted that wireheading would be a bad thing that hedonic measures on their own would tend to support, the total data we support gathering is not going to univocally commend wireheading. Properly designed measures of things like eudaimonia and capabilities will not support it- assessing as they do variables such as functioning, connectedness, striving and community participation. This is another reason to defend the broad approach we have supported in this thesis- measuring effects of policy on a variety of different concepts of subjective-wellbeing. The sheer range of different types of measures- including measures that work on quite abstract concepts like flourishing- makes it unlikely that any wireheading policy would show up as an across the board good, using the plural metrics we are championing.

But this raises another question. The approach we have developed in this thesis works best for creating consensus if different kinds of subjective wellbeing, and different kinds of welfare according to different theories of welfare, are correlated. What happens though under cultural or technological conditions under which they might massively diverge, such as the widespread adoption of wireheading? A cultural conversation about what kinds of happiness were most important would be necessary, and we might hope that the information provided by our method would help with this proposed conversation.

Another closely related strategy of defence against the objection owes to Goodin (1995) and attempts to defend a kind of utilitarianism. What I argue here is a combination of Goodin's views and my own and mixes exegesis with extrapolation. In *Utilitarianism as a Public philosophy* Goodin (1995), as one might expect, suggests that utilitarianism works best as a public philosophy. Most of the various thought experiments marshalled to try and undermine utilitarianism take it out of its original conception as a philosophy for governing public life, and move it into the private sphere for which it was not primarily made. What I take to be important about this approach is that Goodin does not see it as necessary to defend utilitarianism in all possible situations. Rather Goodin, argues that utilitarianism works well in a particular space (contemporary century policy making), against the approach to ethics which sees the role of ethics as trying to work out a set of rules that is defensible in all possible situations.

Something similar could be said about the wireheading case, both as it reflects on utilitarianism proper, and as it reflects on the distant, indeed only slightly related, relative of public utilitarianism we defend here. What we are outlining, particularly in chapter 7 on values, is not a morally valid procedure for all possible conditions, but an attempt to navigate through actually existing ethical views and conditions to allow the welfare economist to say something interesting about policy choice without running aground on ethical sectarianism. This is not the kind of project that can be shipwrecked by a merely possible shoal, and for the most part, concerns about wireheading are hypothetical- and, leaving aside bio-psycho-social plausibility, are politically infeasible.

Now I grant that there is a portion of this problem which really applies today. One might argue for example that propaganda that disseminated comforting but untrue propositions might receive support from utilitarianism under certain circumstances, even as it dulls the minds of the public, but I would suggest that propaganda this encompassing and effective as to shut out alternative unpleasant thoughts is relatively unlikely in contemporary societies. Nonetheless, we can't rule out the possibility that attention to a few subjective-wellbeing metrics or perhaps even all possible subjective-wellbeing metrics might exclude very important facts from consideration in a way that could support repugnant policy recommendations. Here our main line of defence is to simply suggest, again, that we are just providing information and that it is important our approach *never* claims to provide *all* the available ethically important information. This is our strongest and most final line of defence against any argument that tries to show our approach is a downward slope to endorsing deceptive propaganda or even wireheading.

But we have additional lines of defence. One of the most important is a same boats strategy- *cost-benefit analysis is also in the same boat of only being able to provide partial information, and that partial information may prima facie support monstrous deeds.* We are proposing ideas not in a vacuum, but as an alternative to traditional strategies in welfare economics. Take cost-benefit analysis. Who can deny that cost-benefit analysis, if pursued single mindedly, could lead to abominable things? If wireheading is going to be urged against us using far flung situations, let us think about some things that willingness to pay could lead to as a metric. Swift (1729/1997) in *A Modest Proposal For preventing the Children of Poor People From being a Burthen to Their Parents or Country, and For making them Beneficial to the Publick* makes a satirical argument that Irish babies should be farmed for cannibalistic consumption by the English. Is it not entirely possible that this could pass unweighted cost-benefit analysis? Babies are credit constrained so the willingness to pay of babies is notoriously low, and if their parents are in on the scheme as Swift imagines, they will not object either.

Our suggestion that effects on the overall level of subjective-wellbeing be presented alongside predictions about the distribution of subjective-wellbeing helps foreground the fact that this is not the kind of process that outputs a single numerical judgment of social welfare. In this regard it runs ahead of CBA in the race to avoid endorsing monstrous deeds.

10.2 Is weighted cost benefit analysis inefficient?

There is a tradition (Harberger, 1978; Y.-K. Ng, 2003b; Weisbach, 2015) in welfare economics that holds that weighted cost-benefit analysis is inefficient, as income inequality reductions carried out via applying weighting to project appraisal, or analysis of law (Kaplow & Shavell, 1994, 2000, 2002) Is allegedly less efficient than achieving the same reductions via the tax and transfer system, or more generally by some other policy mechanism. Our proposed method can be seen as part of the family of weighted cost-benefit analysis, so despite our method's differences from other forms of cost-benefit analysis (weighted or otherwise), we will endeavour to reply to the challenge.

As a defender of SWBEE, I have a lesser burden to discharge than other authors who appeal to weighted cost-benefit analysis. Our general orientation, as outlined in chapters 4 and 7, is that SWBEE analysis is intended to be *informative* rather than act as a decisionmaker. Thus, all we are arguing is that there is value in presenting an analysis of the first order effects of a policy on a proxy for welfare- subjective wellbeing. Procedures can be developed for considering labour supply effects separately to that, to independently be weighed and considered in the agency decision making process. It is not incumbent upon us to show that decisions must be made or even advised according to a weighted cost-benefit analysis algorithm because we are not proposing decision making or even decision advising by such an algorithm. Nevertheless, the argument that only unweighted cost-benefit analysis leads to coherent decision-making complicates our method and so we must reply.

Notably, it is not true that contemporary cost-benefit analysis is usually strictly unweighted, even in jurisdictions that nominally prefer unweighted cost-benefit analysis, Hammitt (2021) makes the point that:

> Concerns about distributional equity influence BCA as practiced. It is common to use the same monetary value of a good for all individuals, ignoring differences associated with income or other characteristics. (Hammitt, 2021, p. 81)

For example, suppose that we decided to value giving clean water to a community in which water is difficult to access due to an infrastructure failure. Market failure means that clean water is not currently available in the community, hence we have to value it *de novo* perhaps on the basis of stated willingness to pay. A cost-benefit analysis would value the water distributed at some rate, but it would be unlikely to include different prices for the rich and poor citizens receiving water, despite their very different maximum willingness to pay. In other words, Hammit reminds us we are very often already using a form of weighted cost-benefit analysis implicitly, making more alarmist predictions about the potential effects of adopting weighted CBA hard to believe.

The argument we are concerned with has a long history- it is present to some degree in Harberger (1978), Ng (2000) and Weisbach (2015) among others. The argument is that weighted cost-benefit analysis can be misleading in a way that unweighted cost-benefit analysis is not, because the distributional effects might reduce labour supply or investment. Effectively, weighted cost-benefit analysis raises the progressivity of the total effect of government action in an ad hoc manner, causing suboptimal policy results either because there are more efficient mechanism of achieving redistribution, or because the add-hoc nature of these effects means that fine tuning the degree of redistribution is difficult. We will call this the incentive effects concern.

Consider two countries, A & B. Both countries have a progressive income tax system with the same schedule of rates. In country A, the weight the government places on your needs in deciding where to provide its services is proportional to your total income. In country B, the amount of spending you receive in public goods provision is equal no matter what your income. Treat public good provision as part of income. The effective incomes of the rich will be higher in country A than country B, and therefore, the progressivity of government policy is higher in A than B.

It is certainly true that the distributive effects of policy need to be considered, and that, at least in principle, the overall effects of a policy that greatly redistributed resources downward via application of weighted CBA could be negative due to labour disincentive effects. We can imagine, for example, someone who contemplates becoming a medical specialist, but decides it is not worth the effort when he sees that the surroundings of the rich (e.g., the roads and parks near their houses) are not particularly opulent. I will leave it to the reader to judge whether this is plausible. In what follows, I will aim to undermine the incentive effects concern as a universal argument against weighted CBA.

10.2.1 The lump of redistribution fallacy

If redistribution through regulations or infrastructure investment were exactly like redistribution through taxes and transfers in the magnitude of unwanted incentive effects, when producing an equal inequality reduction, or strictly worse, this would certainly be an argument against redistributing in this way. We however have reason to doubt that this is the case.

The most general problem we have with the incentive effects concern is the assumption that the labour supply effects of the well to do receiving, say, more parks, are equivalent in behavioural effect to the labour supply effects of the well to do receiving lower taxes- because it fails to account for possible differences in salience between these redistributive mechanisms, and then the extent to which they may produce unwanted incentive effects.

To assume that the labour supply effects are the same makes very strong assumptions about economic rationality and information availability and processing ability. The effects might well be weaker, because they are less obvious than a change in the bank balance, though they might even be more obvious, though this seems unlikely.

One possible solution to this problem in the case where we struggle to quantify incentive effects is to estimate the effects of the policy on welfare across a range of plausible estimate of the relevant incentive effects. A complimentary approach is to attempt to improve estimates of the incentive effects of redistribution carried out via weighed CBA, which would narrow the range of plausible effects.

This is an area that cries out for empirical study. Does building roads at the same rate for the rich and poor reduce labour supply to a degree that is similar to an equally increase in tax progressivity? We should try to find out. There are available sources of data that could produce estimates based on natural experiments. As we mentioned earlier, Hammitt (2021) notes, much cost-benefit analysis is already weighted in practice. Suppose as in our example discussed earlier that we value clean water at the cost of a water filter per household, and do not adjust for household income. Then we have already performed a weighted cost-benefit analysis. Thus, all we need to do is examine the labour supply effects of such practices. There is also scope for simulation studies. For example, having players take on an avatar and choose its labour/leisure trade-off under imaginary conditions, and see whether direct income matters more than the kinds of indirect benefits assigned through cost-benefit analysis.

It is interesting in this regard to review the literature on salience, and what has been called schemeduling (Liebman & Zeckhauser, 2004) which could suggest that differences in deadweight loss may arise from differences in the salience of tax instruments (Mourao, 2008; Baron & McCaffery, 2004; de Bartolome, 1995; Padovano & Galli, 2002; Saez, 1999; McCaffery & Baron, 2006; Sanandaji & Wallace, 2010). Jolls (1998; Jolls et al., 1998) in a discussion of tort law and redistribution makes a very similar point. The kinds of mental accounting that people do for taxes may be very different to the kinds of mental accounting they do for tort losses. Hence treating an egalitarian ethos in tort law as if it were equivalent to higher taxes is probably a mistake.

Additionally, there are numerous other ways in which differential weighting might not function like an income tax that do not relate to the psychology of salience- at least directly. Consider two regions in a country A & B. Rich people tend to live in A, and poor people tend to live in B *however, interregional mobility is low.* Maybe for cultural reasons people tend to keep living where they were born and perhaps the situation is like Australia, where interstate mobility is only half as common as in the US (Rahman 2005). Living in region A is correlated with being wealthy, but being wealthy does not cause one to live in region A. Suppose that we built a similarly good network of roads in region B as A due to the use of weighted cost-benefit analysis using regional welfare weights, as practiced in the UK (Evans et al., 2005). Is this going to reduce the degree to which people work hard to become rich? No, because becoming rich wouldn't, in all likelihood, change the region they lived in.

These arguments lead us to the conclusion that the incentive effect argument against weighted CBA, based on a claim that a dollar of redistribution via taxes is equivalent to a dollar of redistribution through weighted CBA, is thus dubious.

10.2.2 Kelsey on the argument from control theory

Kelsey (1988) argues in favour of weighted CBA on the grounds that where there are multiple policy instruments, each with a degree of uncertainty in its effects, it is best to utilize a variety of instruments. To this end, Kelsey refers to a number of results from control theory. Ng's proposal, and the proposal of others like Weisbach, that only income tax should be used is thus non-optimal.

It may be obvious enough why there will be uncertainty in the effects of non-tax and transfer policies, but why expect there will be substantial uncertainty around policies aimed at adjusting the income distribution through tax and transfer? Might we not expect any uncertainty around such policies to be a mere rounding error? Kelsey (1988) does not think so:

> Tax law made by governments has effects that were not anticipated when the legislation was designed. Individuals find loopholes, which the government never intended, enabling them to reduce their tax bills or even avoid tax completely. Further uncertainty is introduced by the fact that many individuals use illegal means to avoid paying taxes. The system of allowances is extremely complicated. Many individuals have very imperfect knowledge about the allowances they are entitled to claim. The uncertainty is such that some points have to be settled by test cases in the courts. Neither side knows the outcomes of these cases in advance. (Kelsey, 1988, pp. 580–581)

I would draw particular notice to the failure of citizens, for whatever reason, to claim the full range of benefits that are available to them. Consider, for example, underclaiming of valuable welfare benefits. Against Kelsey, Ng (1988) has a reply in which he argues "we are both correct", but that, in the main, extensive use of weighting is not justified unless certain special conditions are met.

It is important to note that even if Kelsey is right, the optimal weighting in cost-benefit analysis is unlikely to be the weighting that what we might term "naive" utilitarianism would give. Naive utilitarianism can be defined as trying to draw conclusions about what a utilitarian should do directly from a system of weighted cost-benefit analysis with weights set according to marginal utility of income. This approach, in which the CBA weighting function is the same as that used in the global social welfare function, will almost certainly not be vindicated by Kelsey's approach. Instead, if we have utilitarian goals and follow Kelsey's analysis it is likely that we will end up with weights somewhere between unweighted cost-benefit analysis, and naive utilitarian weights.

This gives another reason to be sceptical of the idea that weights in our model can ever fully capture total social costs and benefits. If Kelsey is right, to get at such a number we would have to engage in complex estimates using control theory. It is much easier to assess something like the direct psychological impact of a program than it is to capture all of the flow on effects mediated through factors like labour supply.

It is interesting to combine Kelsey's argument from control theory with arguments we make elsewhere about the political costs of redistribution. In particular, we can imagine an argument which considers political uncertainty in redistribution. If we are unsure which measures can be enacted, or more plausibly still, if we are unsure which of our measures is politically sustainable in the long run, it may be that we are well-advised to try to implement a variety of measures which each help the poor.

10.2.3 The equality intuition

But these positive concerns, based on control theory, relative salience and other factors, are not our only concerns. Although empirical research would be needed to confirm this, I would posit that many people have, and hold quite strongly, what we will term the *equality intuition*:

Ceteris Paribus, the government providing public goods or services for a poorer person is at least as important as the government providing goods or services for a richer person.

To take a relatively soft case, a detailed unweighted cost-benefit analysis using maximum willingness to pay would place most parks in richer areas, but I strongly suspect many people would find this intuitively unfair. In the most extreme case- the case of valuing lives- it has become quite standard to adopt the equality intuition and to place an equal value on all lives, regardless of the wealth of the life saved or lost. But to get this kind of *valuing equally* you will need some kind of weighted cost-benefit analysis. Our own approach- SWBEE, will certainly give you that data you need to treat the poor and the rich equally, if that is your goal. On the other hand, if I am right about the equality intuition, unweighted cost-benefit analysis does not respect popular expectations in decision-making.

I think that there is probably a spectrum of acceptability on unweighted cost-benefit analysis in the eyes of the public. At one end we have, for example, road upkeep- it is not unusual to encounter sentiments like “road upkeep by municipal government should be supplied in proportion to rates paid”. In the middle region of acceptability, I suspect would fall functions like healthcare, education and criminal justice. At the extreme region, which I would guess almost no ordinary person would accept, we would find valuing lives differently based on wealth. My guess is that a majority hold the equality intuition at least starting from the middle type cases, but without empirical data we cannot say.

One dramatic example of political qualms about unweighted versus weighted cost-benefit analysis has been the debate over whether the IPCC should value the lives of citizens of different countries with different income levels differently, as discussed by Frankhauser et al. (1997) for discussion. In this case there was strong ethical impetus to the view that lives should not be valued unevenly.

If one is of a more utilitarian bent and does not care especially about the general public’s intuitions on these cases, one might frame things differently, in terms of a *taste for fairness.* There is evidence both from our everyday experience and from psychology (Bjørnskov et al., 2013; Knight & Gunatilaka, 2022) that the negative effect of income inequality on happiness is partially mediated by perceptions of fairness. Even animals (Brosnan & de Waal, 2003) become upset in the presence of what they consider to be unfairness. This evidence gives a utilitarian reason to favour a distribution of public goods which is seen as fair, as perceived unfairness lowers welfare.

As an expansion of this point, we should consider *positionality*. We know that people’s wellbeing is heavily mediated by their sense of relative income and then social standing (Ball & Chernova, 2008; Blanchflower & Oswald, 2003, 2004; Clark & Oswald, 1996), and being treated as inferior by society is highly detrimental (Anderson et al., 2012; Layte & Whelan, 2014; Marmot, 2004; Marmot & Wilkinson, 2001; Wilkinson & Pickett, 2006). If we do not consider the poor as equally worthy of parks or roads, or even more so, hospitals or schools, what will the effects be on the self-esteem and felt dignity of the poor? Of course, one could say these factors are unlikely to be salient to the poor, but this seems incompatible with a line according to which the difference between weighted and unweighted cost-benefit analysis is likely to be fully salient to the rich!

Another question, more abstract and less direct, is what kind of government we are setting up, if we set that government up to value the needs of the poor less than the rich. What will be, so to speak, the effect on the political morality of the state and the political morality of the polis?

10.2.4 Political feasibility and political costs

Thus far we have framed our *political* arguments about the tax progressivity concern in terms of a political philosophy around legitimacy, but we could just as easily make it an argument about political feasibility. It may well be politically easier to build an equal number of parks, hospitals and schools for rich and poor people than it is to, say, make direct payments to poor people.

Earlier we mentioned Fennell & McAdams (2016) who, in a discussion of law and economics suggest that, for example, redistribution through tort law mechanisms is likely less politically costly than redistribution through taxes and transfer because of perceptions around fairness. A similar 'fairness factor' may be applicable to building additional bridges, hospitals etc. for the poor through weighted CBA. People may feel uncomfortable with monetary transfers to the poor, but may regard, say, equal road building in poor and rich areas as simple equal treatment. That is, not only may cash transfers have relatively high political costs, but equal public good provision may have exceptionally low, even negative political costs, as a method of redistribution.

This might sound like making excuses to embrace sub-optimal policy for the sake of political feasibility. The point, however, is that the public (and their representatives) have particular feelings around redistribution and the "right" way to do it, and that economists shouldn't let the perfect be the enemy of the good by supporting redistribution only if it is done in what they consider the "right" way. Only supporting redistribution if it is done efficiently could well amount to never or rarely supporting redistributive plans which have a chance of success.

I hasten, quickly to add that that this not, by any means, to suggest that we should stop or cut programs moving economic resources from the rich to the poor by way of tax and transfer. Such programs are necessary for countless reasons, and I believe the progressivity of tax and transfer should be increased. Rather merely to suggest that, given public sentiments, the maximum sustainable level of redistribution may be higher using a mixed basket of public goods and cash transfers.

Scitovsky's comments on the political costs of simple redistribution in a free-market economy are worth quoting here:

> *there is a presumption in the free enterprise economy against the State correcting the income distribution brought about by the market mechanism. This militates against any economic policy that would have to be accompanied by a payment of subsidies or compensations. In other words, the effects of an economic policy on efficiency on the one hand and on income distribution on the other hand cannot-as a rule-be separated in a free enterprise economy, because in such an economy compensation payments are not feasible politically. (Scitovsky, 1951, p. 310)*

10.2.5 A house and congress divided

As a side point, we note the importance, in the context of the USA, of disagreement between different arms of government. This point has been made in the context of debates over weighted CBA before (Adler, 2016). In the United States, most CBA is done on behalf of the executive, yet the progressivity of the tax and transfer system is set almost exclusively by the legislature-and, given various institutional and political complexities, does not even necessarily reflect the current views of the legislature. Where the executive is at odds with the legislature, and the executive more progressive, alternatives to unweighted cost-benefit analysis seem a natural choice for the executive. Even if a majority of both houses and the executive are in agreement, changing the tax code may not be a viable option for increasing redistribution, because of the filibuster.

Even in parliamentary systems, it might be impossible for the party of government to coordinate a change to tax policies, especially if parliament is bicameral, but potentially even with a unicameral chamber especially where complex coalition politics are involved.

An obvious counterargument available to partisans of the tax progressivity concern is that if the presidency cannot get a majority together to change the tax code, then the president does not have the mandate to redistribute through CBA. Does the executive have a mandate to insist on greater equality if it cannot get the numbers on the tax system?

The answer to this question will depend on many factors- for example, whether you think legitimacy is primarily *systematic*- dependent on the way the rules are worked out; *democratic*-dependent on the degree to which the popular will is implemented, even if it bends or breaks rules; or *outcome oriented*- dependent on the degree to which the process gets the right results. To answer this question in, for example, an American context we would also have to consider the degree to which the United Sates is, at present, democratic. Most obviously this raises questions about the malapportionment of the American senate, the filibuster, the practice of gerrymandering and the role of the electoral college, but it also raises questions about electoral democracy in general, elite capture of democracy and many other topics. We do not have the space to resolve these questions here, but we are sceptical of any argument that holds democracy or legitimacy forbids the executive from redistributing by way of the provision of public goods.

10.3 Concluding thoughts on the incentive effects argument

10.3.1 Could weighting counter non-transparent biasing towards the rich?

In the context of discussions over weighting, it is important to consider the possibility of implicit interest 'weights' in the process. It is a truism that the interests of the rich have great influence in politics (Bartels, 2004; Gilens, 2005; Karabarbounis, 2011) and the interests of the rich most likely influence what projects are even considered. Thus, in practice, CBA is embedded in a political process that may be weighted even more heavily towards the rich than would occur as result of a democratic application of unweighted CBA. Does this source of implicit bias vindicate weighted cost-benefit analysis as a way of counteracting bias? No, not necessarily though it *may* give us some reason to think weighted cost-benefit analysis is superior to the status quo. But it does add a certain degree of political realism to claims that unweighted cost-benefit analysis reflects nothing more than correction towards efficiency.

10.3.2 A final jab at the incentive effects argument

We can give an additional argument against the incentive effects argument by way of the following example. Consider that a wealthy person decides to give away their money, and they seek out an economist to advise them in this process.

The wealthy person had been planning to give their money to the poorest in society, and this seems like a fairly natural and defensible choice. However, the economist, proceeding on the basis of the incentive effects argument, claims this will have unintended consequences, as it will raise the effective tax rate on the rich, should the rich expect similar beneficiaries in the future to act in the same way. The economist further argues that these incentive effects are likely of the same magnitude as would occur as result of similarly redistributive tax change. Ergo, in making his donations, the wealthy man should treat a dollar as a dollar, no matter who it goes to distributively.

Intuitively there are many things we want to say here, but two of them are:

1. Just like a certain arm of government may not have the means to increase redistribution by raising taxes, so our philanthropist may also lack these capabilities. Hence, just like the government in that case, our philanthropist may not shy away from the prospect of effectively raising tax rates through their philanthropy, and the question of whether or not it would be better to increase redistribution through taxes is largely irrelevant.
2. It is unclear that existing charitable ventures effectively raise the tax rate in a way that influences behaviour, perhaps in part due to salience effects.

10.3.3 In a complex situation with various tradeoffs, simply presenting useful information becomes an even more attractive option

We hope we have made the case that there is no simple answer to the question: Should redistribution be carried out purely through the tax and transfer system?

For this reason, it makes a great deal of sense to have information about the direct SWB effects of a policy available separately from calculations of incentive effects. To establish this point, it is not necessary that SWBEE alone will provide sufficient information to maximise some SWF, for example in the utilitarian case, maximise the sum of SWB.

It is rather only necessary that SWBEE provides useful data for decision makers as an output of the analysis. Indeed, this is essential to our argument in chapters 4 and 7- we deliberately eschew the role of decisionmaker for the role of information provider. We leave the question of how to integrate this information into a balanced decision-making process which accounts for both the direct welfare effects of the policy, and the incentive effects to others to resolve. In my view, this challenge has not been met by anybody in the literature, including advocates of unweighted cost-benefit analysis, because we still do not understand factors like political constraints and salience effects and how these interact with redistribution by tax and transfer, and public good provision. Treating redistribution by regulation and public goods as broadly equivalent to redistribution by tax and transfer in its political costs and effects on the labour supply etc. may be a tempting simplification. Ultimately though, this assumption is no more responsible than treating redistribution by policy and public good as ‘free’ additional redistribution. This is, as best I can tell, an unsolved problem for everybody. If anything, SWBEE gives us an additional resource here- humility- it does not claim to provide an overall evaluation of the correctness of policy, only to provide useful information.

We can make an additional defence of SWBEE along lines similar to an approach outlined by Bronsteen et al. (2013). If unweighted CBA lets us know something valuable about the most efficient way to solve a certain type of market imperfection, something SWB based methods don’t tell us, it is possible to supplement our analysis by performing CBA as well as what we call SWBEE- although see chapters 7 & 8 for concerns about how much unweighted CBA really tells us.

There are broader problems here, problems that also apply to weighted and unweighted cost-benefit analysis and go beyond the questions of policy-progressivity. It is easy to identify direct effects, but every project has an endless array of second, third and nth order effects. Michael Beggs (2023 personal communication) gives the example of the Australian first home buyers grant. Any ‘true’ appraisal of this policy by welfare economics- whether weighted CBA, unweighted CBA or our method would have to grapple with a tangle of effects, distributive, macroeconomic, cultural, political and so on- giving them all their due in analysis would fill volumes. In practice, every analysis in welfare economics has to draw an ‘outside’ and an ‘inside’ scope, and it is an open question whether this can be done in a fully principled way.

10.4 The urgent need for experimental philosophy and the case of leaky buckets

Experimental philosophy (Sytsma & Livengood, 2015) is a paradigm within philosophy that attempts to uncover details about the beliefs of ordinary people on philosophical matters. This is important because philosophers have often appealed to philosophical intuitions -seen as widespread, even universal- to support their positions.

Above I posited that people have what I call an equality intuition- that the poor should receive *no less* in public goods than the rich, at least with regards to certain goods like health and education. This is an empirical claim- it is the sort of thing that we can, in principle, check- and until such time as it is checked, it is only a guess on my part- a reasonable guess, but a guess, nonetheless. Furthermore, it is not the sort of empirical claim that is easy to test, or even to frame with any precision. Work in experimental philosophy has shown that intuitions can vary between demographic groups, between questions that differ only in aspects philosophers would traditionally consider irrelevant (Feltz, 2007), between individuals (Feltz & Cokely, 2008), and on the basis of various experimental interventions like precise wording (Petrinovich & O’Neill, 1996), though Knobe (2019) argues that the importance of demographic effects may sometimes be overstated.

Experimental philosophy has been imagined in a variety of different ways- for example, as a positive project for uncovering new philosophical truths or as a critical lens from which to attack traditional philosophy (Knobe & Nichols, 2017). Empirical normative political philosophy of the sort we are proposing here- what we might call *democratic experimental philosophy*- potentially has a different purpose altogether- *democratic consultation*. We are not so much interested in philosophical truth here as democratic support and legitimacy. Even though welfare economics is intended to aid economic decision making by democratic states, there has been startlingly little research into the public acceptability of the methods and reasoning of welfare economics. The rhetoric of welfare economics often shows an anti-democratic bent, e.g., as noted by Nyborg and Spangen (1997) when they talks about “the” social-welfare function of “the” social-planner.

A notable exception to the claim that there has been little experimental philosophy in relation to welfare economics is a series of *leaky bucket* (Beckman et al., 2004; Pirttilä & Uusitalo, 2010) experiments that, examine questions of the form: “Would you take A dollars from a person with X dollars to give B dollars to a person with Y dollars” in order to determine an implicit marginal elasticity of the social value of income.

Generally speaking, prior results in experimental philosophy should lead us to be very careful about simply giving participants something like leaky bucket experiments and assuming that participants will understand them in the same way that we would. As discussed, a variety of work in experimental philosophy has suggested that intuitions are unstable to context, demographics, exact words and framing (Knobe & Nichols, 2017).

Perhaps unsurprisingly, there is evidence that responses to such questions vary greatly depending on the details of the asking. Pirttilä and Uusitalo (2010) in running both a leaky bucket experiment and a preferred wage distribution experiment find enormous differences between the preferred level of inequality in each:

> For the 'leaky bucket' type of question we estimated that the median inequality aversion parameter [ε in this work] lies below 0.5 (e in Atkinson's social welfare function) ... The results from the wage inequality type of question gave a completely contradictory view of the magnitude of inequality aversion, with e [ε] being greater than 3. (Pirttilä & Uusitalo, 2010, p. 74)

I wish to venture one theory as to why preferred redistributivity in leaky bucket experiments may be so low in comparison to other experiments, for example preferred wage distribution experiments. While the framing of the question is intended to make us think about only the direct benefits to the richer person versus the direct benefits to the poorer person, the participants who favoured the richer person may have been extrapolating out further economic effects on society from the loss of money in the leaky bucket- e.g. they may have been imagining a "trickle-down effect" would be lost if the money were redistributed- perhaps the wealthy will no longer pay for a plumber for bathroom renovations, and so the plumber will be poorer etc.

Consider this question, again by Pirttilä and Uusitalo (2010):

> The taxation of all high-income earners, whose disposable income exceeds €3300 per month, is increased. The money is spent for the benefit of those low-income earners whose disposable income is less than €800 per month . . . The high-income earners can, however, react to the tax increase by reducing their work effort, and part of the money goes to administrative expenses. Therefore, for each €100 paid by the high-income earners, only €25 can be spent for the benefit of low-income earners. (Pirttilä & Uusitalo, 2010, p. 63)

Such a question leaves it open to the respondent to worry about flow on economic effects, such as the 'trickle down' effects we mentioned, political and political economy effects, etc. in a way that a question which asks directly about income distribution doesn't, since it represents, at least in the interim, a final result:

> Let us imagine that in wage negotiations two different alternatives are considered. Which of the following do you prefer?” [two different wage distributions are then outlined].(Pirttilä & Uusitalo, 2010, p. 64)

There is also the possibility that participants were under the influence of loss aversion, a claimed though debated phenomenon (Gal & Rucker, 2018; Yechiam, 2019) in which people weight loss more heavily than gain. An important question here to give us some perspective on the experiment is: how would participants react if you turned it around and suggested *gaining bucket* experiments that transferred from the poor to the rich? On these experiments, one takes A dollars’ worth of income or consumption from a poorer person and gives it to a richer person, who receives B dollars’ worth of consumption or income [with $A > B$]. I would guess that such an experiment would show a cataclysmically large value of ε.

If I am right about this, the difference might be partly attributable to loss aversion, but it may also reflect, a sense that *pre-tax income intrinsically belongs to its “owner”, and the government must defeat a strong moral presupposition if it is to justify redistributing it* (Murphy & Nagel, 2004).

The problem is that between loss aversion and the public’s different philosophical starting points from most economists, the leaky bucket experiment may not mean the same thing to participants as it does to welfare economists. For example, the public may well reject ‘efficient’ policies that greatly enrich the already rich at the cost of the poor because much of the public are deeply sensitive to who starts with ownership. Simply reading off an ε value of 0.5 from the leaky bucket experiments will not reflect the public’s true values over a broader range of experiments. Income distribution experiments likely give the public’s preferred outcomes, but the question of what measures they perceive as legitimate in getting to that outcome is a separate question.

A better paradigm for experimental philosophy on welfare economics might involve *active participation and learning by the subjects* modelled on participatory democracy. A jury of participants are presented with the challenge of establishing a hypothetical procedure for trading off gains and losses between persons. This jury can bring in expert witnesses to give more information. The jury explains not just their preferred procedure, but also the rationale behind it. The advantage of active engagement is that we are not simply throwing questions into the wind, trusting that participants are thinking in the same basic terms that we are. Nor are we assuming that participants start with complete views on these subjects and our role is to draw them out. Nor are we seeking to appropriate the views of respondents as props in our preexisting debate. Rather, we are inviting participants to coherently develop a position on the basis of their underlying values, joining the conversation.

Democratic consultation and experimental philosophy are an urgent priority in this area. Perhaps no area of economics exceeds the conjoint degree of *influence over public policy*, *disputable value judgements* and *obscurity from the public gaze* as applied welfare economics. There is thus an urgent need to study exactly how the public would feel about the ethical foundations of mainstream welfare economics, were the public to understand them fully- and to determine whether they might prefer alternative approaches. Further we must understand how the public would differ among itself on these questions, the lines on which these differences might fall, and how these differences might be incorporated into applied welfare economics

10.5 Should decision makers expect welfare economics to provide an algorithm?

In arguing that the tax progressivity concern undermines weighted cost-benefit analysis, Weisbach (2015) for example raises a broader question with arguments about whether 'maximise welfare' or maximise efficiency' is a better instruction. Should agencies be making decisions according to clear, nearly algorithmic criteria? If agencies were to make decisions on the basis of a singular rule or algorithm, I agree that Weisbach's attack on standard welfare-weightings would be disturbing- for what kind of rule would "maximize welfare on your patch of turf with no regard to overall labour supply incentive effects" be for the Food and Drug Administration or the Securities and Exchange commission?

The approach we have described requires a divide between *welfare economics* and *agency decision making* because it holds that the proper role of applied welfare economics is to give information to agencies, who will then integrate it with other considerations in order to come to an overall decision. But what if you think that agencies should make decisions on the basis of clear, algorithmic rules- albeit perhaps subject to discretion in exceptional circumstances? In this case, our insistence on a divide between applied welfare economics and final decision-making becomes rather a moot point, especially if the information the welfare economist provides forms all or almost all the inputs into the decision algorithm.

We can imagine an alternative pole to this framework, which holds that a better approach is for agencies to make decisions on the basis of an enormous variety of considerations balanced against each other in a way that is likely impossible to formalize. These might include, but not be limited to 1. Prevailing moral, legal, professional and political standards 2. Estimates of the welfare effects of policies 3. Statements by governments about their strategic goals 4. Internal charters outlining values. In extremis it may even be appropriate to appeal to more exotic sources, such as international legal instruments and personal moral judgements.

We are interested here in a debate between *non-algorithmic* or *non-rules-based* approaches to agency decision making and algorithmic approaches. According to the algorithmic view of agency decision making, agencies should make decisions according to prespecified, quantified decision rules where it is possible to do so. According to the non-algorithmic view, agencies should review all available information about the costs, benefits and other consequences of an action, weigh them and consider them in the context of other possible actions, and then make a qualitative judgment on which course of action is desirable.

The argument for a singular decision rule is easy to make. In the absence of a singular decision rule, or more generally fully defined goals of policy, decisions by agencies might simply reflect the whim of the decision maker or analyst. This is analogous to the debate between cost-benefit analysis, which can be taken as a canonical form of a single criteria, algorithmic decision-making approaches, and multicriteria decision analysis. Dobes and Bennett (2009) argue, multiple criteria and other apparently more sophisticated approaches really amount to dodging the question of how to weigh things against each other:

> *Most schoolchildren have learned that it is illogical to try to add apples and oranges. And few biologists would propose adding the mass of a flea to the length of a rat in advancing a new theory, although there is no problem in adding the weight of a flea to the weight of a rat…Multi-criteria analysis breaches this principle of dimensionality. It is not possible, as in the tabular example above, to add hectares of vegetation saved to the number of species recovered.* (Dobes & Bennett, 2009, p. 23)

But there is also a strong argument for non-algorithmic decision making. If processes like cost-benefit analysis that rely on singular decision criteria are underspecified or require supplementary judgment, bureaucrats may effectively fill them in with their discretion, so it might be more honest to admit this is part of the process "upfront", rather than pretending everything is being done by the algorithm and there is no need to look "behind the curtain". Cost-benefit analysis as a form of algorithmic/single criteria decision-making involves an enormous number of discretionary empirical, normative and political decision points in all but trivial cases, so there is much indeed going on behind the curtain.

Theoretical arguments aside, what of the practice "on the ground"? In an intriguing, albeit small scale, study Annema, Mouter and Razaei (2015) consider the Netherlands experience of cost-benefit analysis. Nominally CBA is required for large scale transport infrastructure projects. In practice though, while CBA is seen as *useful* it is hardly considered individually decisive. In particular, politicians find the idea of a single, composite number, representing a total judgment of the value of the project "pretentious" and are more interested "in a clear picture of the trade-offs" (Annema et al., 2015, p. 791).

Should agencies operate according to a formal rule, or according to a multidimensional weighing of contrary considerations? Answering this question decisively is beyond the scope of this thesis, but I do wish to make one point on this debate that I will frame as a response to Bennett & Dobes (2009), who argue that multicriteria analysis is illogical, among other things, because it is impossible to weigh up disparate units against each other. Now it may be true that certain types of multicriteria analysis effectively require the decision maker to weigh apples against oranges mathematically, but there is a more sophisticated version of this approach which avoids this objection, and effectively makes Bennett & Dobes (2009) argument circular.

According to this version of the multicriteria approach *the whole point is to avoid using a decision-making algorithm and rely on a qualitative approach to judgment because this is the more natural mode of operation for many types of decisions given the architecture of the human mind*. Given the limitations of all analyses- whether singular or multicriterial in this fallen world, it may well be that qualitative decision making, though flawed in theory, leads to better outcomes in practice than trying to squeeze everything into a common numeraire. The objection of Bennett and Dobes that the multicriterial approach is unquantifiable then may be circular in the sense that *it was always intended to be unquantifiable.* Saying something makes no mathematical sense is no concern if it wasn't intended to be a mathematical process but a qualitative act of ineliminable judgment.

Let us use an analogy from formal decision theory. Consider two people. Johnny makes his decisions through thinking things through thoroughly but unsystematically. Maya makes her decisions by assigning utilities and probabilities to different outcomes, conditional on her actions, and then working out which action has the highest expected utility. While there are various proofs in formal decision theory of the *rationality* of certain decision frameworks like Maya's in the abstract, this is no guarantee that actual humans deploying them explicitly, on average, generate better decisions. It could be that heuristics work better for humans in practice than formal explicit reasoning in many cases- this is an empirical question that only empirical evidence can resolve, and is likely relative to the decision maker, the decision, and other contextual factors.

In passing, we should mention an astonishing- and alarming study by Sinden (2019) who found: "categories of benefits described by the agency as "important," "significant," "substantial" [were] left unquantified in 80% of CBAs" (Sinden, 2019, p. 73). In other words, 80% of studies excluded factors that were, by their own estimation, important, because they could not come up with a method to quantify them. In theory, of course, the agencies should simply quantify these values, however imperfectly, rather than excluding them. In practice, this ideal seems to be- for whatever psychological or political reasons- unattainable. Perhaps non-algorithmic decision making would prove better able to accommodate such costs. An illustration, then, that in some (perhaps most!) cases, cost-benefit analysis may be avoiding comparing apples to oranges by pretending apples don't exist!

Finally, we raise a democratic concern: what role for the democratic public's view is left in the algorithmic picture of decision-making? We've tied our flag to the view in this thesis that the products of welfare economics should inform rather than dictate decision making, because we think doings so leads to a satisfying answer to the question "How can economics play a fruitful role in decision making without making strong assumptions about what values society should adopt". We agree with Nyborg & Spangen (1997), Singh & Alexandrova (2020) and doubtless others that there are insurmountable contradictions between the ideal of *democratic decision making* and decision-making via welfare economics algorithm.

10.5.1 Colander, the art of policy, and why it is incompatible with decision making by an algorithmically applied rule

Colander (2015) builds an approach to the application of policy well worth discussing and inherently opposed to any algorithmic decision making approach, which draws inspiration from economic authors in the classical liberal tradition such as the Neville Keynes and classical utilitarians like Mill. Colander (2015) distinguishes between *precepts* and *theorems*. A theorem is a result telling us what will happen in a situation if certain formal, often very demanding premises hold. A precept is a general maxim about how policy should be run. To get from theorems to precepts we need both moral judgements and healthy commonsense about what might happen when the demanding conditions stipulated by our theorems are not met. Unfortunately, argues Colander, modern economics often collapses the distinction. For example, some infer from the theorem that a minimum wage causes inefficiency under perfect competition to the claim that the minimum wage reduces the welfare of society- yet this requires two giant leaps, one positive and one both positive and normative:

1. That we can safely assume that the real-world labour market, which is not a paradise of perfect competition will act like the textbook case
2. That inefficiency in the formal economic sense of the theorem necessarily implies a reduction in welfare, or even in *maximum feasible* achievable welfare.

Colander (2018) then goes on to argue for what he calls an engineering approach to applied economics. He makes his argument by identifying two types of difficulty with the social welfare function approach (which presumably any decision rule approach to economics would have to apply. Firstly, values are not set independently of positive discussion- exactly what our values around a topic are *cannot be set a priori of thinking and discussing the topic*. Secondly, even if our values could be set in such a manner, they would be much too complex to be captured by a social welfare function. Instead applied economics should proceed in a method reminiscent of engineering, using heuristics but recognising that every heuristic is fallible.

If Colander is right, the search for a generalized decision rule must be misguided. However, even if he is not right, he well illustrates, I think, why there is nothing inherent to the very idea of economic evaluation which *requires* a formal decision rule.

10.6 The environment and other matters of importance not easily or fully captured by subjective-well being

Our method, like traditional cost-benefit analysis, can only include effects on the environment insofar as they are mediated through effects on human beings. From some points of view (e.g., deep ecology (Naess, 1984) this will be a negative for our approach, although one it shares with traditional CBA. There is a danger our approach might be even *worse* than traditional CBA in some ways- remote wilderness might have very little impact on any measure of wellbeing, yet people might still be willing to pay to sustain it in contingent valuation studies.

This topic deserves a much more thorough investigation. However, our preferred solution is relatively simple. We support providing information to decision makers and the democratic public. Such information can, in a disaggregated form, include information not just on subjective-wellbeing effects, but also on the environment. This includes other factors that are thought to be either of intrinsic value separate to their effects on wellbeing, or which are thought to be too difficult to integrate with the mainline analysis of wellbeing for calculational reasons e.g., effects on national defence (Harberger, 1971). We could also value environmental impacts through classical CBA, if desired, although for reasons raised in chapters 7 and 8, I have reservations over whether willingness to pay for conservation truly reflects the value the public places on the environment.

10.7 Recommendations for future research

We have merely outlined a methodology in this thesis in abstract terms. Thus, the most urgent project arising from this thesis would be the creation of best practices for implementing the methodology we have described.

At the same time, we have focused heavily on cost benefit analysis, but the approach must be extended to areas beyond CBA including the analysis of optimal tax, trade theory, the economic analysis of law, constitutional theory and other aspects of applied welfare economics.

Beyond this, the two most urgent tasks with regards to expanding our approach are: 1. Finding solutions to what we described in section 4.8 as the problem of open requests (briefly discussed again in 10.7.1) and, a task that we would commend as urgent to the whole of welfare economics- 2. The investigation of the political, political economy, salience and other pertinent aspects of redistribution via public goods and regulation versus redistribution by tax and transfer. It simply will not do to assume that building an additional amenity for the less well off is treated by the public in its political, labour and investment choices as identical to tax and transfer redistribution for the less well off.

In chapter nine, we made a case for the desirability of a psychometric measurement of desire satisfaction per se, not merely life satisfaction or domain satisfaction. This seems like a vital-yet surprisingly neglected task. Once we have a measure of perceived overall desire satisfaction it would be useful to research the degree to which life satisfaction and domain satisfaction approximate desire satisfaction.

We have largely taken psychometrics for granted in the context of this thesis, but of course psychometrics raises its own questions. Further research- already thoroughly ongoing- into the philosophy of psychometrics continues to be welcome. Outstanding issues include the nature of validation, measurement theory, the philosophy of statistics, the methodological role of techniques like factor analysis, the usefulness of face validity, the role of folk knowledge in science and so on.

Chapter 6 discussed the problem of cardinality and provides possible solutions but gives no definite answers. The literature on establishing cardinality with psychological measures is surprisingly scarce for such an important question. Further empirical, methodological and philosophical exploration is needed, and I hope to study this in future work. Equally, establishing a zero point for ratio strength subjective-wellbeing scales is a vital task.

We flagged earlier in this chapter a moral case for further investigations using the methodology of experimental philosophy into what the public does and does not want from welfare economics, as employed by government agencies. The democratic public has a right to have its preferences regarding welfare economics taken seriously, if such preferences are to be employed as a core basis for agency decision making. Indeed, as Nyborg and Spangen (1997) note, the wrong sort of approach to making such decisions can amount to a curtailing of democracy- is it not essential then, in trying to get it right, that we consider not only the public's preferences, sentiments and welfare in decision making, but also the public's views on how those first order preferences are used?

In this thesis we have often treated behavioural methods of getting at intensity of feeling and psychometric methods for the same as wholly distinct, even opposed. In truth, it is entirely possible to integrate behavioural methods into a psychometric paradigm. On this approach, the analysis of economic behaviour is not seen as giving revealed preferences- defined through the representational theory of measurement, but as giving data hinting at latent variables (the strength of wanting) in the tradition of the psychometric approach to measurement. We commend further research on this approach, and as we have already noted, it may be possible to combine these investigations with neuroscientific evidence e.g., of the activity of the dopaminergic reward system.

We have already noted ongoing research into what the public considers to constitute the good life, and we commend future research on this subject.

10.7.1 Broad welfarism as a promising avenue for research

There is a limitation with our approach that we raised in chapter four. Our approach will work from the point of view of *evaluating* (or contributing) to the evaluation of policies, but it does not have any inbuilt method of, when given a goal, for example building a bridge somewhere to cross some river, generating methods that will achieve that goal in preferred ways. This is because our method encodes no SWF. By contrast, other approaches, for example, the SWF approach, have a natural methodology for working out the best ways to achieve a goal - when given a goal such as building a river crossing, they will suggest the project that is estimated to maximising the social welfare function.

Earlier in the thesis we suggested that it may be that almost everyone may have a welfarist *aspect* to their ethics. They may have other things they value apart from welfare, and they may place side constraints other than trying to maximise their values, nevertheless, most people care about welfare as a matter of ethics, even if they care about other things as well. We might conceive a view, alternative to the one I develop in this thesis:

> *Broad welfarism:* the aim of welfare economics is to tell us what the welfarist component of our ethics would have us do in any given case.

But don't different welfarisms advise very different conduct? Suppose that we could demonstrate a spectrum of welfarist approaches, ranging from strong prioritarianism to utilitarianism gave, in almost all circumstances the same results. This would amount to saying that the welfarist component of very many people's ethical views pointed in the same direction.

Research by Latty (2015c, 2015a, 2015b), and ongoing research conveyed to me by personal communication, suggested that a very broad range of welfarist views converge to similar recommendations in respect to the warranted degree of income inequality aversion, at least across some broad range of plausible circumstances and parameter values. Perhaps analysis of policy using this generalised welfarism might form exactly the sweet spot we alluded to needing above, allowing a large majority of people to reach, in principle and given all the facts, an important consensus about what a component of their ethical beliefs implies.

If a very large majority of people's views converged *at least when it came to the welfarist component of their ethics* this would be a useful form of overlapping ethical consensus and might give welfare economics a "north star" to orient itself towards when generating solutions.

10.8 Summary

In chapter one, we suggested the primary problem of welfare economics is to aggregate harms and benefits to different individuals, a problem we called the aggregation problem. Many of the classical problems of welfare economics are a part of this larger problem- the problem of interpersonal comparison, the problem of cardinality and the problem of the role of ethical judgements in science.

In chapter two, we taxonomized some prominent approaches to welfare economics. Our method of taxonomy was neither exhaustive nor exclusive but provides a framework for conceptualising the field. Finally, we introduced some of the machinery of Social-Welfare Functions.

In chapter three we reviewed the basic concepts of psychometrics related to our thesis. We then considered a sample of philosophical, empirical, and methodological concerns and thinking on the topic of the psychometrics of wellbeing. We outlined our position that psychometrics as a practice can provide knowledge about the various forms of subjective wellbeing and, most importantly, all the major theories philosophers have put forward about the good life. This came with inevitable caveats, most notably we do not claim to have refuted the field's critics, or all concerns about the field.

In chapter four we outline Subjective-Wellbeing Effect Estimation (SWBEE), an approach to welfare economics which aims to estimate the effects of a proposed policy on the distribution of subjective-wellbeing. Such an approach aims to provide ethically relevant *information* for the public and decisionmakers without engaging in ethical decision-making or privileging one or another set of values or even providing a single scalar 'score' for the proposal. Multiple estimates, made across different forms of subjective-wellbeing, are desirable where resources allow, as not everyone agrees on which forms of subjective-wellbeing provide the most direct evidence about the degree to which people are living good lives. Under some circumstances SWBEE can be evaluated directly through knowledge of the welfare effects of similar past interventions. In other circumstances where less is known it might resemble weighted cost-benefit analysis, weighted to try and anticipate the subjective-wellbeing effects of policies, rather than in line with some social-welfare function and its ethical evaluation.

In chapter five we tried to show how the psychometric approach can bear fruit with regards to a classic problem of welfare economics- the problem of interpersonal comparison. We argued that if we accept the accuracy of psychometrics, plus an assumption that it is reasonable to believe that alike functional states correspond to similar experiential states, the problem of interpersonal comparison is solved. Then, appealing to a combination of reasoning by Harsyani and Lerner, we have gone on to show how our assumptions can be further weakened.

In chapter six we argued that the psychometric approach to welfare economics can also deal with the problem of cardinality. Our strategy was to outline the many prongs of potential assault on the problem, in the hope of convincing the reader that the prospects are ample. Having classified objections to the idea that these instruments are cardinal, we have argued that there are multiple reasons to think such measures may already be cardinal, that even if they are not cardinal, there are likely ways to cardinalize them, and that the evidence suggests that, perhaps surprisingly, any difference the scales have from cardinality may matter relatively little for statistical purposes.

In chapter seven we outlined how SWBEE avoids the problems of value sectarianism in welfare economics- by conceptualising itself as providing *information* relevant to the public and decisionmakers, rather than adopting a specific social welfare function. Finally, we ended with an expression of hope- that presenting information regarding the SWB effects of policy may tend to drive consensus, rather than disagreement.

In chapter eight we argued that cost-benefit analysis (excluding SWBEE), weighted or unweighted, cannot, at least not as it is normally conceptualised, escape the problem of value sectarianism in economics. Attempts to conceptualise CBA as just information provision fails for several reasons, most notably *it is unclear what the value of the information provided is*. Attempts to defend CBA through value containment strategies- the conceptualisation of CBA as only involving modest value judgements that almost everyone would agree with- likewise fail.

In chapter nine we argued that even on the terrain of the desire satisfaction theory of wellbeing, and even though there is not currently a psychometric instrument for measuring desire satisfaction, it is likely that a psychometric measure would be superior in measuring desire satisfaction to revealed and stated preferences measures. Revealed and stated preferences are contextually unstable, and it is unclear what seeking the ‘true’ value across multiple different valuations amounts to. Psychometric methods allow the subject to define their own sense of overall wellbeing and desire satisfaction. Revealed and stated preferences measures tend to provide information about *instrumental* preferences, whereas a psychometric measure can very easily track *non-instrumental* preference fulfilment.

In chapter ten we considered a variety of topics relevant to our approach including wireheading, incentive effects, the possibility of democratic experimental philosophy, the question of whether policy choice should be algorithmic and the environment. By far the largest portion of our time was spent on the question of incentive effects, where we concluded, among other things, that while the measurement of incentive effects is difficult, it is everyone's problem- simply assuming that the unwanted incentive and political effects of redistribution through public good provision (or regulation) are equal to those through redistribution by tax and transfer is insufficient. We considered directions for future research and, in the next section, give an overall conclusion for the thesis.

10.9 Concluding statement for thesis

I have outlined a research program for an alternative approach to applied welfare economics based on psychometrics, SWBEE. This program aims to do more:

1 Give us detailed information on the distribution of mental states that indicate the likely presence or absence of welfare.
2 Give us a way to resolve the interpersonal comparison problem.
3 Give us a method to cardinalize various welfare metrics.

With less

4 Without implicitly, or explicitly, invoking a social welfare function, or dubious assumptions about democratic endorsement of the existing income distribution, the possibility of lump sum transfers etc.

Whether or not the details of this program are correct, it is my strong belief that the future of welfare economics- and perhaps of economics generally- is a collaboration with the psychology of wellbeing. After all, welfare economics is a fascinating study for its own sake, but its highest object must be providing the preconditions for human wellbeing, and it is impossible to know whether we are working towards human wellbeing without knowing a little bit about it.

Democracy is an element frequently ignored in welfare economics and in the detail of welfare economics, we can often lose the larger picture. The method I have described is intended as a workable approach for increasing both democracy and equity here and now, but it is also, in a small way, a brick in the wall of a larger conceptual structure- a small part of a blueprint for those who want to build economic and political democracy.

11 Bibliography

Adler, M. D. (2012). *Well-being and fair distribution: Beyond cost-benefit analysis*. Oxford University Press.

Adler, M. D. (2016). Benefit–Cost Analysis and Distributional Weights: An Overview. *Review of Environmental Economics and Policy*, *10*(2), 264–285. https://doi.org/10.1093/reep/rew005

Adler, M. D., & Posner, E. A. (1999). Rethinking Cost-Benefit Analysis. *The Yale Law Journal*, *109*(2), 165–247. https://doi.org/10.2307/797489

Adler, M. D., & Posner, E. A. (2006). *New foundations of cost-benefit analysis*. Harvard University Press.

Aigner, D. J., & Heins, A. J. (1967). A Social Welfare View of the Measurement of Income Equality. *Review of Income and Wealth*, *13*(1), 12–25.

Alexandrova, A. (2012). Well-Being as an Object of Science. *Philosophy of Science*, *79*(5), 678–689. https://doi.org/10.1086/667870

Alexandrova, A. (2016). Can the Science of Well-Being Be Objective? *The British Journal for the Philosophy of Science*.

Alexandrova, A. (2017). *A Philosophy for the Science of Well-Being*. Oxford University Press.

Alexandrova, A., & Haybron, D. M. (2011). High Fidelity Economics. In J. B. Davis & D. W. Hands (Eds.), *The Elgar companion to Recent Economic Methodology* (pp. 94–121). Edward Elgar.

Alexandrova, A., & Haybron, D. M. (2016). Is Construct Validation Valid? *Philosophy of Science*, *83*(5), 1098–1109. https://doi.org/10.1086/687941

Amiel, Y., Creedy, J., & Hurn, S. (1999). Measuring Attitudes Towards Inequality. *The Scandinavian Journal of Economics*, *101*(1), 83–96.

Anand, S., & Sen, A. (2000). The Income Component of the Human Development Index. *Journal of Human Development*, *1*(1), 83–106.

Anderson, C., Kraus, M. W., Galinsky, A. D., & Keltner, D. (2012). The Local-Ladder Effect: Social Status and Subjective Well-Being. *Psychological Science*, *23*(7), 764–771. https://doi.org/10.1177/0956797611434537

Andreas, J. T. (2010). *The cardinalist manifesto: The epistemology of the measurability of utility* [Ph.D.]. University of Illinois at Chicago.

Andreoni, J., & Sprenger, C. (2012). Risk Preferences Are Not Time Preferences. *American Economic Review*, *102*(7), 3357–3376. https://doi.org/10.1257/aer.102.7.3357

Angelini, V., Cavapozzi, D., Corazzini, L., & Paccagnella, O. (2014). Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases. *Oxford Bulletin of Economics and Statistics*, *76*(5), 643–666. https://doi.org/10.1111/obes.12039

Angner, E. (2013). Is it possible to measure happiness? *European Journal for Philosophy of Science*, *3*(2), 221–240. https://doi.org/10.1007/s13194-013-0065-2

Annema, J. A., Mouter, N., & Razaei, J. (2015). Cost-benefit Analysis (CBA), or Multi-criteria Decision-making (MCDM) or Both: Politicians' Perspective in Transport Policy Appraisal. *Transportation Research Procedia*, *10*, 788–797. https://doi.org/10.1016/j.trpro.2015.09.032

Antiel, R. M., Curlin, F. A., Persad, G., White, D. B., Zhang, C., Glickman, A., Emanuel, E. J., & Lantos, J. D. (2020). Should Pediatric Patients Be Prioritized When Rationing Life-Saving Treatments During COVID-19 Pandemic. *Pediatrics*, *146*(3), e2020012542. https://doi.org/10.1542/peds.2020-012542

Archibald, G. C. (1959). Welfare Economics, Ethics, and Essentialism. *Economica*, *26*(104), 316–327.

Armstrong, W. E. (1951). Utility and the Theory of Welfare. *Oxford Economic Papers*, *3*(3), 259–271.

Arrow, K. J. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, *58*(4), 328–346.

Arrow, K. J. (1951a). Little's Critique of Welfare Economics. *The American Economic Review*, *41*(5), 923–934.

Arrow, K. J. (1951b). *Social choice and individual values*. Wiley.

Asian Development Bank. (2013). *Cost-Benefit Analysis for Development: A Practical Guide*. Asian Development Bank. https://www.adb.org/documents/cost-benefit-analysis-development-practical-guide

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263.

Atkinson, A. B. (2001). The Strange Disappearance of Welfare Economics. *Kyklos*, *54*(2–3), 193–206.

Atkinson, A. B. (2009). Economics as a Moral Science. *Economica*, *76*, 791–804.

Atkinson, A. B. (2011). The Restoration of Welfare Economics. *American Economic Review*, *101*(3), 157–161.

Atkinson, A. B., & Brandolini, A. (2015). Unveiling the Ethics behind Inequality Measurement: Dalton's Contribution to Economics. *The Economic Journal*, *125*(583), 209–234. https://doi.org/10.1111/ecoj.12225

Backhouse, R. E. (2016, August). *The origins of the New Welfare Economics*. www.ier.hit-u.ac.jp/extra/10.Backhouse.pdf

Backhouse, R. E., & Medema, S. G. (2009). Robbins's Essay and the Axiomatization of Economics. *Journal of the History of Economic Thought*, *31*(04), 485.

Ball, R., & Chernova, K. (2008). Absolute Income, Relative Income, and Happiness. *Social Indicators Research*, *88*(3), 497–529. JSTOR.

Baron, J., & Greene, J. (1996). Determinants of Insensitivity to Quantity in Valuation of Public Goods: Contribution, Warm Glow, Budget Constraints, Availability, and Prominence. *Journal of Experimental Psychology: Applied*, *2*, 107–125. https://doi.org/10.1037/1076-898X.2.2.107

Baron, J., & McCaffery, E. J. (2004). *Masking Redistribution (or its Absence)* (SSRN Scholarly Paper ID 528165). Social Science Research Network. http://papers.ssrn.com/abstract=528165

Barron, F. H., Winterfeldt, D. von, & Fischer, G. W. (1984). Empirical and theoretical relationships between value and utility functions. *Acta Psychologica*, *56*(1–3), 233–244. https://doi.org/10.1016/0001-6918(84)90021-0

Barry, D., & Oelschlaeger, M. (1996). A Science for Survival: Values and Conservation Biology. *Conservation Biology*, *10*(3), 905–911.

Bartels, L. M. (2004, February). *Partisan Politics and the U.S. Income Distribution*. Princeton University. http://www.princeton.edu/~bartels/income.pdf

Bartels, L. M. (2005a). Homer Gets a Tax Cut: Inequality and Public Policy in the American Mind. *Perspectives on Politics*, *3*(01). https://doi.org/10.1017/S1537592705050036

Bartels, L. M. (2005b, August). *Economic Inequality and Political Representation*. Princeton University, Research Program in Political Economy. http://econpapers.repec.org/RePEc:ecl:prirpe:08-11-2005

Batz-Barbarich, C., Tay, L., Kuykendall, L., & Cheung, H. K. (2018). A Meta-Analysis of Gender Differences in Subjective Well-Being: Estimating Effect Sizes and Associations With Gender Inequality. *Psychological Science*, *29*(9), 1491–1503. https://doi.org/10.1177/0956797618774796

Baujard, A. (2013). *Welfare Economics* (SSRN Scholarly Paper 2357412). https://doi.org/10.2139/ssrn.2357412

Baujard, A. (2021). Values in Welfare Economics. In C. Heilmann & J. Reiss (Eds.), *The Routledge Handbook of the Philosophy of Economics*. Routledge.

Baumgaertner, B., & Holthuijzen, W. (2017). On nonepistemic values in conservation biology. *Conservation Biology*, *31*(1), 48–55. https://doi.org/10.1111/cobi.12756

Baumol, W. J. (1958). The Cardinal Utility Which is Ordinal. *The Economic Journal*, *68*(272), 665–672. https://doi.org/10.2307/2227278

Beckerman, W. (2010). *Economics as Applied Ethics: Value Judgments in Welfare Economics*. Palgrave Macmillan.

Beckman, S., Formby, J., & Smith, W. (2004). Efficiency, Equity and Democracy: Experimental Evidence on Okun's Leaky Bucket. *Research on Economic Inequality*, *11*, 17–42. https://doi.org/10.1016/S1049-2585(04)11002-8

Bennett, J. (1981). The Probable Gain from Egalitarian Redistribution. *Oxford Economic Papers*, *33*(1), 165–169. https://doi.org/10.2307/2662763

Bergsma, A., & Veenhoven, R. (2011). The happiness of people with a mental disorder in modern society. *Psychology of Well-Being: Theory, Research and Practice*, *1*(1), 2. https://doi.org/10.1186/2211-1522-1-2

Bergson, A. (1938). A Reformulation of Certain Aspects of Welfare Economics. *The Quarterly Journal of Economics*, *52*(2), 310–334. https://doi.org/10.2307/1881737

Bergson, A. (1954). On the Concept of Social Welfare. *The Quarterly Journal of Economics*, *68*(2), 233–252. https://doi.org/10.2307/1884448

Besley, T. J., & Meghir, C. (1998). *Tax Based Savings Incentives* (World Bank Savings Project) [Unpublished paper]. World Bank. http://siteresources.worldbank.org/INTMACRO/Resources/besley.pdf

Betz, G. (2013). In defence of the value free ideal. *European Journal for Philosophy of Science*, *3*(2), 207–220. https://doi.org/10.1007/s13194-012-0062-x

Binmore, K. G. (2009). Interpersonal comparison of utility. In H. Kincaid & D. Ross (Eds.), *The Oxford Handbook of Philosophy of Economics* (pp. 540–559). Oxford University Press.

Birchler-Pedross, A., Schröder, C. M., Münch, M., Knoblauch, V., Blatter, K., Schnitzler-Sack, C., Wirz-Justice, A., & Cajochen, C. (2009). Subjective well-being is modulated by circadian phase, sleep pressure, age, and gender. *Journal of Biological Rhythms*, *24*(3), 232–242. https://doi.org/10.1177/0748730409335546

Bjørnskov, C., Dreher, A., Fischer, J. A. V., Schnellenbach, J., & Gehring, K. (2013). Inequality and happiness: When perceived social mobility and economic reality do not match. *Journal of Economic Behavior & Organization*, *91*, 75–92. https://doi.org/10.1016/j.jebo.2013.03.017

Black, D. (1948). On the Rationale of Group Decision-making. *Journal of Political Economy*, *56*(1), 23–34.

Blackorby, C., Bossert, W., & Donaldson, D. (2002). Utilitarianism and the theory of justice. In B.-H. of S. C. and Welfare (Ed.), *Handbook of Social Choice and Welfare* (Vol. 1, pp. 543–596). Elsevier.

Blackorby, C., & Donaldson, D. (1978). Measures of relative equality and their meaning in terms of social welfare. *Journal of Economic Theory*, *18*(1), 59–80. https://doi.org/10.1016/0022-0531(78)90042-X

Blackorby, C., & Donaldson, D. (1982). Ratio-Scale and Translation-Scale Full Interpersonal Comparability without Domain Restrictions: Admissible Social-Evaluation Functions. *International Economic Review*, *23*(2), 249–268.

Blackorby, C., & Donaldson, D. (1990). The Case against the Use of the Sum of Compensating Variations in Cost-Benefit Analysis. *The Canadian Journal of Economics / Revue Canadienne d'Economique*, *23*(3), 471–494.

Blanchflower, D. G., & Oswald, A. J. (2003). *Does Inequality Reduce Happiness? Evidence from the States of the USA from the 1970s to the 1990s*. Dartmouth college.

Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, *88*(7–8), 1359–1386.

Blanchflower, D. G., & Oswald, A. J. (2008). Hypertension and happiness across nations. *Journal of Health Economics*, *27*(2), 218–233.

Blaug, M. (1978). *Economic Theory in Retrospect*. Cambridge University Press.

Blaug, M. (1992). *The Methodology of Economics: Or, How Economists Explain* (2 edition). Cambridge University Press.

Blaug, M. (2007). The Fundamental Theorems of Modern Welfare Economics, Historically Contemplated. *History of Political Economy*, *39*(2), 185–207.

Boadway, R. (1974). The Welfare Foundations of Cost-Benefit Analysis. *The Economic Journal*, *84*(336), 926–939.

Boadway, R. (2012). *From Optimal Tax Theory to Tax Policy: Retrospective and Prospective Views*. MIT Press. https://mitpress.mit.edu/books/optimal-tax-theory-tax-policy

Boadway, R. (2017). Second-Best Theory: Ageing well at Sixty. *Pacific Economic Review*, *22*(2), 249–270. https://doi.org/10.1111/1468-0106.12225

Boadway, R., & Bruce, N. (1984). *Welfare Economics*. Basil Blackwell.

Bond, T. N., & Lang, K. (2014). *The Sad Truth About Happiness Scales* (Working Paper 19950; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w19950

Bonica, A., McCarty, N., Poole, K. T., & Rosenthal, H. (2013). Why Hasn't Democracy Slowed Rising Inequality? *Journal of Economic Perspectives*, *27*(3), 103–124.

Borsboom, D., & Mellenbergh, G. (2004). Why Psychometrics is Not Pathological A Comment on Michell. *Journal of Dental Research - J DENT RES*, *14*, 105–120. https://doi.org/10.1177/0959354304040200

Boulding, K. E. (1952). Welfare Economics. In H. S. Ellis, B. F. Haley, & M. W. Reder (Eds.), *A survey of contemporary economics* (Vol. 2, pp. 1–33). Irwin for the American Economic Association.

Boulding, K. E. (1969). Economics as a Moral Science. *The American Economic Review*, *59*(1), 1–12.

Braddon-Mitchell, D., & Jackson, F. (1999). The Divide and Conquer Path to Analytical Functionalism. *Philosophical Topics*, *26*(1–2), 71–88. https://doi.org/10.5840/philtopics1999261/235

Brandolini, A., & Carta, F. (2016). *Some Reflections on the Social Welfare Bases of the Measurement of Global Income Inequality* (Working Paper 1070). Bank of Italy. https://papers.ssrn.com/abstract=2867417

Brent, R. J. (1984). Use of Distributional Weights in Cost-Benefit Analysis: A Survey of Schools. *Public Finance Review*, *12*(2), 213–230.

Briggs, R. (2015). Transformative Experience and Interpersonal Utility Comparisons. *Res Philosophica*, *92*(2), 189–216. https://doi.org/10.11612/resphil.2015.92.2.7

Bright, L. K. (2018). Du Bois? Democratic Defence of the Value Free Ideal. *Synthese*, *195*(5), 2227–2245. https://doi.org/10.1007/s11229-017-1333-z

Bromley, D. W. (1990). The ideology of efficiency: Searching for a theory of policy analysis. *Journal of Environmental Economics and Management*, *19*(1), 86–107. https://doi.org/10.1016/0095-0696(90)90062-4

Bronsteen, J., Buccafusco, C., & Masur, J. (2013). Well-Being Analysis vs. Cost-Benefit Analysis. *Duke Law Journal*, *62*(8), 1603–1689.

Broome, J. (2008). Why Economics Needs Ethical Theory. In K. Basu & R. Kanbur (Eds.), *Arguments for a Better World: Essays in Honor of Amartya Sen: Volume I: Ethics, Welfare, and Measurement* (1 edition, pp. 7–14). Oxford University Press.

Broome, J. (2012). *Climate Matters: Ethics in a Warming World: Vol. Norton Global Ethics Series*. W. W. Norton & Company.

Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, *425*(6955), Article 6955. https://doi.org/10.1038/nature01963

Brown, C. (2005). Priority or Sufficiency; or Both? *Economics and Philosophy*, *21*(02), 199–220.

Brown, C. (2007). Prioritarianism for Variable Populations. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *134*(3), 325–361.

Bruni, L., & Porta, P. L. (2005). *Economics and Happiness*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/0199286280.001.0001/acprof-9780199286287

Bruni, L., & Sugden, R. (2007). The Road Not Taken: How Psychology Was Removed from Economics, and How It Might Be Brought Back. *The Economic Journal*, *117*(516), 146–173.

Brunner, E., Smith, G. D., Marmot, M., Canner, R., Bersinka, M., & O'Brien, J. (1996). Childhood social circumstances and psychosocial and behavioural factors as determinants of plasma fibrinogen. *Lancet*, *347*, 1008–1013.

Buttrick, N. R., & Oishi, S. (2017). The psychological consequences of income inequality. *Social and Personality Psychology Compass*, *11*(3). https://doi.org/10.1111/spc3.12304

Byrne, A. (2020). Inverted Qualia. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/qualia-inverted/

Cabanac, M. (1992). Pleasure: The common currency. *Journal of Theoretical Biology*, *155*(2), 173–200. https://doi.org/10.1016/s0022-5193(05)80594-6

Carson, R. T., & Mitchell, R. C. (1995). Sequencing and Nesting in Contingent Valuation Surveys. *Journal of Environmental Economics and Management*, *28*(2), 155–173. https://doi.org/10.1006/jeem.1995.1011

Case, A., & Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism* (1 edition). Princeton University Press.

Cass, D. (1965). Optimal growth in an aggregate model of capital accumulation. *Review of Economic Studies*, *32*, 233–240.

Chalmers, D. J. (2022). *Reality+—Virtual Worlds and the Problems of Philosophy*. W. W. Norton & Company.

Chang, H. (2004). *Inventing Temperature*. Oxford University Press. https://doi.org/10.1093/0195171276.001.0001

Chang, H. (2009). Operationalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/operationalism/

Chapman, S. J. (1913). The Utility of Income and Progressive Taxation. *The Economic Journal*, *23*(89), 25–35.

Charpentier, A. (2015). Insights from life history theory for an explicit treatment of trade-offs in conservation biology: Trade-offs in Biodiversity Conservation. *Conservation Biology*, *29*(3), 738–747. https://doi.org/10.1111/cobi.12442

Chin, C.-L., & Yao, G. (2014). Convergent Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1275–1276). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_573

Chipman, J. S. (1974). The welfare ranking of Pareto distributions. *Journal of Economic Theory*, *9*(3), 275–282.

Chipman, J. S., & Moore, J. C. (1978). The New Welfare Economics 1939-1974. *International Economic Review*, *19*(3), 547–584.

Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature*, *46*(1), 95–144.

Clark, A. E., Knabe, A., & Rätzel, S. (2010). Boon or bane? Others' unemployment, well-being and job insecurity. *Labour Economics*, *17*(1), 52–61. https://doi.org/10.1016/j.labeco.2009.05.007

Clark, A. E., & Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, *61*(3), 359–381.

Clark, A. E., & Senik, C. (2011, February). Is Happiness Different From Flourishing? Cross-Country Evidence from the ESS. *Revue d'Économie Politique*, *1*, 17–34.

Clark, A. E., Senik, C., & Yamada, K. (2017). When experienced and decision utility concur: The case of income comparisons. *Journal of Behavioral and Experimental Economics*, *70*, 1–9. https://doi.org/10.1016/j.socec.2017.07.002

Claveau, F. (2009). Interdependent preferences and policy stances in mainstream economics. *Erasmus Journal for Philosophy & Economics*, *2*(1), 1–28.

Cohen, T. (1992). Consequentialism Implies a Zero Rate of Intergenerational Discount. In *Philosophy, Politics, and Society* (pp. 162–168). Lasletr and Fishkin.

Cohen, T., & Parfit, D. (1992). Against the social discount rate. In *Philosophy, Politics, and Society* (pp. 144–161). Lasletr and Fishkin.

Colander, D. (2009). What Was "it" That Robbins Was Defining? *Journal of the History of Economic Thought*, *31*(04), 437–448.

Colander, D. (2015). Economic Theory Has Nothing to Say about Policy (and Principles Textbooks Should Tell Students That). *Eastern Economic Journal*, *41*(4), 461–465. https://doi.org/10.1057/eej.2015.31

Colander, D. (2018). The Scope and Method of Applied Policy Economics. *The American Economist*, *63*(2), 132–146. https://doi.org/10.1177/0569434518783386

Cook, E. (2022). Efficiently Unequal: The Global Rise of Kaldor-Hicks Neoliberalism. *Global Intellectual History*, 1–23. https://doi.org/10.1080/23801883.2022.2062423

Cooter, R., & Rappoport, P. (1984). Were the Ordinalists Wrong About Welfare Economics? *Journal of Economic Literature*, *22*(2), 507–530.

Copp, D. (1987). The Justice and Rationale of Cost-Benefit Analysis. *Theory and Decision*, *23*(1), 65. https://doi.org/10.1007/bf00127337

Costa, D. L., & Kahn, M. E. (2004). Changes in the Value of Life, 1940–1980. *Journal of Risk and Uncertainty*, *29*(2), 159–180. https://doi.org/10.1023/B:RISK.0000038942.18349.88

Cowell, F. A., & Gardner, A. (1999). *Welfare Weights* (Working Paper 202). Office of Fair Trading. http://www.oft.gov.uk/shared_oft/reports/consumer_protection/oft282.pdf

Cowen, T. (2004). Policy implications of zero discounting: An exploration in politics and morality. *Social Philosophy and Policy*, *21*(1), 121–140. https://doi.org/10.1017/S0265052504211062

Creedy, J. (2007). Policy Evaluation, Welfare Weights and Value Judgements: A Reminder. *Australian Journal of Labour Economics*, *10*(1), 1–15.

Crisp, R. (2003a). Egalitarianism and Compassion. *Ethics*, *114*(1), 119–126.

Crisp, R. (2003b). Equality, Priority, and Compassion. *Ethics*, *113*(4), 745–763.

Crisp, R. (2021). Well-Being. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/well-being/

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Dagum, C. (1990). On the relationship between income inequality measures and social welfare functions. *Journal of Econometrics*, *43*(1–2), 91–102. https://doi.org/10.1016/0304-4076(90)90109-7

Dahl, R. A. (1989). *Democracy and Its Critics*. Yale University Press.

Dalton, H. (1920). The Measurement of the Inequality of Incomes. *The Economic Journal*, *30*(119), 348–361. https://doi.org/10.2307/2223525

Dasgupta, P. (2019). *Time and the Generations: Population Ethics for a Diminishing Planet* (p. 344 Pages). Columbia University Press.

Davis, M. (2012). A Plea for Judgment. *Science and Engineering Ethics*, *18*(4), 789–808. https://doi.org/10.1007/s11948-011-9254-6

de Bartolome, C. A. M. (1995). Which tax rate do people use: Average or marginal? *Journal of Public Economics*, *56*(1), 79–96. https://doi.org/10.1016/0047-2727(93)01409-4

de Van Graaff, J. (1957). *Theoretical welfare economics*. Cambridge University Press.

DeLong, J. B. (2003). A Non-Socratic Dialogue on Social Welfare Functions. *DeLong: Long Form*. https://delong.typepad.com/delong_long_form/2003/10/a-non-socratic-dialogue-on-social-welfare-functions.html

Desser, A. S., Olsen, J. A., & Grepperud, S. (2013). Eliciting Preferences for Prioritizing Treatment of Rare Diseases: The Role of Opportunity Costs and Framing Effects. *PharmacoEconomics*, *31*(11), 1051–1061.

Desvousges, W. H., Johnson, F., Dunford, R. W., Boyle, K. J., Hudson, S. P., & Wilson, K. N. (2010). *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy* (2nd ed.). Research Triangle Institute. https://www.rti.org/rti-press-publication/measuring-nonuse-damages-using-contingent-valuation

Desvousges, W., Mathews, K., & Train, K. (2016). From Curious to Pragmatically Curious: Comment on "From Hopeless to Curious? Thoughts on Hausman's 'Dubious to Hopeless' Critique of Contingent Valuation." *Applied Economic Perspectives and Policy*, *38*(1), 174–182. https://doi.org/10.1093/aepp/ppv023

Di Tella, R., & MacCulloch, R. (2006). Some Uses of Happiness Data in Economics. *Journal of Economic Perspectives*, *20*(1), 25–46. https://doi.org/10.1257/089533006776526111

Diamond, P. A., & Hausman, J. A. (1994). Contingent Valuation: Is Some Number better than No Number? *The Journal of Economic Perspectives*, *8*(4), 45–64.

Diekmann, S., & Peterson, M. (2013). The Role of Non-Epistemic Values in Engineering Models. *Science and Engineering Ethics*, *19*(1), 207–218. https://doi.org/10.1007/s11948-011-9300-4

Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being*, *3*(1), 1–43. https://doi.org/10.1111/j.1758-0854.2010.01045.x

Diener, E., Inglehart, R., & Tay, L. (2012). Theory and Validity of Life Satisfaction Scales. *Social Indicators Research*, *112*(3), 497–527. https://doi.org/10.1007/s11205-012-0076-y

Diener, E., Ng, W., Harter, J., & Arora, R. (2010). Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology*, *99*(1), 52–61. https://doi.org/10.1037/a0018066

Diener, E., & Seligman, M. E. P. (2004). Beyond Money: Toward an Economy of Well-Being. *Psychological Science in the Public Interest*, *5*(1), 1–31. https://doi.org/10.1111/j.0963-7214.2004.00501001.x

Dixit, A., & Norman, V. (1980). *Theory of International Trade: A Dual, General Equilibrium Approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511628627

Dixit, A., & Norman, V. (1986). Gains from trade without lump-sum compensation. *Journal of International Economics*, *21*(1), 111–122. https://doi.org/10.1016/0022-1996(86)90008-5

Dobb, M. (1940). *Political economy and capitalism: Some essays in economic tradition* (2nd Ediiton). Routledge & Kegan Paul.

Dobes, L., & Bennett, J. (2009). Multi-Criteria Analysis: “Good Enough” for Government Work? *Agenda: A Journal of Policy Analysis and Reform*, *16*(3), 7–29.

Dockray, S., Grant, N., Stone, A. A., Kahneman, D., Wardle, J., & Steptoe, A. (2010). A Comparison of Affect Ratings Obtained with Ecological Momentary Assessment and the

Day Reconstruction Method. *Social Indicators Research*, *99*(2), 269–283. https://doi.org/10.1007/s11205-010-9578-7

Dockray, S., & Steptoe, A. (2010). Positive affect and psychobiological processes. *Neuroscience and Biobehavioral Reviews*, *35*(1), 69–75. https://doi.org/10.1016/j.neubiorev.2010.01.006

Dow, S. C. (2012). Babylonian Mode of Thought. In J. E. King (Ed.), *The Elgar Companion to Post Keynsian Economics*. Edward Elgar.

Drakopoulos, S. A. (1989). The Historical Perspective of the Problem of Interpersonal Comparisons of Utility. *Journal of Economic Studies*, *16*(4), EUM0000000000136. https://doi.org/10.1108/EUM0000000000136

Drakopoulos, S. A. (1990). The Implicit Psychology of the Theory of the Rational Consumer: An Interpretation. *Australian Economic Papers*, *29*(55), 182–198.

Drakopoulos, S. A. (2008). *The Concept Of Comparison Income: An Historical Perspective* (MPRA Paper 8713). University Library of Munich, Germany. http://ideas.repec.org/p/pra/mprapa/8713.html

Drakopoulos, S. A. (2010). *The history of the mainstream rejection of interdependent preferences* (MPRA Paper 23980). University Library of Munich, Germany. http://ideas.repec.org/p/pra/mprapa/23980.html

Drakopoulos, S. A. (2011). The neglect of comparison income: An historical perspective. *European Journal of the History of Economic Thought*, *18*(3), 441–464.

Drakopoulos, S. A. (2012). The History Of Attitudes Towards Interdependent Preferences. *Journal of the History of Economic Thought*, *34*(04), 541–557.

Drakopoulos, S. A. (2016). *Comparisons in Economic Thought: Economic Interdependency Reconsidered*. Routledge.

Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In P. A. David & M. Reder (Eds.), *Nations and Households in Economic Growth*.

Easterlin, R. A. (2017). Paradox Lost? *Review of Behavioral Economics*, *4*(4), 311–339.

Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., & Zweig, J. S. (2010). The happiness-income paradox revisited. *Proceedings of the National Academy of Sciences*, *107*(52), 22463–22468. https://doi.org/10.1073/pnas.1015962107

Easterlin, R. A., & O'Connor, K. J. (2020). *The Easterlin Paradox* (Discussion Paper 13923; IZA Discussion Papers). Institution of Labour Economics (IZA). https://docs.iza.org/dp13923.pdf

Eckersley, R. (2013). Subjective Wellbeing: Telling Only Half the Story. *Social Indicators Research*, *112*(3), 529–534. https://doi.org/10.1007/s11205-013-0239-5

Edgeworth, F. Y. (1881). *Mathematical Psychics-An Essay on the Application of Mathematics to the Moral Sciences*. Kegan Paul.

Eeckhoudt, L. R., & Hammitt, J. K. (2001). Background risks and the value of a statistical life. *Journal of Risk and Uncertainty*, *23*(3), 261–279.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology. II. *Journal of Personality and Social Psychology*, *58*(2), 342–353.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*(3), 414–420. https://doi.org/10.1037/0022-3514.54.3.414

Elster, J. (1993). *Interpersonal Comparisons of Well-being*. Cambridge University Press.

European Social Survey Core Scientific Team. (2016). *European Social Survey 2016* [Dataset]. https://www.europeansocialsurvey.org/data

Eva, B. (2019, April 30). *Principles of Indifference* [Preprint]. https://philsci-archive.pitt.edu/16041/

Evans, D. J. (2004). The elevated status of the elasticity of marginal utility of consumption. *Applied Economics Letters*, *11*(7), 443–447. https://doi.org/10.1080/1350485042000189541

Evans, D. J. (2005). The Elasticity of Marginal Utility of Consumption: Estimates for 20 OECD Countries*. *Fiscal Studies*, *26*(2), 197–224. https://doi.org/10.1111/j.1475-5890.2005.00010.x

Evans, D. J., Kula, E., & Sezer, H. (2005). Regional welfare weights for the UK: England, Scotland, Wales and Northern Ireland. *Regional Studies*, *39*(7), 923–937.

Evans, D. J., & Sezer, H. (2004). Social discount rates for six major countries. *Applied Economics Letters*, *11*(9), 557–560. https://doi.org/10.1080/135048504200028007

Evans, D. J., & Sezer, H. (2005). *Social discount rates for member countries of the European Union*. *32*, 47–59.

Fabian, M., Alexandrova, A., Coyle, D., Agarwala, M., & Felici, M. (2023). Respecting the subject in wellbeing public policy: Beyond the social planner perspective. *Journal of European Public Policy*, *30*(8), 1494–1517. https://doi.org/10.1080/13501763.2022.2093947

Fabricatore, A. N., Handal, P. J., & Fenzel, L. M. (2000). Personal spirituality as a moderator of the relationship between stressors and subjective well-being. *Journal of Psychology and Theology*, *28*(3), 221–228.

Fankhauser, S., Tol, R. S. J., & Pearce, D. W. (1997). The Aggregation of Climate Change Damages: A Welfare Theoretic Approach. *Environmental and Resource Economics*, *10*(3), 249–266. https://doi.org/10.1023/A:1026420425961

Feldstein, M. (1998). *Income Inequality and Poverty* (Working Paper 6770). National Bureau of Economic Research. https://doi.org/10.3386/w6770

Feldstein, M. (1999). Reducing poverty, not inequality. *The Public Interest*, *137*.

Fellner, W. (1967). Operational utility: The theoretical background and a measurement. In W. Fellner (Ed.), *Ten Economic Studies in the Tradition of Irving Fisher*. Wiley.

Feltz, A. (2007). The Knobe Effect: A Brief Overview. *The Journal of Mind and Behavior*, *28*(3/4), 265–277.

Feltz, A., & Cokely, E. (2008). The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions. *Cognitive Science Society. For Evaluation Only. Copyright*, 1771–1776.

Fennell, L., & McAdams, R. (2016). The Distributive Deficit in Law and Economics. *Minnesota Law Review*. https://scholarship.law.umn.edu/mlr/199

Ferrer-i-Carbonell, A. (2013). Happiness economics. *SERIEs*, *4*(1), 35–60.

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How Important Is Methodology for the Estimates of the Determinants of Happiness? *The Economic Journal*, *114*(497), 641–659.

Fetherstonhaugh, D., Slovic, P., Johnson, S., & Friedrich, J. (1997). Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing. *Journal of Risk and Uncertainty*, *14*(3), 283–300. https://doi.org/10.1023/A:1007744326393

Fisher, F. M. (1956). Income Distribution, Value Judgments, and Welfare. *The Quarterly Journal of Economics*, *70*(3), 380–424. https://doi.org/10.2307/1884231

Fisher, I. (1927). A statistical method for measuring “marginal utility” and testing the justice of a progressive income tax. In *The Economic Essays Contributed in Honour of J. Bates* (pp. 157–193). Macmillan.

Fitoussi, J. P., & Stiglitz, J. E. (2011). *On the measurement of social progress and well being: Some further thoughts* (Documents de Travail de l’OFCE 2011–19). Observatoire Francais des Conjonctures Economiques (OFCE). http://ideas.repec.org/p/fce/doctra/1119.html

Flavin, P. (2012). Income Inequality and Policy Representation in the American States. *American Politics Research*, *40*(1), 29–59.

Flavin, P., & Franko, W. W. (2016). Government's Unequal Attentiveness to Citizens' Political Priorities. *Policy Studies Journal*, n/a-n/a.

Fletcher, G. (2016). *The Philosophy of Well-Being: An Introduction* (1st edition). Routledge.

Fleurbaey, M. (2003). On the informational basis of social choice. *Social Choice and Welfare*, *21*(2), 347–384.

Fraassen, V., & C, B. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199278220.001.0001

Frank, M. G., & Ekman, P. (1993). Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles. *Humr*, *6*(1), 9–26. https://doi.org/10.1515/humr.1993.6.1.9

Frank, R. H. (2000). Why is Cost-Benefit Analysis so Controversial? *The Journal of Legal Studies*, *29*(S2), 913–930. https://doi.org/10.1086/468099

Frankfurt, H. (1987). Equality as a Moral Ideal. *Ethics*, *98*(1), 21–43. https://doi.org/10.1086/292913

Fraser, C. (2023). Mohism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/mohism/

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55. https://doi.org/10.1037/0022-3514.65.1.45

Frey, B. S. (2008). *Happiness: A revolution in economics*. MIT Press.

Frey, B. S., & Stutzer, A. (2002). *Happiness and Economics*. Princeton University Press.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, *3*(10), e1701381. https://doi.org/10.1126/sciadv.1701381

Friedman, M. (1953). The Methodology of Positive Economics. In *Essays in Positive Economics* (pp. 1–43). Chicago University Press.

Friedman, M., & Savage, L. J. (1948). The Utility Analysis of Choices Involving Risk. *Journal of Political Economy*, *56*(4), 279–304.

Frijters, P., Clark, A. E., Krekel, C., & Layard, R. (2020). A happy choice: Wellbeing as the goal of government. *Behavioural Public Policy*, *4*(2), 126–165. https://doi.org/10.1017/bpp.2019.39

Frijters, P., Krekel, C., Frijters, P., & Krekel, C. (2021). *A Handbook for Wellbeing Policy-Making: History, Theory, Measurement, Implementation, and Examples*. Oxford University Press.

Frisch, R. (1932). *New Methods of Measuring Marginal Utility*. J. C. B. Morh.

Frisch, R. (1959). A Complete Scheme for Computing All Direct and Cross Demand Elasticities in a Model with Many Sectors. *Econometrica*, *27*(2), 177–196. JSTOR. https://doi.org/10.2307/1909441

Gal, D., & Rucker, D. D. (2018). The Loss of Loss Aversion: Will It Loom Larger Than Its Gain? *Journal of Consumer Psychology*, *28*(3), 497–516. https://doi.org/10.1002/jcpy.1047

Geary, R. C. (1950). A Note on “A Constant-Utility Index of the Cost of Living.” *The Review of Economic Studies*, *18*(1), 65–66. https://doi.org/10.2307/2296107

Gedikli, C., Miraglia, M., Connolly, S., Bryan, M., & Watson, D. (2023). The relationship between unemployment and wellbeing: An updated meta-analysis of longitudinal evidence. *European Journal of Work and Organizational Psychology*, *32*(1), 128–144. https://doi.org/10.1080/1359432X.2022.2106855

Gibbard, A. (1986). Interpersonal Comparisons: Preference, Good, and the Intrinsic Reward of a Life. In J. Elster & A. Hylland (Eds.), *Foundations of social choice theory* (pp. 196–194). Cambridge University Press.

Gilens, M. (2005). Inequality and Democratic Responsiveness. *Public Opinion Quarterly*, *69*(5), 778–796. https://doi.org/10.1093/poq/nfi058

Gilens, M. (2009). Preference Gaps and Inequality in Representation. *PS: Political Science & Politics*, *42*(02), 335–341.

Gilens, M. (2014). *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press.

Gilens, M., & Page, B. I. (2014). Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. *Perspectives on Politics*, *12*(3), 564–581. https://doi.org/10.1017/S1537592714001595

Ginsburgh, V. (2017). Contingent Valuation, Willingness to Pay, and Willingness to Accept. In B. S. Frey & D. Iselin (Eds.), *Economic Ideas You Should Forget* (pp. 65–66). Springer International Publishing. https://doi.org/10.1007/978-3-319-47458-8_26

Goodin, R. E. (1995). *Utilitarianism as a Public Philosophy*. Cambridge University Press.

Gopavaram, S., Dev, J., Das, S., & Camp, L. J. (2021). *IoT Marketplace: Willingness-To-Pay vs. Willingness-To-Accept* (SSRN Scholarly Paper 3867652). https://doi.org/10.2139/ssrn.3867652

Gould, S. J. (1996). *The Mismeasure of Man* (Revised and Expanded edition). W. W. Norton & Company.

Greene, P., & Sullivan, M. (2015). Against Time Bias. *Ethics*, *125*(4), 947–970.

Gul, F., & Pesendorfer, W. (2008). The Case for Mindless Economics. In A. Caplin & A. Schotter (Eds.), *The Foundations of Positive and Normative Economics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195328318.001.0001

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior*, *37*(1), 29–41. https://doi.org/10.1007/s10919-012-0139-4

Gyrd-Hansen, D., Kjær, T., & Nielsen, J. S. (2012). Scope insensitivity in contingent valuation studies of health care services: Should we ask twice? *Health Economics*, *21*(2), 101–112.

Haack, S. (2009). *Evidence and inquiry: A pragmatist reconstruction of epistemology* (2nd, expanded ed ed.). Prometheus Books.

Haase, C. G., Long, A. K., & Gillooly, J. F. (2016). Energetics of stress: Linking plasma cortisol levels to metabolic rate in mammals. *Biology Letters*, *12*(1), 20150867. https://doi.org/10.1098/rsbl.2015.0867

Hagger, M. S., Gucciardi, D. F., & Chatzisarantis, N. L. D. (2017). On nomological validity and auxiliary assumptions: The importance of simultaneously testing effects in social cognitive theories applied to health behavior and some guidelines. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01933

Hájek, A. (2023). Interpretations of Probability. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/probability-interpret/

Hall, R. L. (1959). Reflections on the Practical Application of Economics. *The Economic Journal*, *69*(276), 639–652. https://doi.org/10.2307/2227663

Hamilton, C. (1998). Measuring Changes in Economic Welfare: The Genuine Progress Indicator for Australia. In R. Eckersley (Ed.), *Measuring Progress: Is life getting Better?* CSIRO.

Hammitt, J. K. (2021). Accounting for the Distribution of Benefits and Costs in Benefit–Cost Analysis. *Journal of Benefit-Cost Analysis*, *12*(1), 64–84. https://doi.org/10.1017/bca.2020.29

Hammond, P. J. (1991). Interpersonal comparisons of utility: Why and how they are and should be made. In J. Elster & J. E. Roemer (Eds.), *Interpersonal Comparisons of Well-Being* (pp. 200–254). Cambridge University Press. https://doi.org/10.1017/CBO9781139172387.008

Hands, D. W. (2009). Effective Tension in Robbins' Economic Methodology. *Economica*, *76*, 831–844.

Hands, D. W. (2010). Economics, psychology and the history of consumer choice theory. *Cambridge Journal of Economics*, *34*(4), 633–648.

Hands, D. W. (2011). Back to the Ordinalist Revolution: Behavioral Economic Concerns in Early Modern Consumer Choice Theory. *Metroeconomica*, *62*(2), 386–410.

Hands, D. W. (2014). Mark Blaug on the normativity of welfare economics. *Erasmus Journal for Philosophy and Economics*, *6*(3), 1–25. https://doi.org/10.23941/ejpe.v6i3.148

Hansson, S. O. (2004). Welfare, Justice, and Pareto Efficiency. *Ethical Theory and Moral Practice*, *7*(4), 361–380.

Harberger, A. C. (1971). Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay. *Journal of Economic Literature*, *9*(3), 785–797.

Harberger, A. C. (1978). On the Use of Distributional Weights in Social Cost-Benefit Analysis. *Journal of Political Economy*, *86*(2), S87–S120. https://doi.org/10.2307/1829758

Harris, M. N., Knott, R. J., Lorgelly, P. K., & Rice, N. (2020). Using externally collected vignettes to account for reporting heterogeneity in survey self-assessment. *Economics Letters*, *194*, 109325. https://doi.org/10.1016/j.econlet.2020.109325

Harrod, R. F. (1938). Scope and Method of Economics. *The Economic Journal*, *48*(191), 383–412.

Harsanyi, J. C. (1955). Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, *63*(4), 309–321.

Harsanyi, J. C. (1958). Ethics in Terms of Hypothetical Imperatives. *Mind*, *67*(267), 305–316.

Harsanyi, J. C. (1975). Nonlinear social welfare functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality? *Theory and Decision*, *6*(3), 311–332.

Harsanyi, J. C. (1976). Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality? In J. C. Harsanyi (Ed.), *Essays on Ethics, Social Behavior, and Scientific Explanation* (pp. 64–85). Springer Netherlands. https://doi.org/10.1007/978-94-010-9327-9_5

Harvard, S., Werker, G. R., & Silva, D. S. (2020). Social, ethical, and other value judgments in health economics modelling. *Social Science & Medicine*, *253*, 112975. https://doi.org/10.1016/j.socscimed.2020.112975

Hausman, D. M. (1992). *The Inexact and Separate Science of Economics*. Cambridge University Press.

Hausman, D. M. (1995). The Impossibility of Interpersonal Utility Comparisons. *Mind*, *104*(415), 473–490.

Hausman, D. M. (2010). Hedonism and Welfare Economics. *Economics & Philosophy*, *26*(3), 321–344. https://doi.org/10.1017/S0266267110000398

Hausman, D. M. (2011). *Why Satisfy Preferences?* (Working Paper 1124; Papers on Economics and Evolution). Max Planck Institute of Economics, Evolutionary Economics Group.

Hausman, D. M. (2012a). *Preference, value, choice, and welfare*. Cambridge University Press.

Hausman, D. M. (2012b). Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis, Adler. Oxford University Press, 2012, 634 pages. *Economics and Philosophy*, *28*(03), 435–443. https://doi.org/10.1017/S0266267112000326

Hausman, D. M., & McPherson, M. S. (1996). *Economic analysis and moral philosophy*. Cambridge University Press.

Hausman, D. M., McPherson, M., & Satz, D. (2016). *Economic Analysis, Moral Philosophy, and Public Policy*. Cambridge University Press.

Hawthorne, G., & Griffith, P. (2000). *The Friendship Scale: Development and properties*. https://doi.org/10.4225/03/59360c301845a

Hawtrey, R. G. (1946). The Need for Faith. *The Economic Journal*, *56*(223), 351–365. https://doi.org/10.2307/2226044

He, J., Van de Vijver, F. J. R., Fetvadjiev, V. H., de Carmen Dominguez Espinosa, A., Adams, B., Alonso–Arbiol, I., Aydinli–Karakulak, A., Buzea, C., Dimitrova, R., Fortin, A., Hapunda, G., Ma, S., Sargautyte, R., Sim, S., Schachner, M. K., Suryani, A., Zeinoun, P., & Zhang, R. (2017). On Enhancing the Cross–Cultural Comparability of Likert–Scale Personality and Value Measures: A Comparison of Common Procedures. *European Journal of Personality*, *31*(6), 642–657. https://doi.org/10.1002/per.2132

Heathwood, C. (2006). Desire Satisfactionism and Hedonism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *128*(3), 539–563.

Hendren, N., & Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies. *Quarterly Journal of Economics*, *135*(3), 1209–1318.

Hersch, G. (2015). Can an evidential account justify relying on preferences for well-being policy? *Journal of Economic Methodology*, *22*(3), 280–291. https://doi.org/10.1080/1350178X.2015.1071507

Hicks, J. R. (1941). The Rehabilitation of Consumers' Surplus. *The Review of Economic Studies*, *8*(2), 108–116. https://doi.org/10.2307/2967467

HM Treasury. (2003). *The Green Book: Appraisal and Evaluation in Central Government* (Treasury Guidance). HM Treasury. https://www.gov.uk/government/uploads/.../green_book_complete.pdf

HM Treasury. (2022). *The Green Book: Appraisal and Evaluation in Central Government* (Treasury Guidance). HM Treasury. https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent/the-green-book-2020

Hodgson, G. M. (2012). On the Limits of Rational Choice Theory. *Economic Thought*, *1*(1).

Horowitz, J. M., Igielnik, R., & Kochhar, R. (2020, January 9). *Most Americans Say There Is Too Much Economic Inequality in the U.S., but Fewer Than Half Call It a Top Priority*. Pew

Research Centre. https://www.pewresearch.org/social-trends/2020/01/09/most-americans-say-there-is-too-much-economic-inequality-in-the-u-s-but-fewer-than-half-call-it-a-top-priority/

Hubley, A. M. (2014). Discriminant Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1664–1667). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_751

Hudík, M. (2014). Reference-Dependence and Marginal Utility: Alt, Samuelson, and Bernardelli. *History of Political Economy*, *46*(4), 677–693. https://doi.org/10.1215/00182702-2815611

Huta, V. (2020). How distinct are eudaimonia and hedonia? It depends on how they are measured. *Journal of Well-Being Assessment*, *4*(3), 511–537. https://doi.org/10.1007/s41543-021-00046-4

Hwang, Y. K., & Lee, C. (2018). Relationship between Stress and Happiness in Middle School Students: Dual Mediation Effect of Growth Mindset and Self-Esteem. *Medico-Legal Update*, *18*, 248. https://doi.org/10.5958/0974-1283.2018.00053.1

Intemann, K. (2015). Distinguishing between legitimate and illegitimate values in climate modeling. *European Journal for Philosophy of Science*, *5*(2), 217–232. https://doi.org/10.1007/s13194-014-0105-6

Jayawickreme, E., & Pawelski, J. O. (2013). Positivity and the capabilities approach. *Philosophical Psychology*, *26*(3), 383–400. https://doi.org/10.1080/09515089.2012.660687

Johansson-Stenman, O. (2005). Distributional Weights in Cost-Benefit Analysis: Should We Forget about Them? *Land Economics*, *81*(3), 337–352.

Jolls, C. (1998). Behavioral economics analysis of redistributive legal rules. *Vanderbilt Law Review*, *51*(6), 1653–1677.

Jolls, C., Sunstein, C. R., & Thaler, R. (1998). A Behavioral Approach to Law and Economics. *Stanford Law Review*, *50*(5), 1471–1550. https://doi.org/10.2307/1229304

Jones, C. (2002). *The “Boadway Paradox” Revisited* (Working Paper 421; The Faculty of Economics and Commerce). Australian National University. https://openresearch-repository.anu.edu.au/handle/1885/40256

Jorm, A. F., & Ryan, S. M. (2014). Cross-national and historical differences in subjective well-being. *International Journal of Epidemiology*, *43*(2), 330–340. https://doi.org/10.1093/ije/dyt188

Kahneman, D. (1986). Comments on the contingent valuation method. In R. G. Cummings, D. S. Brookshire, W. D. Schulze, & N. J. Totowa (Eds.), *Valuing environmental goods: A state of the arts assessment of the contingent valuation method* (pp. 185–194). Roweman and Allanheld.

Kahneman, D. (2000). Experienced utility and objective happiness: A moment-based approach. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames* (pp. 673–692). Cambridge University Press. (Original work published 1984)

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489–16493. https://doi.org/10.1073/pnas.1011492107

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, *22*(1), 57–70. https://doi.org/10.1016/0095-0696(92)90019-S

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science (New York, N.Y.)*, *306*(5702), 1776–1780. https://doi.org/10.1126/science.1103572

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated Willingness to Pay for Public Goods: A Psychological Perspective. *Psychological Science*, *4*(5), 310–315. https://doi.org/10.1111/j.1467-9280.1993.tb00570.x

Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues. *Journal of Risk and Uncertainty*, *19*(1–3), 203–235. https://doi.org/10.1023/A:1007835629236

Kahneman, D., & Sugden, R. (2005). Experienced Utility as a Standard of Policy Evaluation. *Environmental and Resource Economics*, *32*(1), 161–181.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics*, *112*(2), 375–406. https://doi.org/10.1162/003355397555235

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(42), e2210412119. https://doi.org/10.1073/pnas.2210412119

Kaiser, M. I. (2019). Normativity in the Philosophy of Science. *Metaphilosophy*, *50*(1–2), 36–62. https://doi.org/10.1111/meta.12348

Kaldor, N. (1939). Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *The Economic Journal*, *49*(195), 549–552. https://doi.org/10.2307/2224835

Kalecki, M. (1943). Political Aspects of Full Employment. *The Political Quarterly*, *14*(4), 322–330. https://doi.org/10.1111/j.1467-923X.1943.tb01016.x

Kanya, L., Sanghera, S., Lewin, A., & Fox-Rushby, J. (2019). The criterion validity of willingness to pay methods: A systematic review and meta-analysis of the evidence. *Social Science & Medicine (1982)*, *232*, 238–261. https://doi.org/10.1016/j.socscimed.2019.04.015

Kaplow, L. (2005). The Value of a Statistical Life and the Coefficient of Relative Risk Aversion. *Journal of Risk and Uncertainty*, *31*(1), 23–34.

Kaplow, L., & Shavell, S. (1994). Why the Legal System Is Less Efficient than the Income Tax in Redistributing Income. *The Journal of Legal Studies*, *23*(2), 667–681.

Kaplow, L., & Shavell, S. (2000). Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income. *The Journal of Legal Studies*, *29*(2), 821–835.

Kaplow, L., & Shavell, S. (2002). *Fairness versus welfare*. Harvard University Press.

Karabarbounis, L. (2011). One Dollar, One Vote. *The Economic Journal*, *121*(553), 621–651. https://doi.org/10.1111/j.1468-0297.2010.02406.x

Kaysen, C. (1969). Model Makers and Decision Makers. In R. L. Heilbroner (Ed.), *Economic Means and Social Ends* (pp. 132–137). Prentice Hall.

Kelman, S. (1981). Cost-benefit analysis: An ethical critique. *Regulation*, *5*(1), 33.

Kelsey, D. (1988). Policies to Achieve a Better Distribution of Income: Or is a Dollar a Dollar? *Oxford Economic Papers*, *40*(3), 577–583.

Kemp, M. C., & Ng, Y.-K. (1976). On the Existence of Social Welfare Functions, Social Orderings and Social Decision Functions. *Economica*, *43*(169), 59–66.

Kennedy, C. (1952). The Economic Welfare Function and Dr. Littles Criterion. *The Review of Economic Studies*, *20*(2), 137–142.

Keynes, J. N. (1890). *The scope and method of political economy* (First Edition). Macmillan and Co.

Killingsworth, M. A. (2021). Experienced well-being rises with income, even above $75,000 per year. *Proceedings of the National Academy of Sciences*, *118*(4), e2016976118. https://doi.org/10.1073/pnas.2016976118

Kim, D. A., Benjamin, E. J., Fowler, J. H., & Christakis, N. A. (2016). Social connectedness is associated with fibrinogen level in a human social network. *Proceedings. Biological Sciences*, *283*(1837). https://doi.org/10.1098/rspb.2016.0958

Kingdon, G. G., & Knight, J. (2007). Community, comparisons and subjective well-being in a divided society. *Journal of Economic Behavior & Organization*, *64*(1), 69–90. https://doi.org/10.1016/j.jebo.2007.03.004

Kirk, R. (2023). Zombies. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/zombies/

Kirk, R., & Squires, R. (1974). Zombies v. Materialists. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *48*, 135–163.

Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology*, *72*(6), 1245–1267. https://doi.org/10.1037/0022-3514.72.6.1245

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology*, *91*(5), 890–903. https://doi.org/10.1037/0022-3514.91.5.890

Kitcher, P. (2011). *Science in a democratic society*. Prometheus Books.

Klappholz, K. (1964). Value Judgments and Economics. *The British Journal for the Philosophy of Science*, *15*(58), 97–114.

Klappholz, K., & Agassi, J. (1959). Methodological Prescriptions in Economics. *Economica*, *26*(101), 60–74.

Knight, J., & Gunatilaka, R. (2022). Income inequality and happiness: Which inequalities matter in China? *China Economic Review*, *72*, 101765. https://doi.org/10.1016/j.chieco.2022.101765

Knight, J., Song, L., & Gunatilaka, R. (2009). Subjective well-being and its determinants in rural China. *China Economic Review*, *20*(4), 635–649. https://doi.org/10.1016/j.chieco.2008.09.003

Knobe, J. (2019). Philosophical Intuitions Are Surprisingly Robust Across Demographic Differences. *Epistemology and Philosophy of Science*, *56*(2), 29–36. https://doi.org/10.5840/eps201956225

Knobe, J., & Nichols, S. (2017). Experimental Philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/

Kong, F., Hu, S., Wang, X., Song, Y., & Liu, J. (2015). Neural correlates of the happy life: The amplitude of spontaneous low frequency fluctuations predicts subjective well-being. *NeuroImage*, *107*, 136–145. https://doi.org/10.1016/j.neuroimage.2014.11.033

Koopmans, T. C. (1965). On the concept of optimal economic growth. In *The econometric approach to development planning*. North Holland.

Kozma, A., & Stones, M. J. (1983). Predictors of happiness. *Journal of Gerontology*, *38*(5), 626–628. https://doi.org/10.1093/geronj/38.5.626

Krauss, A., & Graham, C. (2013). *Subjective Wellbeing in Colombia: Some Insights on Vulnerability, Job Security, and Relative Incomes*. https://doi.org/10.1596/1813-9450-6672

Kristoffersen, I. (2015a). *Economic analysis of happiness and satisfaction: Selected methodological issues and applications* [Ph.D. dissertation, University of Western Australia]. https://research-repository.uwa.edu.au/en/publications/economic-analysis-of-happiness-and-satisfaction-selected-methodol/fingerprints/

Kristoffersen, I. (2015b). *Economic analysis of happiness and satisfaction: Selected methodological issues and applications*. https://www.semanticscholar.org/paper/Economic-analysis-of-happiness-and-satisfaction%3A-Kristoffersen/435df097970f6f35af5f7ab229cd99465612148e

Krutilla, J. V. (1981). Reflections of an applied welfare economist: Presidential Address Presented at the Annual Meeting of the Association of Environmental and Resource

Economists September 6, 1980, Denver, Colorado. *Journal of Environmental Economics and Management*, *8*(1), 1–10. https://doi.org/10.1016/0095-0696(81)90052-8

Kubiszewski, I., Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., & Aylmer, C. (2013). Beyond GDP: Measuring and achieving global genuine progress. *Ecological Economics*, *93*, 57–68. https://doi.org/10.1016/j.ecolecon.2013.04.019

Kyngdon, A. (2008). The Rasch Model from the Perspective of the Representational Theory of Measurement. *Theory & Psychology*, *18*(1), 89–109. https://doi.org/10.1177/0959354307086924

Lai, L. C. H., Cummins, R. A., & Lau, A. L. D. (2013). Cross-Cultural Difference in Subjective Wellbeing: Cultural Response Bias as an Explanation. *Social Indicators Research*, *114*(2), 607–619.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157.

Lange, O. (1936). On the Economic Theory of Socialism: Part One. *The Review of Economic Studies*, *4*(1), 53–71.

Lange, O. (1937). On the Economic Theory of Socialism: Part Two. *The Review of Economic Studies*, *4*(2), 123–142.

Latty, K. (2011). *Income distribution, growth and social-welfare: Towards an economic solution to the growth-equality trade-off problem* [Honours Thesis, University of Sydney]. http://hdl.handle.net/2123/8260

Latty, K. (2014, October 29). *A Kaleckian theory of tax inefficiency: How elite influence favours inefficient redistributive mechanisms* [Conference paper]. Getting the G20 to work against Corporate Tax Avoidance, Sydney University. https://www.academia.edu/10644811/A_Kaleckian_theory_of_tax_inefficiency_How_elite_influence_favours_inefficient_redistributive_mechanisms

Latty, K. (2015a). *A Generalised Logistic Prioritarian Social Welfare Function*. https://www.academia.edu/11250950/A_Generalised_Logistic_Prioritarian_Social_Welfare_Function

Latty, K. (2015b). *Optimal inequality on the Utilitarian-Prioritarian-Egalitarian spectrum*. https://www.academia.edu/9448100/Optimal_Inequality_on_the_Utilitarian-Prioritarian-Egalitarian_Spectrum

Latty, K. (2015c, July 17). *A five parameter Atkinson like index featuring relative income effects, with a seven parameter extension for nonlinear (prioritarian) social welfare functions*. https://www.academia.edu/6099318/A_five_parameter_Atkinson_like_index_featuring_relative_income_effects_with_a_seven_parameter_extension_for_nonlinear_prioritarian_social_welfare_functions

Lau, A. L. D., Cummins, R. A., & McPherson, W. (2005). An investigation into the cross-cultural equivalence of the personal wellbeing index. *Social Indicators Research*, *72*(3), 403–430. https://doi.org/10.1007/s11205-004-0561-z

Layard, R. (2005). *Happiness: Lessons from a New Science*. Penguin.

Layard, R., Mayraz, G., & Nickell, S. (2008). The marginal utility of income. *Journal of Public Economics*, *92*(8–9), 1846–1857.

Layte, R., & Whelan, C. T. (2014). Who Feels Inferior? A Test of the Status Anxiety Hypothesis of Social Inequalities in Health. *European Sociological Review*, *30*(4), 525–535. https://doi.org/10.1093/esr/jcu057

Lee, C.-S. (2005). Income Inequality, Democracy, and Public Sector Size. *American Sociological Review*, *70*(1), 158–181.

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a likert scale. *Research in Nursing & Health*, *25*(4), 295–306. https://doi.org/10.1002/nur.10041

Lerner, A. P. (1944). *The Economics of Control*. Macmillan.

Lerner, A. P. (1969). On Instrumental Analysis. In R. L. Heilbroner (Ed.), *Economic Means and Social Ends* (pp. 131–136). Prentice Hall.

Lewis, D. K. (1966). An Argument for the Identity Theory. *Journal of Philosophy*, *63*(1), 17–25. https://doi.org/10.2307/2024524

Liebman, J. B., & Zeckhauser, R. J. (2004, October). *Schmeduling*. Harvard University. http://www.hks.harvard.edu/jeffreyliebman/schmeduling.pdf

Linley, P., Maltby, J., Wood, A., Osborne, G., & Hurling, R. (2009). Measuring happiness: The higher order factor structure of subjective and psychological well-being measures. *Personality and Individual Differences*, *47*, 878–884. https://doi.org/10.1016/j.paid.2009.07.010

Lipsey, R. G. (2007). Reflections on the general theory of second best at its golden jubilee. *International Tax and Public Finance*, *14*(4), 349–364.

Lipsey, R. G., & Lancaster, K. (1956). The General Theory of Second Best. *The Review of Economic Studies*, *24*(1), 11–32.

Little, I. M. D. (1950). *A critique of welfare economics* (1st edition). Clarendon Press.

Little, I. M. D. (2002a). *Ethics, Economics, and Politics*. Oxford University Press. https://doi.org/10.1093/0199257043.001.0001

Little, I. M. D. (2002b). *Ethics, economics, and politics: Principles of public policy*. Oxford University Press.

Liuzzo, C. (2020, October 11). *A question for the treasurer: How do you treat mental health without measuring well-being?* The Conversation. http://theconversation.com/a-question-for-the-treasurer-how-do-you-treat-mental-health-without-measuring-well-being-147572

Lloren, A., & Wuest, R. (2013). Who Represents the Poor? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2321442

Lopez, H. (2008). *The Social Discount Rate:* (4639; Policy Research Working Paper). The World Bank. http://wwwwds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2008/06/03/000158349_20080603084938/Rendered/PDF/wps4639.pdf

Lowe, A. (1969). Towards a Science of Political Economics. In R. L. Heilbroner (Ed.), *Economic Means and Social Ends* (pp. 1–36). Prentice Hall.

Lu, L. (1995). The Relationship Between Subjective Well-Being and Psychosocial Variables in Taiwan. *The Journal of Social Psychology*, *135*(3), 351–357. https://doi.org/10.1080/00224545.1995.9713964

Lu, L., & Gilmour, R. (2004). Culture and conceptions of happiness: Individual oriented and social oriented swb. *Journal of Happiness Studies*, *5*(3), 269–291. https://doi.org/10.1007/s10902-004-8789-5

Lu, L., & Gilmour, R. (2006). Individual-oriented and socially oriented cultural conceptions of subjective well-being: Conceptual analysis and scale development. *Asian Journal Of Social Psychology*, *9*(1), 36–49. https://doi.org/10.1111/j.1467-839X.2006.00183.x

Lucas, C. G., Kemp, C., & Griffiths, T. L. (2011). From preferences to choices and back again: Evidence for human inconsistency and its implications. In L. A. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *CogSci 2011*. cognitivesciencesociety.org. https://mindmodeling.org/cogsci2011/papers/0274/index.html

Lusk, G. (2021). Does democracy require value-neutral science? Analyzing the legitimacy of scientific information in the political sphere. *Studies in History and Philosophy of Science Part A*, *90*, 102–110. https://doi.org/10.1016/j.shpsa.2021.08.009

Lyubomirsky, S., & Lepper, H. S. (1999). A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation. *Social Indicators Research*, *46*(2), 137–155. https://doi.org/10.1023/A:1006824100041

Mach, E. (1903). Critique of the Concept of Temperature. (Concluded.). *The Open Court*, *1903*(3). https://opensiuc.lib.siu.edu/ocj/vol1903/iss3/4

Machlup, F. (1969). Positive and Normative Economics. In R. L. Heilbroner (Ed.), *Economic Means and Social Ends* (pp. 99–129). Prentice Hall.

Mack, R. P. (1952). Economics of Consumption. In H. S. Ellis, B. F. Haley, & M. W. Reder (Eds.), *A survey of contemporary economics* (Vol. 2, pp. 39–83). Irwin for the American Economic Association.

MacKerron, G. (2012). Happiness Economics from 35 000 Feet. *Journal of Economic Surveys*, *26*(4), 705–735. https://doi.org/10.1111/j.1467-6419.2010.00672.x

Marmot, M. (2004). *The Status Syndrome: How Social Standing Affects Our Health and Longevity*. Bloomsbury.

Marmot, M., & Wilkinson, R. G. (2001). Psychosocial and material pathways in the relation between income and health: A response to Lynch et al. *BMJ*, *322*(7296), 1233–1236. https://doi.org/10.1136/bmj.322.7296.1233

Martín-Fernández, J., López-Nicolás, Á., Oliva-Moreno, J., Medina-Palomino, H., Polentinos-Castro, E., & Ariza-Cardiel, G. (2021). Risk aversion, trust in institutions and contingent valuation of healthcare services: Trying to explain the WTA-WTP gap in the Dutch population. *Cost Effectiveness and Resource Allocation : C/E*, *19*, 27. https://doi.org/10.1186/s12962-021-00281-9

Masini, F. (2009). Economics and Political Economy in Lionel Robbins's Writings. *Journal of the History of Economic Thought*, *31*(04), 421. https://doi.org/10.1017/S105383720999023X

Matsumori, K., Iijima, K., Yomogida, Y., & Matsumoto, K. (2021). *Interpersonal Comparison of Utility by Measuring Neural Activity* (p. 2021.06.04.447048). bioRxiv. https://doi.org/10.1101/2021.06.04.447048

McCaffery, E. J., & Baron, J. (2006). Thinking about tax. *Psychology, Public Policy, and Law*, *12*(1), 106–135. https://doi.org/10.1037/1076-8971.12.1.106

McDonald, M. P. (2005). Validity, Data Sources. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 939–948). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00046-3

McFadden, D., & Leonard, G. (1993). Issues in the contingent valuation of environmental goods: Methodologies for data collection and analysis. In J. A. Hausman (Ed.), *Contingent Valuation: A Critical Assesment* (pp. 162–215). North Holland.

McFadden, D., Machina, M. J., & Baron, J. (2000). Rationality for economists? *Elicitation of Preferences*, 73–110.

McGregor, T., Smith, B., & Wills, S. (2019). Measuring inequality. *Oxford Review of Economic Policy*, *35*(3), 368–395. https://doi.org/10.1093/oxrep/grz015

McKenna, P. (2008). Vote of no confidence. *New Scientist*, *198*(2651), 30–33. https://doi.org/10.1016/S0262-4079(08)60914-8

Meade, J. E. (1959). Review of A Critique of Welfare Economics. *Economic Journal*, *1*(69), 124–129.

Medin, H., Nyborg, K., & Bateman, I. (2001). The assumption of equal marginal utility of income: How much does it matter? *Ecological Economics*, *36*(3), 397–411.

Michell, J. (2000). Normal Science, Pathological Science and Psychometrics. *Theory & Psychology*, *10*(5), 639–667. https://doi.org/10.1177/0959354300105004

Milanovic, B. (2007). Why We All Care about Inequality (But Some of Us Are Loathe to Admit It). *Challenge*, *50*(6), 109–120.

Mill, J. S. (1909). *Principles of political economy with some of their applications to social philosophy* (W. J. Ashley, Ed.). Longmans Green & Co. (Original work published 1848)

Mirrlees, J. A. (1969, December 9). *Information and Incentives: The Economics of Carrots and Sticks*. Nobel Prize. https://www.nobelprize.org/uploads/2018/06/mirrlees-lecture.pdf

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies*, *38*(2), 175–208.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics*, *6*(4), 327–358.

Mirrlees, J. A. (1986). The theory of optimal taxation. In *Handbook of mathematical economics*.

Mishan, E. J. (1957). An Investigation into Some Alleged Contradictions in Welfare Economics. *The Economic Journal*, *67*(267), 445–454. https://doi.org/10.2307/2227361

Mishan, E. J. (1960). A Survey of Welfare Economics, 1939-59. *The Economic Journal*, *70*(278), 197–265. https://doi.org/10.2307/2228726

Mishan, E. J. (1967). Pareto Optimality and the Law. *Oxford Economic Papers*, *19*(3), 255–287.

Morreau, M. (2019). Arrow's Theorem. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/arrows-theorem/

Morton, J. (n.d.). *Welfare Economics: For Light or for Fruit*. Trinity College, Dublin. https://www.tcd.ie/Economics/assets/pdf/SER/1994/Justin_Morton.html

Mosini, V. (2012). *Reassessing the Paradigm of Economics Bringing Positive Economics Back into the Normative Framework* (Vol. 12). Routledge.

Mourao, P. R. (2008). Towards a Puviani's Fiscal Illusion Index. *Hacienda Pública Española*, *187*(4), 49–86.

Mòzǐ. (2020). *The Essential Mòzǐ: Ethical, Political, and Dialectical Writings* (C. Fraser, Trans.; Abridged edition). Oxford University Press. (Original work published 5th–3rd centuries BC)

Murali, V., & Oyebode, F. (2004). Poverty, social inequality and mental health. *Advances in Psychiatric Treatment*, *10*(3), 216–224. https://doi.org/10.1192/apt.10.3.216

Murphy, L. B., & Nagel, T. (2004). *The Myth of Ownership: Taxes and Justice*. Oxford University Press.

Myrdal, G. (1970). *Objectivity in social research: Vol. 1967.* Duckworth.

Myrdal, G. (with Swedberg, R.). (1953). *The Political Element in the Development of Economic Theory*. Routledge & Kegan Paul.

Naess, A. (1984). A Defence of the Deep Ecology Movement. *Environmental Ethics*, *6*(3), 265–270. https://doi.org/10.5840/enviroethics19846330

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World.

Nath, S. K. (1969). *A reappraisal of welfare economics*. Routledge & K. Paul.

Ne'eman, A., Bell, E., Schneider, M. C., & Strolovitch, D. (2022). Identifying And Exploring Bias In Public Opinion On Scarce Resource Allocation During The COVID-19 Pandemic. *Health Affairs*, *41*(10), 1513–1522. https://doi.org/10.1377/hlthaff.2022.00504

Nelson, R. H. (1987). The Economics Profession and the Making of Public Policy. *Journal of Economic Literature*, *25*(1), 49–91.

Nemati, S., & Mehdipour, F. (2016). The Relationship between Life Satisfaction and Happiness: The Mediating Role of Resiliency. *International Journal of Psychological Studies*, *8*, 194. https://doi.org/10.5539/ijps.v8n3p194

Ng, N. Y., & Ruger, J. P. (2011). Global Health Governance at a Crossroads. *Global Health Governance : The Scholarly Journal for the New Health Security Paradigm*, *3*(2), 1–37.

Ng, Y.-K. (1972). Value Judgments and Economists' Role in Policy Recommendation. *The Economic Journal*, *82*(327), 1014–1018.

Ng, Y.-K. (1975). Bentham or Bergson? Finite Sensibility, Utility Functions and Social Welfare Functions. *The Review of Economic Studies*, *42*(4), 545–569.

Ng, Y.-K. (1982). Beyond Pareto Optimality: The Necessity of Interpersonal Cardinal Utilities in Distributional Judgements and Social Choice. *Zeitschrift Für Nationalökonomie / Journal of Economics*, *42*(3), 207–233.

Ng, Y.-K. (1984a). Expected Subjective Utility: Is the Neumann-Morgenstern Utility the Same as the Neoclassical's? *Social Choice and Welfare*, *1*(3), 177–186.

Ng, Y.-K. (1984b). Interpersonal Level Comparability Implies Comparability of Utility Differences. *Theory and Decision*, *17*(2), 141–147.

Ng, Y.-K. (1985). Some fundamental issues in social welfare. In G. R. Feiwel (Ed.), *Issues in Contemporary Microeconomics and Welfare* (pp. 435–469). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-06876-0

Ng, Y.-K. (1988). Is a Dollar a Dollar? A Response. *Oxford Economic Papers*, *40*(3), 584–586.

Ng, Y.-K. (1995). Towards welfare biology: Evolutionary economics of animal consciousness and suffering. *Biology and Philosophy*, *10*(3), 255–285.

Ng, Y.-K. (1996). Happiness surveys: Some comparability issues and an exploratory survey based on just perceivable increments. *Social Indicators Research*, *38*(1), 1–27. https://doi.org/10.1007/BF00293784

Ng, Y.-K. (1997). A Case for Happiness, Cardinalism, and Interpersonal Comparability. *The Economic Journal*, *107*(445), 1848–1858.

Ng, Y.-K. (1999). Utility, informed preference, or happiness: Following Harsanyi's argument to its logical conclusion. *Social Choice and Welfare*, *16*(2), 197–216.

Ng, Y.-K. (2000). *The Optimal Size of Public Spending and the Distortionary Cost of Taxation* (Monash Economics Working Paper archive-19). Monash University, Department of Economics. http://ideas.repec.org/p/mos/moswps/archive-19.html

Ng, Y.-K. (2002). East-Asian Happiness Gap. *Pacific Economic Review*, *7*(1), 51–63. https://doi.org/10.1111/1468-0106.00149

Ng, Y.-K. (2003a). From preference to happiness: Towards a more complete welfare economics. *Social Choice and Welfare*, *20*(2), 307–350.

Ng, Y.-K. (2003b). *Welfare Economics*. Palgrave Macmillan.

Ng, Y.-K. (2011). Consumption tradeoff vs. catastrophes avoidance: Implications of some recent results in happiness studies on the economics of climate change. *Climatic Change*, *105*(1–2), 109–127. https://doi.org/10.1007/s10584-010-9880-z

Ng, Y.-K. (2015). Some Conceptual and Methodological Issues on Happiness: Lessons from Evolutionary Biology. *Singapore Economic Review*, *60*(4), 1–17.

Ng, Y.-K., & Ho, L. (2006). *Happiness and public policy: Theory, case studies and implications*. Palgrave Macmillan. http://ezproxy.library.usyd.edu.au/login?url=http://www.palgraveconnect.com/doifinder/10.1057/9780230288027

Nordhaus, W., & Tobin, J. (1972). Is Growth Obsolete. In *Economic Research: Retrospect and Prospect* (Vol. 5, pp. 1–80). NBER. http://www.nber.org/chapters/c7620

Norinder, A., Hjalte, K., & Persson, U. (2001). Scope and scale insensitivities in a contingent valuation study of risk reductions. *Health Policy*, *57*(2), 141–153. https://doi.org/10.1016/S0168-8510(01)00117-8

Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.

Nussbaum, M. C. (1992). Human Functioning and Social Justice: In Defense of Aristotelian Essentialism. *Political Theory*, *20*(2), 202–246.

Nussbaum, M. C. (2007). *Frontiers of Justice: Disability, Nationality, Species Membership: 4* (1st edition). *Belknap Press.

Nyborg, K., & Spangen, I. (1997). *Cost-Benefit Analysis and the Democratic Ideal* [Discussion Paper]. Statistics Norway, Research Department. https://econpapers.repec.org/paper/ssbdispap/205.htm

Nyborg, K., & Spangen, I. (2000). Cost-Benefit Analysis and the Democratic Ideal. *Nordic Journal of Political Economy*, *26*, 83–93.

OECD. (2013). *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing. http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm

Oishi, S. (2002). The Experiencing and Remembering of Well-Being: A Cross-Cultural Analysis. *Personality and Social Psychology Bulletin*, *28*(10), 1398–1406. https://doi.org/10.1177/014616702236871

Oishi, S., & Diener, E. (2014). Can and Should Happiness Be a Policy Goal? *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 195–203. https://doi.org/10.1177/2372732214548427

Olsen, J. A., Donaldson, C., & Pereira, J. (2004). The insensitivity of `willingness-to-pay' to the size of the good: New evidence for health care. *Journal of Economic Psychology*, *25*(4), 445–460. https://doi.org/10.1016/S0167-4870(03)00029-1

Orben, A., & Lakens, D. (2020). Crud (Re)Defined. *Advances in Methods and Practices in Psychological Science*, *3*(2), 238–247. https://doi.org/10.1177/2515245920917961

Osmani, S. R. (1978). On the Normative Measurement of Inequality. *The Bangladesh Development Studies*, *6*(4), 417–442.

Oswald, A. J. (2008). On the curvature of the reporting function from objective reality to subjective feelings. *Economics Letters*, *100*(3), 369–372.

Padovano, F., & Galli, E. (2002). Comparing the growth effects of marginal vs. Average tax rates and progressivity. *European Journal of Political Economy*, *18*(3), 529–544. https://doi.org/10.1016/S0176-2680(02)00104-0

Page, B. I., & Gilens, M. (2017). *Democracy in America?: What Has Gone Wrong and What We Can Do About It*. University of Chicago Press.

Panagiotakos, D. B., Pitsavos, C., Chrysohoou, C., Tsetsekou, E., Papageorgiou, C., Christodoulou, G., Stefanadis, C., & ATTICA study. (2004). Inflammation, coagulation, and depressive symptomatology in cardiovascular disease-free people; the ATTICA study. *European Heart Journal*, *25*(6), 492–499. https://doi.org/10.1016/j.ehj.2004.01.018

Pareto, V. (2018). *Manuale di Economia Politica: Con una Introduzione Alla Scienza Sociale*. Forgotten Books.

Parfit, D. (1984). *Reasons and persons*. Clarendon.

Parfit, D. (1995). *Equality or Priority?* University of Kansas.

Parfit, D. (1997). Equality and Priority. *Ratio*, *10*(3), 202–221.

Parks, R. P. (1976). An Impossibility Theorem for Fixed Preferences: A Dictatorial Bergson-Samuelson Welfare Function. *The Review of Economic Studies*, *43*(3), 447–450. https://doi.org/10.2307/2297221

Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, *5*(2), 164–172. https://doi.org/10.1037/1040-3590.5.2.164

Percoco, M. (2008). A social discount rate for Italy. *Applied Economics Letters*, *15*(1), 73–77. https://doi.org/10.1080/13504850600706537

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology & Sociobiology*, *17*(3), 145–171. https://doi.org/10.1016/0162-3095(96)00041-6

Phelan, M., & Buckwalter, W. (2012). Analytic Functionalism and Mental State Attribution. *Philosophical Topics*, *40*(2), 129–154. https://doi.org/10.5840/philtopics201240217

Philippi, C. L. (2021). On Measurement Scales: Neither Ordinal Nor Interval? *Philosophy of Science*, *88*(5), 929–939. https://doi.org/10.1086/714873

Pigou, A. C. (1903). Some Remarks on Utility. *The Economic Journal*, *13*(49), 58–68.

Pigou, A. C. (1912). *Wealth and Welfare*. McMillan.

Pigou, A. C. (1932). *The Economics of Welfare* (4th ed.). Macmillan. (Original work published 1920)

Pigou, A. C. (1951). Some Aspects of Welfare Economics. *The American Economic Review*, *41*(3), 287–302.

Pirttilä, J., & Uusitalo, R. (2010). A ‘Leaky Bucket’ in the Real World: Estimating Inequality Aversion using Survey Data. *Economica*, *77*(305), 60–76. https://doi.org/10.1111/j.1468-0335.2008.00729.x

Pivato, M. (2013). Risky social choice with incomplete or noisy interpersonal comparisons of well-being. *Social Choice and Welfare*, *40*(1), 123–139. https://doi.org/10.1007/s00355-011-0590-x

Plant, M. (2020a). *A Happy Possibility About Happiness (And Other Subjective) Scales: An Investigation and Tentative Defence of the Cardinality Thesis 1,2*. https://www.semanticscholar.org/paper/A-Happy-Possibility-About-Happiness-%28And-Other-An-Plant/ba55211c1782257d568c597d345b36158b8892d6

Plant, M. (2020b). *Can I Get a Little Less Satisfaction, Please?* https://philarchive.org/rec/PLACIG

Polinsky, A. M. (1972). Probabilistic Compensation Criteria. *The Quarterly Journal of Economics*, *86*(3), 407. https://doi.org/10.2307/1880801

Ponthière, G. (2003). *Should we discount future generations’ welfare? A survey on the “pure” discount rate debate* (CREPP Working Paper 0302). Centre de Recherche en Economie Publique et de la Population (CREPP) (Research Center on Public and Population Economics) HEC-Management School, University of Liège. https://ideas.repec.org/p/rpp/wpaper/0302.html

Posner, E. A. (1996). Law, Economics, and Inefficient Norms. *University of Pennsylvania Law Review*, *144*(5), 1697–1744. https://doi.org/10.2307/3312637

Posner, R. A. (1977). *Economic analysis of Law* (2nd ed.). Little, Brown and Company.

Prati, A., & Senik, C. (2020). *Feeling good or feeling better?* [Working Paper]. HAL. https://econpapers.repec.org/paper/halwpaper/halshs-02545228.htm

Prescott, S., & Csikszentmihalyi, M. (1981). ENVIRONMENTAL EFFECTS ON COGNITIVE AND AFFECTIVE STATES: THE EXPERIENTIAL TIME SAMPLING APPROACH.

Social Behavior and Personality: An International Journal, *9*(1), 23–32. https://doi.org/10.2224/sbp.1981.9.1.23

Price, D. D., Barrell, J. E., & Barrell, J. J. (1985). A quantitative-experiential analysis of human emotions. *Motivation and Emotion*, *9*(1), 19–38. https://doi.org/10.1007/BF00991548

Proctor, C., & Tweed, R. (2016). Measuring eudaimonic well-being. In *Handbook of eudaimonic well-being* (pp. 277–294). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-42445-3_18

Quggin, J. (2014). What happiness conceals. *Aeon*. https://aeon.co/essays/misery-has-far-more-to-teach-us-than-happiness-does

Rakotonarivo, O. S., Schaafsma, M., & Hockley, N. (2016). A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. *Journal of Environmental Management*, *183*, 98–109. https://doi.org/10.1016/j.jenvman.2016.08.032

Ramsey, F. P. (1928). A Mathematical Theory of Saving. *The Economic Journal*, *38*(152), 543–559.

Rand, A. (2011). *The Objectivist Ethics*. Literary Licensing, LLC.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*(1), 3–8. https://doi.org/10.1016/0304-3959(96)02994-6

Reder, M. W. (1952). Comment on "Welfare Economics" by Melvin W. Reder. In H. S. Ellis, B. F. Haley, & M. W. Reder (Eds.), *A survey of contemporary economics* (Vol. 2, pp. 34–36). Irwin for the American Economic Association.

Reder, M. W. (1958). Theoretical Welfare Economics: A Comment. *Kyklos*, *11*(1), 84–87. https://doi.org/10.1111/j.1467-6435.1958.tb00249.x

Reiss, J., & Ankeny, R. A. (2022). Philosophy of Medicine. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/medicine/

Reiss, J., & Sprenger, J. (2014). *Scientific Objectivity*. https://stanford.library.sydney.edu.au/archives/win2018/entries/scientific-objectivity/#IntProProObj

Ribeiro, W. S., Bauer, A., Andrade, M. C. R., York-Smith, M., Pan, P. M., Pingani, L., Knapp, M., Coutinho, E. S. F., & Evans-Lacko, S. (2017). Income inequality and mental illness-related morbidity and resilience: A systematic review and meta-analysis. *The Lancet Psychiatry*, *4*(7), 554–562. https://doi.org/10.1016/S2215-0366(17)30159-1

Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science* (1st edition). Macmillan.

Robbins, L. (1938). Interpersonal Comparisons of Utility: A Comment. *The Economic Journal*, *48*(192), 635–641.

Robbins, L. (1981). Economics and Political Economy. *The American Economic Review*, *71*(2), 1–10.

Roberts, K. W. S. (1980a). Interpersonal Comparability and Social Choice Theory. *The Review of Economic Studies*, *47*(2), 421–439. https://doi.org/10.2307/2297002

Roberts, K. W. S. (1980b). Possibility Theorems with Interpersonally Comparable Welfare Levels. *The Review of Economic Studies*, *47*(2), 409–420. https://doi.org/10.2307/2297001

Robertson, D. H. (1954). Utility and All What? *The Economic Journal*, *64*(256), 665–678.

Robertson, D. H. (1962). Welfare Criteria: An Exchange of Notes: A Note. *The Economic Journal*, *72*(285), 226–229.

Robeyns, I., & Byskov, M. F. (2023). The Capability Approach. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023). Metaphysics

Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/capability-approach/

Robinson, J. (1964). *Economic philosophy*. Penguin.

Rolls, E. T. (2007). *Emotion Explained* (1st edition). Oxford University Press.

Rosset, J. (2016). Economically based inequalities in political representation: Where do they come from? In A. Machin & N. Stehr (Eds.), *Understanding Inequality: Social Costs and Benefits* (pp. 241–256). Springer Fachmedien Wiesbaden. http://link.springer.com/10.1007/978-3-658-11663-7_12

Rossi, M., & Tappolet, C. (2022). Well-Being as Fitting Happiness. In R. Cosker-Rowland & C. Howard (Eds.), *Fittingness: Essays in the Philosophy of Normativity* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780192895882.003.0012

Roster, C., Albaum, G., & Rogers, R. (2006). Can Cross-National/Cultural Studies Presume Etic Equivalency in Respondents' Use of Extreme Categories of Likert Rating Scales? *International Journal of Market Research*, *48*(6), 741–759. https://doi.org/10.1177/147078530604800608

Rowbottom, J. (2010). *Democracy distorted: Wealth, influence and democratic politics*. Cambridge University Press.

Ryff, C. D. (2014). Psychological well-being revisited: Advances in the science and practice of eudaimonia. *Psychotherapy and Psychosomatics*, *83*(1), 10–28. https://doi.org/10.1159/000353263

Saez, E. (1999). *Do Taxpayers Bunch at Kink Points?* (Working Paper 7366). National Bureau of Economic Research. https://doi.org/10.3386/w7366

Safdar, S., Friedlmeier, W., Matsumoto, D., Yoo, S. H., Kwantes, C. T., Kakai, H., & Shigemasu, E. (2009). Variations of emotional display rules within and across cultures: A comparison between Canada, USA, and Japan. *Canadian Journal of Behavioural Science / Revue*

Canadienne Des Sciences Du Comportement, *41*(1), 1–10. https://doi.org/10.1037/a0014387

Salkind, N. (2010). Content Validity. In *Encyclopedia of Research Design* (pp. 239–243). SAGE Publications, Inc. https://doi.org/10.4135/9781412961288

Samuelson, P. A. (1947). *Foundations of Economic Analysis*. Harvard University Press.

Sanandaji, T., & Wallace, B. (2010). *Fiscal Illusion and Fiscal Obfuscation: An Empirical Study of Tax Perception in Sweden* (SSRN Scholarly Paper ID 1619268). Social Science Research Network. https://papers.ssrn.com/abstract=1619268

Scarantino, A. (2009). On the Role of Values in Economic Science: Robbins and His Critics. *Journal of the History of Economic Thought*, *31*(04), 449. https://doi.org/10.1017/S1053837209990253

Schlozman, K. L., Verba, S., & Brady, H. E. (2013). *The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy*. Princeton University Press.

Scitovsky, T. (1941a). A Note on Welfare Propositions in Economics. *The Review of Economic Studies*, *9*(1), 77–88. https://doi.org/10.2307/2967640

Scitovsky, T. (1941b). A Note on Welfare Propositions in Economics. *The Review of Economic Studies*, *9*(1), 77–88. https://doi.org/10.2307/2967640

Scitovsky, T. (1951). The State of Welfare Economics. *The American Economic Review*, *41*(3), 303–315.

Sen, A. (2011). The Informational Basis of Social Choice. In *Handbook of Social Choice and Welfare* (Vol. 2, pp. 29–46). Elsevier. http://www.sciencedirect.com/science/article/pii/S0169721810000146

Sen, A. K. (1970). Interpersonal Aggregation and Partial Comparability. *Econometrica*, *38*(3), 393–409.

Sen, A. K. (1973). On Ignorance and Equal Distribution. *The American Economic Review*, *63*(5), 1022–1024.

Sen, A. K. (1979a). *Equality of What ?* Stanford University: Tanner Lectures on Human Values.

Sen, A. K. (1979b). The Welfare Basis of Real Income Comparisons: A Survey. *Journal of Economic Literature*, *17*(1), 1–45.

Sen, A. K. (1982). Ethical Measures of Inequality. In A. K. Sen (Ed.), *Choice, Welfare and Measurement*. Basil Blackwell.

Sen, A. K. (1985). *Commodities and capabilities* (Vol. 7). North-Holland.

Sen, A. K. (1987). *On Ethics and Economics*. Blackwell.

Sheshinski, E. (1972). Relation between a social welfare function and the gini index of income inequality. *Journal of Economic Theory*, *4*(1), 98–100. https://doi.org/10.1016/0022-0531(72)90167-6

Shoemaker, S. (1982). The Inverted Spectrum. *Journal of Philosophy*, *79*(July), 357–381. https://doi.org/10.2307/2026213

Sidgwick, H. (1874). *The methods of ethics*. Macmillan.

Silberberg, E. (1972). Duality and the Many Consumer's Surpluses. *The American Economic Review*, *62*(5), 942–952.

Simonson, I., Drolet, A., & article.], [Dawn Iacobucci served as editor and Barbara Kahn served as associate editor for this. (2004). Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research*, *31*(3), 681–690. https://doi.org/10.1086/425103

Sinden, A. (2019). The Problem of Unquantified Benefits. *Environmental Law*, *49*(1), 73–129.

Singh, R., & Alexandrova, A. (2020). Happiness economics as technocracy. *Behavioural Public Policy*, *4*(2), 236–244. https://doi.org/10.1017/bpp.2019.46

Sippel, R. (1997). An Experiment on the Pure Theory of Consumer's Behaviour*. *The Economic Journal*, *107*(444), 1431–1444. https://doi.org/10.1111/j.1468-0297.1997.tb00056.x

Skinner, Q. (1973). The Empirical Theorists of Democracy and Their Critics: A Plague on Both Their Houses. *Political Theory*, *1*(3), 287–306.

Smart, J. J. C. (1959). Sensations and Brain Processes. *The Philosophical Review*, *68*(2), 141–156. https://doi.org/10.2307/2182164

Søgaard, R., Lindholt, J., & Gyrd-Hansen, D. (2012). Insensitivity to Scope in Contingent Valuation Studies. *Applied Health Economics and Health Policy*, *10*(6), 397–405. https://doi.org/10.1007/BF03261874

Somanathan, E. (2006). Valuing lives equally: Distributional weights for welfare analysis. *Economics Letters*, *90*(1), 122–125. https://doi.org/10.1016/j.econlet.2005.07.015

Soulé, M. E. (1985). What is Conservation Biology?A new synthetic discipline addresses the dynamics and problems of perturbed species, communities, and ecosystems. *BioScience*, *35*(11), 727–734. https://doi.org/10.2307/1310054

Stauffer, W. R., Lak, A., & Schultz, W. (2014). Dopamine reward prediction error responses reflect marginal utility. *Current Biology: CB*, *24*(21), 2491–2500. https://doi.org/10.1016/j.cub.2014.08.064

Steptoe, A., Dockray, S., & Wardle, J. (2009). Positive Affect and Psychobiological Processes Relevant to Health. *Journal of Personality*, *77*(6), 1747–1776. https://doi.org/10.1111/j.1467-6494.2009.00599.x

Steptoe, A., Gibson, E. L., Hamer, M., & Wardle, J. (2007). Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. *Psychoneuroendocrinology*, *32*(1), 56–64. https://doi.org/10.1016/j.psyneuen.2006.10.001

Steptoe, A., Kunz-Ebrecht, S., Owen, N., Feldman, P. J., Rumley, A., Lowe, G. D. O., & Marmot, M. (2003). Influence of socioeconomic status and job control on plasma fibrinogen responses to acute mental stress. *Psychosomatic Medicine*, *65*(1), 137–144.

Steptoe, A., O'Donnell, K., Badrick, E., Kumari, M., & Marmot, M. (2008). Neuroendocrine and inflammatory factors associated with positive affect in healthy men and women: The

Whitehall II study. *American Journal of Epidemiology*, *167*(1), 96–102. https://doi.org/10.1093/aje/kwm252

Steptoe, A., & Wardle, J. (2005). Positive affect and biological function in everyday life. *Neurobiology of Aging*, *26 Suppl 1*, 108–112. https://doi.org/10.1016/j.neurobiolaging.2005.08.016

Steptoe, A., Wardle, J., & Marmot, M. (2005). Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(18), 6508–6512. https://doi.org/10.1073/pnas.0409174102

Stigler, G. J. (1943). The New Welfare Economics. *The American Economic Review*, *33*(2), 355–359.

Stiglitz, J. E., Sen, A. K., & Fitoussi, J.-P. (2015). *The Measurement of Economic Performance and Social Progress Revisited*. CMEPSP. http://www.stiglitz-sen-fitoussi.fr/documents/overview-eng.pdf

Stone, R. (1954). Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand. *The Economic Journal*, *64*(255), 511–527. https://doi.org/10.2307/2227743

Strathern, M. (1997). ‘Improving ratings’: Audit in the British University system. *European Review*, *5*(3), 305–321. https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4

Streeten, P. (1953). Appendix: Recent Controversies. In G. Myrdal (Ed.), *The Political Element in the Development of Economic Theory*. Routledge & Kegan Paul.

Streeten, P. (1981). Values, Facts and the Compensation Principle. In P. Streeten (Ed.), *Development Perspectives* (pp. 35–51). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-05341-4_2

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The Shifting Basis of Life Satisfaction Judgments Across Cultures: Emotions Versus Norms. *Journal of Personality*, *74*(2), 482–493.

Summers, L. H., & Zeckhauser, R. J. (2008). *Policymaking for Posterity* (Working Paper 14359). National Bureau of Economic Research. https://doi.org/10.3386/w14359

Swift, J. (1997). *A Modest Proposal For preventing the Children of Poor People From being a Burthen to Their Parents or Country, and For making them Beneficial to the Publick*. Gutenburg. https://www.gutenberg.org/files/1080/1080-h/1080-h.htm (Original work published 1729)

Sytsma, J., & Livengood, J. (2015). *The Theory and Practice of Experimental Philosophy*. Broadview Press.

Tal, E. (2013). Old and New Problems in Philosophy of Measurement. *Philosophy Compass*, *8*(12), 1159–1173. https://doi.org/10.1111/phc3.12089

Tal, E. (2020). Measurement in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/measurement-science/

Temkin, L. S. (2003). Egalitarianism Defended. *Ethics*, *113*(4), 764–782. https://doi.org/10.1086/373955

Thurow, L. C. (1971). The Income Distribution as a Pure Public Good. *The Quarterly Journal of Economics*, *85*(2), 327–336. https://doi.org/10.2307/1880707

Tintner, G. (1946). A Note on Welfare Economics. *Econometrica*, *14*(1), 69–78. https://doi.org/10.2307/1905704

Tipoe, E., Adams, A., & Crawford, I. (2022). Revealed preference analysis and bounded rationality
[Consume now or later? Time inconsistency, collective choice and revealed preference]. *Oxford Economic Papers*, *74*(2), 313–332.

Torija, P. (2013). *Do Politicians Serve the One Percent? Evidence in OECD Countries* (CITYPERC Working Paper Series 2013–04). Department of International Politics, City University London. https://ideas.repec.org/p/dip/dpaper/2013-04.html

Tov, W. (2009). Culture and subjective well-being. In E. Diener (Ed.), *Culture and well-being: The collected works of Ed Diener* (pp. 9–41). Springer Science + Business Media. https://doi.org/10.1007/978-90-481-2352-0_2

Tov, W., & Nai, Z. L. S. (2018). Cultural differences in subjective well-being: How and why. In *Subjective well-being and life satisfaction* (pp. 50–73). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781351231879-3

Uchida, Y., Norasakkunkit, V., & Kitayama, S. (2004). Cultural constructions of happiness: Theory and emprical evidence. *Journal of Happiness Studies*, *5*(3), 223–239. https://doi.org/10.1007/s10902-004-8785-9

van Praag, B. M. S. (1991). Ordinal and cardinal utility. *Journal of Econometrics*, *50*(1), 69–89. https://doi.org/10.1016/0304-4076(91)90090-Z

van Praag, B. M. S., & Frijters, P. (1999). The measurement of welfare and well-being The Leyden approach. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 413–433). Russell Sage Foundation.

Veenstra, G. (2011). Race, gender, class, and sexual orientation: Intersecting axes of inequality and self-rated health in Canada. *International Journal for Equity in Health*, *10*(1), 3. https://doi.org/10.1186/1475-9276-10-3

Vendrik, M. C. M., & Woltjer, G. B. (2007). Happiness and loss aversion: Is utility concave or convex in relative income? *Journal of Public Economics*, *91*(7), 1423–1448. https://doi.org/10.1016/j.jpubeco.2007.02.008

Vickrey, W. (1960). Utility, Strategy, and Social Decision Rules. *The Quarterly Journal of Economics*, *74*(4), 507–535.

Volokh, A. (2011). Rationality or Rationalism: The Positive and Normative Flaws of Cost-Benefit Analysis. *Houston Law Review*, *48*(1), 79–100.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Waldner, I. (1972). The Empirical Meaningfulness of Interpersonal Utility Comparisons. *The Journal of Philosophy*, *69*(4), 87–103. https://doi.org/10.2307/2024617

Walen, A. (2015). Proof Beyond a Reasonable Doubt: A Balanced Retributive Account. *Louisiana Law Review*, *76*, 355–446.

Waterman, A. S., Schwartz, S. J., Zamboanga, B. L., Ravert, R. D., Williams, M. K., Agocha, V. B., Kim, S. Y., & Donnellan, M. B. (2010). The Questionnaire for Eudaimonic Well-Being: Psychometric properties, demographic comparisons, and evidence of validity. *The Journal of Positive Psychology*, *5*(1), 41–61. https://doi.org/10.1080/17439760903435208

Weber, M. (1949). The meaning of “ethical neutrality” in Sociology and Economics. In M. Weber, E. Shils, & H. A. Finch (Eds.), *The methodology of the social sciences*. The Free Press.

Weber, M. (1949). *The methodology of the social sciences* (E. Shils & H. A. Finch, Trans.). The Free Press. (Original work published 1922)

Weintraub, R. (1996). The Impossibility of Interpersonal Utility Comparisons: A Critical Note. *Mind*, *105*(420), 661–665.

Weinzierl, M. (2017). *A Welfarist Role for Nonwelfarist Rules: An example with envy* (23587; NBER Working Papers). National Bureau of Economic Research, Inc. https://ideas.repec.org/p/nbr/nberwo/23587.html

Weisbach, D. A. (2015). Distributionally Weighted Cost–Benefit Analysis: Welfare Economics Meets Organizational Design. *Journal of Legal Analysis*, *7*(1), 151–182. https://doi.org/10.1093/jla/lau009

Weitzman, M. L. (2007). A Review of the *Stern Review on the Economics of Climate Change*. *Journal of Economic Literature*, *45*(3), 703–724. https://doi.org/10.1257/jel.45.3.703

Weston, S. C. (1994). Toward a Better Understanding of the Positive/Normative Distinction in Economics. *Economics and Philosophy*, *10*(01), 1–17. https://doi.org/10.1017/S0266267100001681

Whitehead, J. C. (2016). Plausible responsiveness to scope in contingent valuation. *Ecological Economics*, *128*, 17–22. https://doi.org/10.1016/j.ecolecon.2016.03.011

Wilkinson, R. G. (1996). *Unhealthy Societies: The afflictions of inequality*. Routledge.

Wilkinson, R. G., & Pickett, K. (2017a). Inequality and mental illness. *The Lancet Psychiatry*, *4*(7), 512–513. https://doi.org/10.1016/S2215-0366(17)30206-7

Wilkinson, R. G., & Pickett, K. E. (2006). Income inequality and population health: A review and explanation of the evidence. *Social Science & Medicine*, *62*(7), 1768–1784. https://doi.org/doi: 10.1016/j.socscimed.2005.08.036

Wilkinson, R. G., & Pickett, K. E. (2017b). The enemy between us: The psychological and social costs of inequality. *European Journal of Social Psychology*, *47*(1), 11–24. https://doi.org/10.1002/ejsp.2275

Wodak, D. (2019). What If Well-Being Measurements Are Non-Linear? *Australasian Journal of Philosophy*, *97*(1), 29–45. https://doi.org/10.1080/00048402.2018.1454483

Yang, Y. C., Li, T., & Frenk, S. M. (2014). Social network ties and inflammation in U.S. adults with cancer. *Biodemography and Social Biology*, *60*(1), 21–37. https://doi.org/10.1080/19485565.2014.899452

Yechiam, E. (2019). Acceptable losses: The debatable origins of loss aversion. *Psychological Research*, *83*(7), 1327–1339. https://doi.org/10.1007/s00426-018-1013-8

Yew-Kwang Ng. (1989). What Should We Do About Future Generations?: Impossibility of Parfit’s Theory X. *Economics & Philosophy*, *5*(2), 235–253. https://doi.org/10.1017/S0266267100002406

Zerbe, R. O., & Scott, T. (2014, July 14). *(Almost) Everybody Wins: A True Pareto Justification for Practical Welfare Economics and Benefit-Cost Analysis*. Social Science Research Network. http://papers.ssrn.com/abstract=2466101

Appendix A: An ordinal, Kaldor-Hicks approach to subjective wellbeing

Let us imagine a happiness scale. To emphasize that it is ordinal (following Wodak 2019) we will use letters rather than numbers. The scale goes A, B, C, D, E, with letters further in the alphabet being better.

Mary is a D, we estimate that a policy will move her down to a C. Johnathan is also a D, but we estimate the same policy will move him to an A.

Now we estimate that it will cost $500 dollars to move Mary back up to a D again, after introducing this policy. We also estimate that Johnathan could afford to lose $500 dollars again without going back to a D.

Hence the policy is a *subjective-wellbeing Kaldor-Hicks Improvement (SWBKHI).*

There are going to be some difficulties with this approach. With normal Kaldor-Hicks we only require one estimate for each person, how much they would be willing to pay to implement the policy, or how much they would be willing to pay to avoid the policy. With this approach, we need two distinct estimates- the effect the policy will have on subjective-wellbeing, and the amount of money it will cost to reverse that effect. Knowing one does not imply knowledge of the other.

Another reason it's not going to work so well- if we want it to be a *truly* ordinal approach, without any implicit interpersonal comparison, we cannot just work out how much money it would take to get a representative person affected by the policy back to their previous happiness level. We have to work it out for every single person affected by the policy. There will be the odd individual who is profoundly affected by the policy, for whom it would cost vast sums of money to get back to their prior state. Perhaps no finite amount of money would be enough in some circumstances.

A closely related problem: People's veto power in normal Kaldor-Hicks is limited by their money. If I only have 50,000, the maximum I can be willing to pay to stop a policy is 50,000. But there's no limit on how much money someone might need to get them back to their prior psychological state. Of course, the advantage here is the greater equity with which the population will be treated. This will break down in certain cases- the cost of compensating a billionaire for his unhappiness in losing his favourite golf course might be obscene, although the same is true of standard Kaldor-Hicks tests.

And of course, many of the normal criticisms of Kaldor-Hicks apply- e.g. the impossibility of costless lump sum transfers- though some will be removed or alleviated.