

# THE CONSTRUCTION OF WELL-BEING

MARY MITCHELL

Submitted for the degree of Doctor of Philosophy

University College London (UCL)



I, Mary Mitchell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



## Abstract

My thesis develops an alternative to orthodox theories of well-being. I argue that well-being is not a property of people or the world that exists separately from attempts to define and measure it. Instead, assessments of well-being are largely shaped by the purposes and interests of the people making the measurement.

In part I, I argue that philosophy of well-being should take at face value the variety of ways that well-being is understood and measured. This points towards a pluralist account of well-being. I go on to argue that a theory of well-being is incomplete unless it says something about how to determine the extent to which life is going well for someone. I argue that the identification of well-being in individuals amounts to a form of measurement. Theories of well-being must therefore be theories of the measurement of well-being.

In part II, I look closely at three approaches to defining and measuring well-being—in terms of objective goods, preferences, and subjective experience. I argue that, in each case, appraisals of whether someone is doing well or not, and how well they are doing, depend on the context of measurement, the tools used to measure well-being, and the goals and purposes of the people making the assessment.

In part III, I propose that well-being should not be treated as a property of people or of the world, but rather something which is largely constructed in the process of measurement. I draw on contemporary model theories of measurement to argue that well-being is best understood as representing a relation between the person whose

well-being is being measured, a measuring instrument, and the environment. This relation is modelled by the people who are making the measurement in order to produce information about well-being and ascriptions of well-being are therefore unavoidably attitude dependent.

## Impact statement

As of April 2018, UCL requires an impact statement to be included in all PhD theses, which should describe how the expertise, knowledge, analysis, discovery or insight presented in the thesis could be put to beneficial use, both within and outside of academia.

My thesis offers an account of the nature of well-being which diverges appreciably from orthodox—and even unorthodox—theories of well-being within academic philosophy. It adds, I hope, something quite distinctive to the discipline, building on and responding to recent work from other theorists. My characterisation of well-being attempts to draw together philosophical ideas about prudential value, meta-ethics, and measurement with ideas in economics, psychology and psychometrics, and consequently reaches out beyond the confines of my immediate discipline. My suggestion that philosophers of well-being cannot avoid thinking about its measurement actively supports dismantling some of the boundaries between philosophy and other disciplines. Inside academia, impact would be most readily achieved via the publication of selected elements of the thesis in peer-reviewed journals.

Outside of academia, my view has potential implications for public policy design, particularly in those areas where the measurement of well-being is already a mainstay of decision making, such as healthcare and international development contexts, but also in those areas where the measurement of well-being is emerging, such as in the workplace. My thesis offers justification for the practice of measuring well-being in a

diversity of ways, but also places the onus on those who are measuring well-being to scrutinise their own purposes and circumstances to determine how well-being ought to be theorised and measured. It thus cautions against a 'one-size-fits-all' approach to the measurement of well-being, or at least demands solid justification for using the same measure across different contexts. Impact in these areas would be best achieved via engagement with public policy makers.



# Contents

ACKNOWLEDGEMENTS	11
INTRODUCTION	13

## Part I

1 THE CASE AGAINST WELL-BEING INVARIANTISM	29
1.1 The philosophical orthodoxy	32
1.2 Well-being invariantism and well-being variantism	41
1.3 Defending invariantism I: Multiple bearers of well-being	47
1.4 Defending invariantism II: Circumscribing well-being	57
1.5 Stance independence and stance dependence	64
1.6 Invariantism, variantism, and the philosophical schism	73
2 WELL-BEING AND MEASUREMENT	80
2.1 The scope of theories of well-being	82
2.2 Well-being and measurement	86
2.3 What is measurement?	96
2.4 Realism, instrumentalism, and well-being	105
2.5 Why measurement?	112

## Part II

INTRODUCTION TO PART II	119
3 OBJECTIVE GOODS	123
3.1 Objective list theories	124
3.2 Prudential goods and well-being	128

3.3	Identifying objective goods I: Individuating the goods	134
3.4	Identifying objective goods II: Specifying the goods	140
3.5	Specification and attitude independence	147
3.6	The attitude dependence of ‘objective’ goods	154
4	PREFERENCES	161
4.1	Desire- and preference-satisfaction theories of well-being	162
4.2	Utility theory	167
4.3	Measuring health state utility	171
4.4	Systematic bias and framing effects	175
4.5	Eliminating irrational preferences	181
5	SUBJECTIVE EXPERIENCE	192
5.1	Experienced utility	195
5.2	Divergent experienced utility values	203
5.3	Temporal monotonicity and remembered utility	207
5.4	Distinguishing remembered from moment utility	212
5.5	Framing and focussing effects in experienced utility measures	216
5.6	The varieties of experienced utility	223
<b>Part III</b>		
6	THE CONSTRUCTION OF WELL-BEING	229
6.1	Well-being represents a relation	230
6.2	The role of the measuring subject	237
6.3	Stance independent pluralism I: McDowell	245
6.4	Stance independent pluralism II: Cornell realism	252
6.5	The extent and limits of well-being	261
	CONCLUSION	274
	BIBLIOGRAPHY	280

## Acknowledgements

I joined the UCL Department of Philosophy, with James Wilson as my primary supervisor, on the recommendation of Albert Weale. While James and I had met once during the application process, and unearthed common interests, our very limited knowledge of one another meant that we both took something of a gamble. The extent of our shared philosophical concerns slowly dawned upon me over my first year at UCL. James surprised me by matching what I had taken to be my radical pragmatist and constructivist tendencies, and encouraged me in divergent interdisciplinary adventures. He continues to amaze me with the enormous breath of his philosophical interests and knowledge. I could not have asked for a better supervisor. He has been unfailingly generous with his time and unerringly instructive with his guidance and direction. Above anyone, I owe the completion of this thesis to him.

I am always privately astonished when people read my written work and actually understand what I'm trying to say. Albert has the ability not only to comprehend what I've put down on the paper, but to grasp what I was thinking when I put it there. This stems not, I think, from some latent telepathic power, but from his immense philosophical generosity. Albert does not—and I think probably cannot—read philosophy without making a concerted effort to fully appreciate and charitably reconstruct the intellectual motives of the author. His synoptic abilities are, as a result, pretty well unmatched. He has a formidable comprehension of the relationship between ethics, economics, and public policy. As my second supervisor, Albert has

given me invaluable counsel, particularly in the final year of this project, and has enabled me to turn a collection of thoughts into an argument.

Under the supervision of John Vorhaus I wrote an ambitious chapter, which ultimately did not make it into the final manuscript. Nonetheless, John's guidance and conversation throughout my time at UCL has been hugely influential on me. He has shown me an alternative way of doing philosophy—thoughtful, critical, humble, attentive.

Many members of the UCL philosophy community have suffered through my half-baked drafts and ideas, and provided generous commentary. In particular I have benefitted from the insight and friendship of Showkat Ali, Julian Bacharach, Vanessa Carr, Catherine Dale, Pete Falconbridge, Ben Fardell, Mog Hampson, Charles Jansen, Andrew Knox, Edgar Phillips and Laura Silva, who have all helped me to weather the challenges of writing a PhD. Luke Fenton-Glynn provided incisive comments on chapter 2, and helped me to get my head around measurement theory.

Sam Rose has been an invaluable friend, conversation partner and sounding board throughout my philosophical development. Finally, my partner Peter Shenai and my sister Lizzie Mitchell have given me nothing but love, encouragement and stability in return for four years of hand-wringing and volatility. No amount of thanks can do justice to the support they have given me.

This thesis was researched and written with the support of an Arts and Humanities Research Council Studentship, awarded by the London Arts and Humanities Partnership.

## Introduction

Well-being is a much debated and much disputed concept in philosophical theories of value. Theories of well-being tell us, in some sense, what makes something good or bad for a person or what makes a life, or part of a life, go well for someone who is living it. Getting the right account of well-being is important because promoting or protecting people's well-being is typically understood to be part of the good, or part of what it is to do the right thing. Though few theories of value take well-being, or *prudential value*, to be exhaustive of value, it is widely taken to be partly constitutive thereof. A poorly justified or incorrect theory of well-being will likely leave us with an impoverished or incomplete theory of value.

But well-being is of practical, not just theoretical, concern. Increasingly measures of well-being and quality of life are used as measures of positive outcome in decision making, notably in public policy contexts, but also by individuals and commercial organisations. Well-being is used in both the design of new public policies and the evaluation of enacted policies, as a means of predicting consequences and assessing impact. In a public context, getting the right account of well-being is crucial not only in order to meet a philosophical standard of truth or coherence—although this may also be necessary—but also to meet a standard of accountability. If it is alleged that public policy decisions are based on well-being, or assessed on the basis of well-being, but in fact use tools which measure something else, or which measure well-being inaccurately, they will fail to be adequately justified and may therefore fail to secure

legitimacy. Furthermore, the concept of well-being used in public decision making will have widespread effects. Given its use in decision making across the public policy spectrum—in health policy, international development, benefits and pensions, environmental policy—the concept, or concepts, of well-being used will impact on the course, length and quality of people’s lives. If the wrong concept of well-being is used for such decisions, the course and content of people’s lives will be unjustifiably affected. Securing and justifying the right concept of well-being is not, then, simply a matter of intellectual achievement, it is also a pressing social and political concern. Developing a clear and convincing account of well-being is a central part of such a justification.

Part of the reason that well-being is a philosophically interesting and perplexing concept is, I think, the position it occupies in relation to fact and value. Assessments of well-being make and depend upon factual claims about the kinds of beings that we are, and the particular properties, resources, characteristics and attitudes that we have. Well-being can be applied to people only in virtue of some set of features that they have. To assert that someone is doing well, or that one person is doing better than another, but to be able to point to nothing about them or their context which justifies this claim, is to fail to say anything meaningful about well-being. However, assessments of well-being also make and depend upon normative claims about what is *good* for us, about what properties, resources, characteristics and attitudes we *ought* to have. Well-being is, in this sense, a thick concept: a concept which has both descriptive and evaluative elements (Roberts 2013). That well-being is a thick concept need not imply that the descriptive and evaluative content of well-being claims can be reductively analysed into separate components, nor need it imply that the descriptive elements of well-being claims are fully non-evaluative. Friendship, achievement, knowledge, happiness and other complex, evaluative goods might be identified as things that contribute to people’s well-being. Or there might be no complete

specification of the things that contribute to or constitute well-being, despite it being possible to point to some features that play this role in particular contexts.

Some theories of well-being do seek to specify non-evaluative prudential properties—this is exhibited, for example, by those perfectionist theories which attempt give a fully naturalist explanation of prudential value, grounding it in facts about human natural history and biological capacity (Foot 2001). Such an analysis is also seen in certain hedonist theories of well-being, particularly those that take pleasure to be a feeling rather than an attitude (Crisp 2006), and theories which equate well-being with the satisfaction of preferences, where preferences are understood in wholly behavioural terms (Sugden 2006). Other theories analyse well-being in terms which are less easily described as non-evaluative, or which are straightforwardly evaluative—in terms of desires (Heathwood 2005; Bruckner 2016), complex goods such as knowledge, friendship, rational activity (Fletcher 2015), and in terms of value and reflective endorsement (Tiberius 2011; Brandt 1979). But even in the latter cases, providing evidence of these goods—even if it is not free from evaluation—is crucial to justifying claims of well-being and distinguishing actual well-being from apparent well-being. And in the case of those theories which seek to provide a non-evaluative definition of well-being, the identification of a set of properties as constitutive of well-being seems to require some assessment of them as valuable. The philosophy of well-being engages, then, in the manifestation and realisation of normative concepts in the world.

My own view seeks to explore and problematise the fact-value distinction in the context of well-being, and to ultimately reject the possibility of a non-evaluative account of prudential value. In this thesis, I develop an account of well-being which presents an alternative to orthodox philosophical theories of well-being and prudential value: I argue that well-being is not a property of people or the world that exists separately from attempts to define and measure it. Instead, assessments of well-being are largely shaped by the purposes and interests of those people who measure it. There

is an element of ‘wait and see’ to the argument that I present, and my positive view is not fully developed until the final chapter. Such argumentative approaches work best when readers have some knowledge of the direction of travel from the outset, and thus are able to follow and understand the relevance of the narrative progression. Surprise is not, on the whole, a philosophical virtue. With this in mind I begin with an outline of the project, which calls attention to some of my underlying methodological and philosophical commitments.

My argument ultimately stems from a dissatisfaction with the way that contemporary philosophers of well-being approach their subject. There are three common features of theories of well-being that strike me as misguided, or at least as limited in their efficacy. First, theories of well-being tend to operate with a globalist conception of well-being, that is, well-being is understood to be a global assessment of a person’s life. This entails that context-specific assessments of well-being are either not assessments of well-being at all, or merely partial contributors to well-being. Secondly, theories of well-being tend to be reductive, offering an account of an essence which characterises all true instances of well-being, and so providing a single, clear definition of well-being which applies across all contexts. Thirdly, theories of well-being often use intuitive and *a priori* reasoning in the development and justification of their respective concepts of well-being. Intuitions about particular thought experiments are used as the basis for general claims about the definition of well-being. And the specification of a general definition of well-being is taken to regulate the use of the concept, and to police the appropriateness of well-being claims. This often manifests itself in the development of a broad and vague concept of well-being and its constituents, with little or no discussion of the manifestation and application of these concepts in people and in the world. While not all contemporary theories of well-being demonstrate all of these features, and not necessarily explicitly or to the same degree, there is a broad tendency to think in these terms. Moreover, it is not only philosophers who exhibit these



tendencies; as I argue in Part II, some economists and psychologists have been led to similar claims, despite being embedded in particular practices of well-being measurement.

The primary motivation for moving away from such a picture is its failure to make sense of the way that the concept of well-being is invoked, and the content of well-being claims, outside of philosophy. Across other disciplines—in the clinical, social and political sciences, in development economics and in microeconomics, for example—and in everyday conversation and personal decision making, well-being is defined and operationalised in a variety of ways, for a proliferation of different purposes. These are often quite specific, and are not suitable outside of a fairly restricted context of application, nor are they necessarily intended to be more widely applicable. Not only does an investigation of well-being in practice point towards a non-global account of well-being, it also casts doubt on the value of a reductive approach to well-being. The concepts of well-being used in different contexts appear to exhibit variety not only in their scope, but also in their theoretical content. Such variety points to a conceptual pluralism about well-being, rather than a single, exhaustive definition. Looking to the use of the concept of well-being to inform philosophical discussion itself constitutes an alternative approach to thinking about well-being, in contrast to *a priori* and intuition-based approaches. It holds the observation of practice to be an important way of informing our philosophical and normative concepts, and foregrounds inductive reasoning. I do not have a conclusive reason for favouring such an approach—though I argue that the alternative requires an improbable error theory—so perhaps this is best understood as a central methodological commitment of my view. That is, I explore how an account of well-being would look if it took at face value the variation in its definition and conception in practice, where taking such variation at face value just seems like a sensible thing to do or, at very least, an alternative worth exploring.

My concern with context-specific definitions and applications of well-being leads me to focus on its measurement, something which is side-lined or altogether left out of many philosophical accounts of well-being, but is typically central to the operationalisation of well-being. I suggest that theories of well-being need to say something about the identification of well-being, that is, they need to answer the question “how do I determine the extent to which life is going well for someone?” I also argue for an inclusive view of measurement, such that nominal and ordinal scales count as measurement scales, as a result of which even minimal identification of the presence and extent of well-being counts as a form of measurement. These two arguments lead me to conclude that any theory of well-being must also be a theory of the measurement of well-being. Conceiving of well-being in terms of measurement helps to focus on an aspect of well-being that I think is overlooked in much philosophical discussion: measurement draws attention to the ontological aspects of conceptual philosophy of well-being, that is, it reminds us that high-level conceptual definition and analysis must be matched with scrutiny of the realisation or manifestation of the concept in the world. It is all very well insisting that well-being is pleasure, but what has to be the case for someone to have pleasure, and what has to be the case for them to have more or less of it, and how can I determine whether they have it or not, and how much?

Taking the variation in the definition and conception of well-being in practice at face value and conceiving of well-being in terms of measurement still leaves substantive options on the table for an overall account of well-being. I cast the relevant options as *stance independent pluralism* about well-being, on the one hand, and *stance dependent pluralism* about well-being, on the other. The crucial distinction here is whether the appropriate characterisation of well-being in particular cases is dependent on people’s attitudes or not. Note that this is not a dispute about the content of well-being claims—that is, whether well-being is assessed via self-report or the measurement of

psychological phenomena, or whether it is assessed via the measurement of external goods. Rather it is a disagreement about whether the things that constitute well-being in particular cases are good for us independently of whether we think that they are, or not.

In order to explore the plausibility of these two options, and ultimately to offer a defence of the stance dependent pluralist view, the central part of the thesis considers, in some detail, three attempts to operationalise well-being: in terms of objective goods, preferences and subjective experience. In line with my broadly inductive approach to thinking about well-being, I consider these chapters to be case studies which provide evidence for a stance dependent pluralism about well-being. I do not seek to criticise any of these three accounts as theories of well-being. Instead I argue that, in each case, all is not as it seems, or rather, all is not as it is presented to be: when well-being or a constitutive of well-being is identified and measured, what is identified and measured is not a property or state which exists prior to the act of identifying and measuring it. Well-being is constructed through its measurement, and so comes into existence as and when it is defined and identified in particular contexts and for particular purposes. These three chapters are fairly divergent with respect to the literature upon which they draw and the particular arguments that I pursue, but I endeavour to demonstrate structural similarities across the three cases.

All this forms the groundwork for the more general account of well-being that I defend in the final chapter. I draw on contemporary measurement theory to give some substance to the idea that well-being is constructed through its measurement. There are two aspects to the measurement of well-being. First, there is a relation between an object of measurement (the person whose well-being is being measured), a measuring instrument (the survey, preference elicitation scenario, electroencephalogram, or index specifying prudential goods, for example), and the environment within which well-being is assessed. This relation produces certain measurement indications: Felix

scoring 4 in question about depression, Aaron having a strong preference for preserving his physical function, Melanie exhibiting high baseline asymmetry in her anterior cortical function, or Eloise having access to high quality healthcare facilities. But these indicators don't alone provide information about well-being. So, secondly, in order to produce assessments of well-being, the interaction between the relata, and the resulting measurement indications, are abstracted or modelled to produce claims about well-being, or measurement outcomes. This abstraction will draw on theoretical and statistical assumptions about the relationship between measurement indications and well-being—the relevance of depression scores or EEG readings to well-being, for example—and will reflect the purposes of the person or people making the measurement. Theorising about well-being in this way involves placing the object of measurement in a parameter space, which represents a range of possible states that it can be in or characteristics that it can have.

This might sound quite abstract, but the bottom line is this: the measuring subject—the person who is measuring well-being—has a particularly important role to play in producing assessments of well-being. They not only have a role to play in choosing or designing the measuring instrument, determining the measuring environment, and selecting the appropriate object of measurement, but their purposes and theoretical framework are also crucial in shaping the possible claims that can be made about well-being. Some aspects of this theoretical framework will be within the control of the measuring subject, and other aspects will be fixed by disciplinary, social and linguistic convention. But the theoretical framework is largely attitude dependent, and to the extent that it *is* dependent on features of the object of measurement, it is dependent on theorised accounts of those features. The implication of all of this is that there is no definite account of how much well-being someone has, of how well they are doing, outside of particular contexts in which it is defined, identified and measured. Not only is well-being non-reductive in that there isn't a single way of characterising well-being

across all well-being claims, but well-being is non-reductive in that there isn't a single thing that well-being refers to in particular cases. Well-being is multiply realisable, and assessments of well-being are constructed in and through their measurement. Capturing the concept of well-being in the language of measurement helps to make sense of the attitude dependence of well-being claims, while not losing sight of the centrality of the object of measurement and their features. There is no objective way of characterising a person in terms of their well-being, whether in general or in particular contexts, and they have to be modelled in a well-being-relative fashion prior to any such characterisation.

The view that I present amounts to a fairly stark contrast to much of the characterisation of well-being in contemporary philosophy. While I defend my account against anticipated objections, there are aspects of my approach that are likely to be outright rejected by some of the well-being theorists whose discussion motivates my view. Most manifest is my claim that normative concepts should be informed and policed by linguistic practice, rather than employing a top down analytic method which takes concepts to have a definitional structure. I take an account of practical differences to be essential for settling metaphysical, and in this case normative and metaethical, disputes. This is itself controversial. Given this, my account does not stand as a straightforward criticism of orthodox philosophy of well-being, though it is contrary to it. Instead, I present an alternative explanation, a different account to occupy the same conceptual space. My view is *variantist* and *pluralist*, insofar as it takes the appropriate theory and concept of well-being to be context-dependent. It is *stance dependent*, insofar as it takes the appropriate theory and concept of well-being for a given context to be dependent on the attitudes and purposes of those people who are seeking to measure well-being, rather than being an objective facts about the world. And it is *pragmatist*, insofar as it takes theories and concepts of well-being to be better understood to be justified and expedient than true. While parts of the project lead me

deep into metaethical and occasionally metaphysical discussion about the nature of well-being, this is firmly in pursuit of practical ends. I endeavour to develop an account of well-being which accommodates the variety of ways in which well-being is conceptualised and measured, rather than attempting to explain them away, and which provides some tools for critically evaluating the appropriateness of particular measures of well-being.

To conclude these introductory remarks, a brief note about terminology. Philosophers use varied language to talk about well-being. Many speak explicitly about well-being, others about welfare, what is good for us, what makes life good for someone, what counts as a good life, what benefits people, what makes someone's life go well, and what it takes for someone to be doing well. Each of these formulations has slightly different connotations, and each appears to be correlated with different theories of well-being. Perfectionist and objectivist theories are more likely to talk in terms of the good: what is good or bad for someone, what makes a good life (Foot 2001, Ch.6; Kraut 2007; Rice 2013). Hedonist theories, which I take to be a variety of objectivist theory—I discuss this further in chapter 3—also tend to think of well-being in terms of good lives or what is good for people (Bradley 2009; Bramble 2016; Crisp 2006). Economists and psychologists, but also some philosophers, speak of utility (Broome 1991; Griffin 1988; Kahneman 1999; Kahneman and Tversky 1979). Desire based accounts often speak in terms of what makes one's life go well (Bruckner 2016; Heathwood 2005, 2006; Parfit 1984) or sometimes what it takes for someone to be doing well (Bruckner 2016; Heathwood 2005). Shelley Kagan (1994) makes a distinction between well-being and how well a person's life is going—the former being dependent on facts internal to a person, that is, facts about how well he thinks he is doing, and the latter being dependent on facts external to the person, but internal to his life—but he is unusual in making this distinction.

I use all of these formulations, as well as talking directly about well-being. It is no coincidence that I use ‘good for’ more frequently in relation to objectivist theories, and ‘going well’ and ‘doing well’ more frequently in relation to desire and preference-based theories. The variety of intelligible ways of speaking about well-being is significant for understanding the concept, and it seems to me that these different linguistic formulations convey conceptual differences. A claim about the good life implies an attitude independent conception of prudential goodness far more than a claim about what it means for life to go well, or for someone to be doing well. My non-prescriptive attitude therefore partly reflects a desire not to load the die in favour of any particular conception of well-being by either avoiding or insisting on certain ways of talking. Furthermore, direct talk about well-being—what enhances well-being, what contributes to well-being, how much well-being someone has—might sometimes be thought to imply that well-being is a specifiable, quantifiable or aggregative property. While less determinate talk about good lives and lives that go well often grounds theories that *do* characterise well-being in terms of particular properties or states (pleasure, satisfaction of desires, fulfilment of characteristically human traits), many well-being theorists avoid language which explicitly implies this. Insisting on translating well-being theorists’ claims into direct claims about contributions to well-being or quantities of well-being might, then, presume in favour of my own critique, so I avoid this where I can. Ultimately, though, I take the large linguistic variation in the philosophy of well-being to broadly support my pluralist conception, and to be itself an indicator of the conceptual variety that well-being manifests.

Finally, I use two concepts throughout the thesis that are closely related—attitude (in)dependence and stance (in)dependence. It is worth mentioning the ways in which I take them to be different, and the reason that I retain both formulations. The concept of *stance* dependence arises in metaethics about the status and justification of moral claims in general, and represents the idea that the truth of moral principles is

dependent on people's attitudes. Normative principles that are stance dependent are not necessarily dependent on the beliefs and attitudes of particular individuals, though they are dependent on the beliefs and attitudes of some group of people. Stance dependence does not, then, necessarily imply comprehensive subjectivism about prudential value, though it does imply at least intersubjectivism. The concept of *attitude* dependence emerges in the objective list literature in part in response to desire-based accounts of well-being, which take well-being claims to depend on individual attitudes. While I ultimately take attitude independence to imply independence from any attitudes, including those held intersubjectively, attitude dependence about well-being carries implications of subjectivism about prudential value. Stance dependence and attitude dependence characterise two ends of the same spectrum, rather than two distinctive positions—the former is more implicative of dependence on shared attitudes, and the latter more implicative of dependence on individual attitudes. Ultimately, I prefer the concept of stance dependence to characterise my own account because—and this will become clear in chapter 6—I understand individual attitudes towards well-being to be embedded in broader socio-linguistic contexts, which shape and delimit them. Even when they are idiosyncratic, individually held conceptions of well-being have to be understood in these contexts in order to assess their content and intelligibility.

### *Roadmap*

The thesis is divided into three parts. Part I comprises two chapters, which outline my approach to thinking about well-being. Chapter 1 makes a case for taking the variety of concepts and measures of well-being at face value *as* measures of well-being, and not as measures of some other concept. Chapter 2 argues that theories of well-being must also be theories of the measurement of well-being.

Part II comprises three chapters, which explore in depth three different approaches to conceptualising and measuring well-being, and in which I begin to establish what a



close consideration of such practices can tell us about the nature of well-being. In Chapters 3, 4 and 5, I consider theories of well-being based on objective goods, preferences and subjective experience respectively.

Part III accounts for just one chapter, in which I draw together my conclusions from part II, and develop my positive view. I set my view against two alternative positions, and defend it against criticism from both. The first is the stance independent pluralist account, which maintains that well-being is indeed plural, but rejects the attitude dependence of well-being assertions. The second, which represents a view more sympathetic to my own, is the view that the truth and content of well-being assertions is dependent on the objective context of the speaker, rather than their psychological state.

Finally, I offer some brief concluding remarks.



# Part I



## CHAPTER 1

### The case against well-being invariantism

Philosophical theories of well-being tend to give definitions of well-being which specify conditions that are met by all and only instances of well-being, and where well-being is assessed over human lives, considered globally.<sup>1</sup> So, for example, a *hedonist* account of well-being might say that well-being is a matter of the felt-experience of pleasure: lives that contain more pleasure contain greater well-being, and lives that

---

<sup>1</sup> While discussion of well-being is a longstanding feature of western philosophy, my particular interest is in well-being theorists working in the last couple of decades. My focus in this chapter lies with contemporary well-being theorists such as Ben Bradley (2009), Ben Bramble (2016), Donald Bruckner (2016), Roger Crisp (2006), Fred Feldman (2002, 2004), Guy Fletcher (2013), Philippa Foot (2001), Chris Heathwood (2005, 2006), Thomas Hurka (1993), Richard Kraut (2007), Marc C. Murphy (1999), Christopher Rice (2013), Valerie Tiberius (2011). In addition to this cluster, there are a some slightly earlier accounts which take a roughly similar approach, such as those defended by Richard Brandt (1979), John Finnis (1980), James Griffin (1988), Martha Nussbaum (1988, 1992), Derek Parfit (1984) and Peter Railton (1986). These are essentially the views that I take to represent the current philosophical orthodoxy about well-being, and many of them are the philosophers of whom I am thinking when I speak of well-being invariantists.

contain less pleasure contain less well-being.<sup>2</sup> Satisfaction of desires, personal flourishing, and securing goods such as health, education, friendship, love, and political power only contribute to well-being insofar as they contribute to the amount of pleasure experienced by individuals. Goods that do not bring pleasure do not contribute to well-being. Or, alternatively, a *desire-satisfaction* account of well-being might take well-being to consist in the satisfaction of desires, or the satisfaction of rational desires.<sup>3</sup> On such accounts, someone is doing well to the extent that their desires, or their rational desires, are satisfied, and doing badly to the extent that they are not. Not all philosophical theories of well-being specify a single constituent of well-being. Pluralist *objective list* accounts take well-being to consist in the securement of some set of irreducible objective goods.<sup>4</sup> On these accounts, well-being consists in the securement of some or all of the items on the specified list of goods, where the items

---

<sup>2</sup> Classic examples of hedonist theorists include, of course, Jeremy Bentham (1843) and John Stuart Mill (1998); contemporary examples include Bradley (2009), Bramble (2016), Crisp (2006), and Feldman (2002, 2004). There is also a strain of psychologist-philosophers attempting to reinstate direct measures of positive subjective psychological experience as the central component of well-being, including Daniel Kahneman (1999).

<sup>3</sup> Brandt (1979), Bruckner (2016), Heathwood (2005, 2006), Murphy (1999) and Railton (1986) all defend desire-satisfaction accounts of well-being. Heathwood and Murphy propose relatively unrestricted views, whereas Brandt, Bruckner and Railton impose restrictions on the desires whose satisfaction can contribute to well-being.

<sup>4</sup> See Fletcher (2013) and Rice (2013). Some perfectionist accounts, such as Finnis (1980) and Nussbaum (1988; 1992) take a similar form, though tend to take the prudential value of the goods to be explained by their relation to an account of human flourishing or functioning. Perfectionist theories have Aristotelian roots (Aristotle 2009, particularly Book I), and typically provide naturalist explanations of prudential value, grounded in human natural history. Non-perfectionist objective list theories do not necessarily seek to provide such an explanation of the prudential value of the goods.

on the list constitute or contribute to well-being regardless of our attitudes towards them. Other theories of well-being propose slightly more complex conditions, often taking well-being to consist in some combination of these factors.<sup>5</sup>

By contrast, outside of philosophical contexts, well-being is measured and defined in a wide variety of ways, which use different metrics and evaluations depending on the object of concern and the context. The concepts of well-being used in the social and clinical sciences, and in public policy, do not presuppose universal conditions of application, nor do they measure well-being only across whole lives. Some are general, and supposed to apply across all human beings, and others are very specific, relevant only to people with very particular characteristics, at particular points in time. Some take well-being to be a subjective indicator, and others to be an objective construct or set of constructs (Alexandrova 2012b, p.680). For example, development economists may measure the well-being of a population by measuring their access to a specified set of resources, whereas clinicians might measure well-being by measuring the physical and psychosocial function of patients who suffer from a particular disease. Measures of well-being in children will likely differ markedly from measures of well-being in adults (Raghavan and Alexandrova 2015; Tomlin 2018). In more quotidian contexts, for a friendly stranger on the street who asks how you are when you trip up, the markers of your well-being and the standards for evaluating it will be quite different from the markers and standard used by a friend concerned about your mental

---

<sup>5</sup> Mixed theories include Parfit's account (1984, p.502), which takes well-being to consist in securing objectively good things (Parfit suggests having knowledge, engaging in rational activity, experiencing mutual love, and being aware of true beauty as examples) while strongly wanting just these things. Shelly Kagan (2009) defends something similar, arguing that well-being consists in having objective goods in my life in which I take pleasure. Idealised subjective theories, which place ideal constraints on the elements of a person's psychological state that can contribute to well-being, are fairly common too (see Brandt 1979, Griffin 1986, Railton 1986, Sumner 1996 and Tiberius 2011).

health, and again from those used by a social worker looking to ascertain your ability to cope as a new parent (Alexandrova 2013, p.311–313).

In this chapter, I start to address some of the considerations involved in developing and justifying an account of well-being which meets both philosophical and practical demands. In 1.1, I begin with a characterisation of the contemporary orthodoxy in philosophy with respect to theories of well-being. My characterisation of the philosophical landscape is similar to accounts given by Anna Alexandrova and Guy Fletcher, and in 1.2 I outline their views. My own account is most similar to Alexandrova's distinction between invariantist and variantist theories of well-being, and I retain much of her terminology, despite minor differences. In 1.3 and 1.4, I suggest two ways in which the invariantist might try to account for the variety of measures of well-being in use, and argue that both are unconvincing. The failure of well-being invariantism to either accommodate or dismiss the variety of measures of well-being, lends credence to well-being variantism. In 1.5, I suggest that well-being variantists can take either a *stance dependent* or *stance independent* view of the content of well-being claims, and outline these two possibilities. In 1.6, I argue that the divergence between invariantism and variantism about well-being is grounded in a deeper disagreement about the enterprise of moral philosophy, and ultimately about nature of concepts and their relation to linguistic use. This makes the disagreement between the two approaches pretty intractable, and convincing the invariantist is going to require more than a few counter-examples. My thesis develops a stance dependent pluralist account of well-being, which I advance as a plausible alternative to invariantist theories.

### 1.1 *The philosophical orthodoxy*

My argument ultimately stems from a dissatisfaction with the way that contemporary philosophers of well-being approach their subject. My characterisation is not necessarily one that is explicitly endorsed by the philosophers whose views I take to be



representative of this approach, though I argue that it is *implicitly* endorsed by them. The approach in question has, however, also been identified and criticised by other philosophers (Alexandrova 2012a, 2012b, 2013, 2017b; Fletcher 2009). More recently something like it has also been acknowledged and defended (Lin 2018). There are three common features of many theories of well-being that strike me as misguided, or at least as limited in the extent to which they should form a central part of theorising about well-being. Not all contemporary theories of well-being demonstrate all of these features, and not necessarily explicitly or to the same degree, but I take there to be a broad tendency to think in these terms, and I believe this to be borne out by the examples that I explore in this section. It is not only philosophers who exhibit these tendencies, and later on, in Part II, I examine similar claims made by certain economists and psychologists. But, for now, my focus is philosophers and, specifically, a cluster of well-being theorists working in the last 20-or-so years. The three features of well-being theories that I pick out here are *globalism*, *reductionism* or *uniqueness*, and the use of *a priori reasoning*.

First, then theories of well-being tend to operate with a globalist conception of well-being, that is, well-being is understood to be a global assessment of a person's life, a claim about how someone's life is going overall. This entails that context-specific assessments of well-being are either not assessments of well-being at all, or merely partial contributors to well-being. Alexandrova calls this feature of theories of well-being the *deathbed assumption* (2012a) or *circumscription* (2013, 2017b). Many well-being theorists explicitly or implicitly invoke a form of globalism about well-being. For Roger Crisp, a hedonist, the key question that arises for theories of well-being is: 'What makes a life good for an individual?' (2006, p.101). Chris Heathwood, a desire-satisfaction theorist, explains this theory as the view that 'one's life goes well to the extent that one's desires are satisfied' (2005, p.487). Ben Bramble states: 'Theories of well-being attempt to explain what it is in virtue of which lives can be good or bad for

their subjects' (2016, p.136). Derek Parfit, who defends a mixed theory that characterises well-being in terms of the attainment of objective goods that are also desired, asks '[w]hat would be best for someone, or would be most in this person's interests, or would make this person's life go, for him, as well as possible?' (1984, p.493).

These are all fairly explicit expressions of globalism; for each, the scope of a theory of well-being is a person's life. Assessments of well-being are judgements about a person's life as a whole, not judgements about a particular area of their life: their health, or their career, or their personal relationships. Moreover, assumptions or statements of globalism take root across different kinds of theories of well-being. Objective list theorists often write in terms which, while not necessarily contradictory to globalism, are less explicitly supportive of it. Guy Fletcher, for example, speaks of well-being as what is 'good for us' (2013). Christopher Rice suggests that an account of well-being is an account 'of what ultimately benefits people' (2013, p.197). Neither of these articulations explicitly imply globalism, though they are consistent with it. However, Rice slips into talk of well-being reflecting 'human life in general' and what makes 'life go well' (2013, p.204; p.211), and Fletcher takes well-being to be a matter of 'how well one's life goes,' and for theories of well-being to specify the constituents of a 'good life' (2013, p.216; p.214). So such theories appear to subscribe to a form of globalism too.<sup>6</sup>

---

<sup>6</sup> Some philosophers acknowledge the existence of temporal well-being (Broome 2004; Kagan 1992; Velleman 1991) and entertain the idea that well-being at a particular time is best understood in subjective terms, whereas well-being over a lifetime is best understood in more objective terms. While accounts of temporal well-being might take an appropriate subjective assessment to be an assessment of all aspects of life at that point in time, they are not strictly globalist, as they do not take as their subject a life as a whole. Insofar as they adopt more than one characterisation of well-being, neither are they

Secondly, theories of well-being tend to seek to define the essence that characterises all true instances of well-being, and so to provide a single, clear definition of well-being which applies across all contexts. On such views, whatever enhances well-being in one context, for one person, must enhance well-being in all contexts, for all people. Alexandrova calls this feature of theories of well-being the *uniqueness* criterion (2012a, p.627). Fletcher calls it *invariabilism* (Fletcher 2009).<sup>7</sup> Sometimes this reductiveness is stated baldly. Ben Bradley defines hedonism as ‘the view that states of affairs consisting of a person getting some pleasure—pleasure states—are the only positive value atoms, and pain states are the only negative value atoms’ (Bradley 2009, p.8). (For Bradley, ‘value atoms are what fundamentally and completely determine how well things go for us’ which seems to commit him to globalism too (2009, p.5).) In fact, Bradley suggests that a theory of well-being will necessarily be reductive in this sense: ‘One might wonder about the possibility of a theory of well-being. It might be thought: What counts as a good life for one person would not be a good life for someone else’ (2009, p.3). This implies that if different things constitute a good life for different people, we are left without the possibility of a theory of well-being, that is, theories of well-being must specify a definition which applies to all instances of well-being.

---

fully reductive. I understand such views to take a step towards contextualism, although, in my view, they don’t go far enough.

<sup>7</sup> Eden Lin (2018) uses *invariabilism* to refer to the view that the same theory of welfare is true of every welfare subject. Lin claims that this is distinct from Fletcher’s sense of invariabilism, which takes theories of well-being to be kind-relative. I am not convinced that Fletcher’s conception of invariabilism *does* take well-being to be kind relative, though it is true that most well-being theorists restrict their discussion to considerations of human well-being. In any case, Lin’s offers an explicit defence of the reductiveness of well-being, arguing that the same theory and concept of well-being should apply to all subjects, including non-human ones, in all contexts.

Other hedonists make similarly clear claims about uniqueness. Fred Feldman says: 'I want to know, in the abstract, what features make a life a good one for the one who lives it' (2004, p.13). Crisp claims that a theory of well-being is a theory of 'what of is ultimately good for any individual,' and defines hedonism as the view that 'what is good for any individual is the enjoyable experience in her life, what is bad is the suffering in that life, and the life best for an individual is that with the greatest balance of enjoyment over suffering' (2006, p.102). Chris Heathwood defines the desire-satisfaction view in similarly reductive terms:

The desire-satisfaction theory of welfare says, roughly, that one's life goes well to the extent that one's desires are satisfied. On standard 'actualist' versions of the theory, it doesn't matter what you desire. So long as you are getting what you actually want—whatever it is—things are going well for you. (2005, p.487)

For Thomas Hurka, who defends a perfectionist account of well-being, or what he prefers to call *the human good*, 'theories of what can be called either "well-being" or "the human good"...connote what's ultimately desirable or worth pursuing as an end for yourself and for other people' (Hurka 2016, p.379). In each of these cases, defining well-being is an undertaking that can be done in abstraction from the particulars of individual cases. Well-being is something that is shared by all human beings (and perhaps, in some cases, by non-human animals and non-animal organisms too). Of course, for all of these theories, determining whether someone is doing well or not requires consideration of particular people and their circumstances. But this would involve identification of the same state, property or characteristic in each person.

Those theories that define well-being in terms of a plurality of goods leave some room for variation among lives that go well and may, to that extent, be thought to avoid commitment to a reductive approach. However, Fletcher argues that objective lists that

characterise well-being in terms of attitude independence are invariabilist (2009, pp.26–7). He cites James Griffin:

So the most plausible form of objectivism allows that many values—many more than the ones linked to basic needs—are objective. The thought behind forming them into a standard of well-being would be this: when they appear in a person's life, then whatever his tastes, attitudes, or interests, his life is better. (1988, p.54)

Fletcher also cites John Finnis, who argues that the goods on his list are constitutive of well-being 'whether I like it or not' (1980, p.72). The suggestion here is that even though one can have a good life without one or some of the goods that are constitutive of well-being, the list characterises a single, invariant concept of well-being, because one's life is always better in virtue of having any of the specified goods in it. Fletcher's own objective list theory, which includes only goods that have pro-attitudes as necessary components, is constructed in order to avoid this implication (2013, p.216). If the prudential goods are constituted by pro-attitudes, then people cannot be alienated from their own good—a concerning possibility attached to objective list theories whose goods don't involve pro-attitudes.

However, there remains a sense in which Fletcher's list of prudential goods *is* attitude independent. That is, the fact that it is these goods, and not some other set of goods, that constitute well-being is not an attitude dependent matter. These goods constitute well-being because they are attitude dependent, but their constituting well-being is not attitude dependent. For Fletcher it is only achievement, friendship, happiness, pleasure, self-respect and virtue that enhance well-being. Other goods—health, political and educational resources, knowledge, parenthood—don't enhance well-being unless they lead to, or can be understood as instances of, the pro-attitude goods. So it looks as though it remains the case that, for Fletcher, one's life is always better in

virtue of having any of the goods on his list. Of course, one's life is only better insofar as one has pro-attitudes towards the goods, but if having the goods necessitates having pro-attitudes towards them, because of the kind of goods they are, then one's life is necessarily better in virtue of having them. So there is a case to be made for thinking of such theories as offering a pluralist, and yet nonetheless reductive definition of well-being—in the sense of specifying a finite list of goods which constitute or contribute to well-being, and whose contribution to well-being is attitude independent.

Thirdly, theories of well-being use *a priori* reasoning in the development and justification of their respective concepts of well-being. The specification of a definition of well-being is taken to regulate the use of the concept and the appropriateness of well-being claims. So the definition of the concept of well-being is understood to be something like a rule which, when invoked in particular contexts, determines whether a particular case is an instance of well-being or not. This, I take it, is similar to Alexandrova's idea of circumscription. I prefer to keep this aspect of theories of well-being separate from globalism, because I take it to reflect an attitude towards well-being, and perhaps more widely, that is prior to a particular commitment to both globalism and reductionism, and upon which both of these commitments rest. Eden Lin comes close to an explicit endorsement of *a priori* reasoning as a means of coming to know instances of well-being when he writes: 'If invariabilism is true, then the one true theory of welfare enables us to identify all of the welfare subjects' (2018, p.322). Gerry Cohen (2003) also endorses such an approach in a broader sense when he argues that our most basic moral principles do not reflect or respond to truths about matters of fact. Cohen gives an example of a young woman who comes to believe that abortion is wrong, after falling pregnant unexpectedly. He characterises her newfound belief thus:

She has come to believe that it is wrong to take a human life, as such, and she has come to believe this as a result of the experience of not being able to take

(or direct the taking of) a human life herself. But, if she is (perhaps unusually) reflective, she will realize that she has come to believe a principle (do not take lives of creatures with humanlike features) whose authority for her is independent of any facts of experience that she or others might have learned. She must believe that it would have been right for her to accept that principle whether or not she had had any experience that induced a belief that lives of human beings in particular should not be taken.<sup>8</sup> (2003, p.232)

For Cohen, moral reasoning is necessarily *a priori* reasoning. To think that something is wrong or right just is, for him, to hold a general principle which can be invoked in particular contexts to make moral claims about the world, but is not itself answerable to experience or facts. Similarly, responses to particular examples or problems are taken to reveal general truths about well-being.

The commitment to giving a context-independent account of the necessary and sufficient conditions of well-being seems to lean on what Mark Wilson (2006) calls the ‘classical picture’ of concepts. Wilson (2006, p.5) attributes this view of concepts to Russell; Ludwig Wittgenstein traces a similar view back to Plato (Forster 2010, p.71). On this account, concepts are constant across time, provide us with exact guidance about their application, and have a fixed scope. A concept picks out some particular characteristic or set of characteristics which objects or states of affairs can have or fail to have. Using a concept involves assessment as to whether some state of affairs or objects satisfies its requirements, where apprehension of the concept and the objects under consideration will determine whether the objects lie within or outside of its scope (Wilson 2006, pp.90–3). This means that a definition provides necessary and

---

<sup>8</sup> Cohen’s emphasis.

sufficient conditions for its own application, and is used to adjudicate the correctness of any usage of the term.

Of course, the definition of well-being doesn't appear out of thin air, and for most of the theorists in question it appears to be grounded in intuitions or responses to particular cases and thought experiments. The development of general principles out of particular examples does not indicate that the principles are not a priori. In Cohen's example the young woman changes her beliefs following a particular experience, but the authority of principle that she acquires is independent of any facts of experience. So particular experiences and responses may be thought of as revealing general principles about well-being to us, rather than justifying or grounding them. Sometimes the intuitions about well-being are very general: Crisp, for example, thinks that '[e]njoyment seems just obviously part of well-being (2006, p.124). Finnis claims that '[i]t is obvious that a man who is well-informed, etc., simply is better-off (other things being equal) than a man who is muddled, deluded, and ignorant' (1980, p.72). Finnis takes the contribution of knowledge to well-being to be self evident: 'It cannot be demonstrated, but equally it needs no demonstration' (1980, pp.64–5). But intuitive responses to particular thought experiments also play an important role for many well-being theorists. Nozick's experience machine, for example, is often cited as a counter-example to hedonism, or defended against such claims (Bramble 2016; Crisp 2006; Haybron 2016; Hurka 2016). And intuitive counter-examples play an important role in ruling out certain classes of theory—the existence of irrational, sadistic, base mindless and false desires and pleasure for are taken to present particular problems for desire-satisfaction theories and hedonism (Fletcher 2009; Heathwood 2006; Hurka 2016; Tiberius 2011). It is difficult to conceive of ethical, indeed philosophical, reasoning that does not depend on intuitions in one way or another—so that is not my



particular concern here.<sup>9</sup> However, using intuitions about particular cases as grounds for making generalisations about the concept of well-being in general, which apply in all other cases, presupposes that the concept of well-being is reductive and that it is self-regulating. The reductiveness of the concept is implied by the suggestion that counter-examples to particular accounts of well-being rule it out as a definition of well-being in all cases. The self-regulation of the concept is implied by the suggestion that comprehension of particular cases is sufficient for determining that they do not lie within the scope of the concept of well-being.<sup>10</sup>

### 1.2 *Well-being invariantism and well-being variantism*

I have already noted that my characterisation of theories of well-being is similar to that of Alexandrova and Fletcher. Alexandrova (2012a, p.625–6) argues that philosophers standardly take well-being to be a global evaluation of what is good for a person—what makes their life as a whole go well. Though there is disagreement about what this good specifically consists in, most theories of well-being take there to be a single, correct substantive definition of well-being. She calls this approach *well-being invariantism*. Invariantism is characterised by two key claims: the deathbed claim—that the best characterisation of a person’s well-being is the one which considers and evaluates their entire life—and the uniqueness claim—that a theory of well-being specifies conditions

---

<sup>9</sup> I do find the invocation of self-evidence as a form of justification, as seen in the second quotation from Finnis here, troubling, and I offer a criticism of the invocation of self-evidence in normative reasoning in chapter 3.

<sup>10</sup> Sometimes counter-examples and intuitive responses to particular cases are used in order to refute the universality of particular theories of well-being. I take this approach to avoid the concern about *a priori* reasoning, to the extent that it is not used as the basis for generalisations about well-being. Fletcher (2009, 2016) makes effective use of such arguments.

that apply in all and only instances of well-being (Alexandrova 2012a, p.627).<sup>11</sup> Elsewhere Alexandrova calls the idea that well-being always refers to a general, all-things-considered evaluation ‘circumscription,’ because it entails that evaluations of someone’s life that are not global specify something other than well-being, thus narrowly circumscribing the definition of well-being (2013, p.313; 2017b p.5). Guy Fletcher (2009) outlines and critiques *well-being invariabilism*, a view which seems to be roughly in line with Alexandrova’s uniqueness claim. Fletcher takes well-being invariabilism to be the view that ‘[a]ny X that non-instrumentally enhances well-being in one context must enhance well-being in any other’ (2009, p. 22). Fletcher does not appear to take invariabilism to be committed to a globalist assumption as well. Indeed, as discussed above, Fletcher’s own objective list view (2013) appears to endorse a form of globalism.

As discussed in the previous section, many theories of well-being do take assessments of well-being to require a God’s-eye-view, global evaluation of someone’s life as a whole. However, the sense in which the evaluation is global is different in different cases, which may call into question Alexandrova’s deathbed nomenclature.<sup>12</sup> On some theories, for example, hedonic theories of well-being, this global evaluation may be done by adding up all of the momentary pleasures and pains experienced by an individual across their lifetime, though the value of those pleasures and pains must be determined at the time of experience, not at the time of summation. In such cases, an accurate (as far as it goes) assessment of well-being can be given at any point in a

---

<sup>11</sup> Though one could be invariantist or variantist about concepts other than well-being, I use ‘invariantism’ and ‘variantism’ as a shorthand for ‘well-being invariantism’ and ‘well-being variantism’ throughout the thesis.

<sup>12</sup> Indeed, she too appears to have dropped this designation in her recent work (2017b) and to have settled on ‘circumscription’ instead.

person's life, with their deathbed, or the end of their perceptual experience, being the point at which the final reading can be taken. On other theories, however, the distribution of well-being across a lifetime, or the order in which certain key events occur may affect the overall assessment, so the relative value of moments can only be determined retrospectively, at the end of their life. Here, an assessment of well-being won't be complete until the end of life, as any occurrences are liable to re-evaluation in light of later events or evidence. On still other theories—some desire-fulfilment models of well-being, for example—events occurring after someone's death will affect their well-being if they result in, or preclude, their living desires being satisfied. This means that even after a person's life has ended, there may be further changes, or the possibility of further changes, to a full evaluation of their well-being. A global assessment of well-being will require consideration of these states of affairs, so the jury can remain out on the status of someone's well-being indefinitely, or at least for a long time, after their death, particularly if their most deeply held desires involve long-term projects or outcomes.

What it means for well-being to be a global assessment is therefore variable. For this reason I prefer not to think of this aspect of theories of well-being as a deathbed assumption. I also take circumscription about well-being to be a broader feature of their accounts, and not merely a feature of globalism: reductionism also implies a form of circumscription about the content of well-being claims. So I am not entirely content with Alexandrova's designation of this assumption as 'circumscription' either—though I do think that the commitment to globalism *is* an example of circumscription. My use of 'globalism' reflects this dissatisfaction with Alexandrova's terminology. I will, however, continue to use 'uniqueness,' in line with Alexandrova, to refer to the assumption that well-being does not vary across context and that a reductive definition of well-being can be specified. Furthermore, I will use Alexandrova's terminology of *invariantism* to refer to the orthodox theories of well-being which are typically

characterised by commitments to globalism, uniqueness, and use of *a priori* reasoning, and *variantism* to refer to the view which rejects this approach to theorising about well-being. Despite the close relation of Fletcher's invariabilism to invariantism, his exclusive focus on the uniqueness claim leads me away from his characterisation and towards Alexandrova's. I am less willing than Alexandrova to specify fixed conditions for invariantism, as I think that those theories of well-being that embody some, but not all, of these features are also misguided to that extent. Nonetheless, invariantism captures my inclination pretty well, and there is, I think, some merit in adopting established language. The views that I characterise as stance independent pluralist (see 1.5) often retain features of invariantist reasoning about well-being—many remain committed to globalism and certain forms of *a priori* reasoning, despite embracing a more or less limited pluralism about the definition of well-being. And in some sense I take these factors to be more important in characterising the kinds of views that they are than the fact that they are not strictly invariantist.

Alexandrova and Fletcher reject invariantism and invariabilism respectively. Alexandrova (2012a, 2012b) argues that, in non-philosophical contexts, well-being is measured and discussed in a large number of different ways, which use different metrics and standards depending on the object of concern and the context. These metrics and standards appear to rely on different substantive theories of well-being, not just different methods of realising the same substantive account (2012a, p.639). Some are general, supposed to apply across all human beings, and some are very specific, relevant only to people with very particular characteristics. Some take well-being to be a subjective indicator, and others take it to be an objective construct or set of constructs (Alexandrova 2012b, p.680). While markers of well-being in a development economics context may measure the quality of life of an entire population in terms of their access to healthcare, nutrition, resources, and public goods, such as clean air and water (Anand and Sen 1994), markers of well-being in a

clinical context might instead measure the quality of life and symptoms of particular individuals or small groups of individuals in relation to a disease or course of treatment.<sup>13</sup> Someone who measures highly on the first set of markers might measure very poorly on the second set, and vice versa. In ordinary contexts, the standard of well-being used by a university tutor asking a first-year undergraduate how they are doing at the end of their first term is different from the standard used by a close friend enquiring about their well-being following a difficult break-up with their boyfriend from home, and different again from the standard used by their GP when they ask them how they are doing in an appointment scheduled to discuss a long-term health condition.

Alexandrova defines *well-being variantism* in contrast to invariantism (2012a, 2012b, 2013, 2017b). The well-being variantist rejects the invariantist's circumscriptionist and uniqueness claims. Instead arguing that different accounts of what it is to do well are appropriate in different contexts. Variantism is also characterised by two claims: first, that well-being can invoke either general or local evaluations, depending on the context. Secondly, that there is no single substantive theory that characterises every

---

<sup>13</sup> Furthermore, clinical measures of well-being vary in their specificity. Some measures of health-related quality of life are intended to be used across all patients, for example, the EQ-5D, SF-36, and Nottingham Health profile (EuroQol Group 1990; Ware and Sherbourne 1992; Hunt et al 1980). Others are intended for use in a clinical subgroup, for example, the Quality of Life Index-Cardiac Version, Cardiac Depression Scale and Cardiac Symptoms Scale are all designed to measure health-related quality of life in patients with heart disease (Ferrans and Power 1985; Hare and Davis 1996; Plach and Heidrich 2001). Still others are intended for use in groups of patient with a particular disease, or who have undergone a specific intervention, for example, the Kansas City Cardiomyopathy Questionnaire, Minnesota Living with Heart Failure Questionnaire, and Cardiac Surgery Symptom Inventory (Green et al 2000; Rector et al 1987; Miller and Grindel 2004). As the titles of some of these measures indicate, they also focus on different aspects of health-related quality of life—some are more symptom focussed, others focus on psychological state, others on a patient's ability to execute daily tasks and activities.

concept of well-being (Alexandrova 2017b, p.43). Determining the right account of well-being to use in a given case requires assessment of the particular characteristics of the population and circumstances under consideration, and an evaluation of the ends and aims of measurement. The well-being variantist takes the global accounts of well-being proposed by well-being invariantism to be only a small subset of many accounts of well-being, often developed for specific purposes or applicable only for particular groups of people. She thus doesn't deny that there are valid global conceptions of well-being, just that any of them is the sole valid conception of well-being. Alexandrova proposes that it is better to embrace this variation as central to the notion of well-being, and to accept that different substantive conceptions of well-being are at play in underpinning the wide variety of measures of well-being in use, rather than to insist on a single, invariant concept of well-being that explains them all (2012a, p.628). Alexandrova argues for well-being variantism as a methodological thesis about how best to make sense of the conceptions of well-being that are used in the social and clinical sciences. Taking measures of well-being at face value, as indeed capturing well-being, does away with the need to explain away the difference between philosophical and ordinary language about well-being. Furthermore, taking a circumscriptionist attitude excludes philosophers from the majority of the debate about well-being, which occurs across the social and clinical sciences, and in policy-making contexts (2012a, p.638).

Fletcher's *well-being variability* is, again, quite similar to Alexandrova's variantism. The well-being variability, for Fletcher, is committed to the claim that 'Any X that non-instrumentally enhances well-being in one context may fail to enhance well-being, or detract from well-being, in another context' (2009, p.22). Well-being, on such a view, may differ depending on the time and context of measurement. Fletcher's argument proceeds on the grounds that objections to competing invariabilist theories of well-being typically accept that well-being is *sometimes* constituted by the good in question

but not always. This, he thinks, should not lead us to reject the theories altogether, but to reject them as universally applicable accounts of well-being. Fletcher does not, unlike Alexandrova, emphasise that the different aims of measurement—whether collected for a development project, a public health intervention or an evaluation of the effectiveness of a therapy—will partly determine the appropriate account of well-being. However, allowing for a reasonably broad interpretation of ‘context’ Fletcher’s definition easily accommodates this. Fletcher’s account of variability, which is developed as an alternative to invariabilism, is chiefly negative and he does not focus on the specifics of defining and measuring well-being in practical contexts. Nor does he include a condition which rejects globalism in his characterisation of variability. Once again, while Fletcher’s critique is highly relevant to my argument, I take Alexandrova’s variantism to be closer to my own account than variability, and I therefore adopt her terminology in my discussion.

Invariantism and variantism are not straightforwardly accounts of what well-being consists in. They don’t tell us whether well-being should be understood in terms of pleasure, or flourishing, or the fulfilment of basic needs. We might instead think of them as meta-theoretical specifications of what kind of a concept well-being is, and as constraining rather than determining the possible content of well-being. The plausibility of variantism would generate significant ramifications for philosophy of well-being, and the orthodox framework for understanding well-being would be disrupted. First, then, the invariantist should be given the chance to respond to the variantist challenge and to explain the apparent variation in the concept of well-being.

### *1.3 Defending invariantism I: Multiple bearers of well-being*

There are a number of strategies the well-being invariantist may pursue in order to attempt to account for or explain away the variety of ways in which well-being is measured and defined across the clinical and social sciences. This and the following section will address the two most plausible approaches for the invariantist. On the first

approach he may argue that the different measures of well-being are in fact different bearers of a shared foundational concept, so they each measure the same invariantist concept of well-being via different means. On the second approach the well-being invariantist may claim that the supposed measures of well-being identified by the variantist account measure not well-being, but other related concepts. True well-being is captured uniquely by a particular invariantist account. The latter approach requires an error theory in order to explain why existing measures that allegedly capture well-being in fact fail to do so, and why well-being is misconceived to such a great extent.

First, then, the invariantist might argue that all of the different measures used across the social and clinical sciences and in development economics are, in fact, different ways of capturing the same overarching concept of well-being. Despite apparent surface level differences, they all measure aspects of the same base value, which is an invariantist concept of well-being. Metrics of well-being may also, on such an invariantist account, turn out to be proxies for well-being, which do not capture its substantive content exactly, but approximate or partially represent it. Such a view might take it that well-being is conceptually stable, but different contexts have different markers of well-being. So, for example, well-being may be conceived of in terms of flourishing or the development of characteristically human capabilities, but what this comprises in different contexts and for different people may end up looking quite different.

The following four examples of context-specific concepts of well-being give an idea of the kinds of differences the invariantist needs to subsume under a single theory of well-being:

- i. *End-of-life*: Improving well-being for someone in the final stages of a terminal disease might include minimisation of symptoms and pain; treatment in accordance with their autonomous wishes, for example about the place and time



of their death and the medical care they receive; and treatment with dignity and respect, for example if they need help washing, dressing, eating or communicating. At the end of life, some of the traditional clinical indicators of health-related well-being, such as disease prognosis and life expectancy, are less likely to be key considerations, and length of life may be traded off against suffering. Well-being may consist, partly, in securing a good death rather than extension of life. The subjective psychological state and preferences of the patient are likely to be the focus of well-being, rather than objective indicators.

- ii. *Anorexia*: For someone suffering from life-threatening anorexia, restoring them to a healthy weight and improving their physical functioning, or limiting loss in physical functioning, is likely to be of primary concern. As people with anorexia typically have body dysmorphia, and may have desires which endanger their lives or put them at risk of permanent disability, their desires may be taken to be largely irrelevant to their well-being, and their interests may be best partly or entirely determined by their medical and social care team. Treatment in accordance with the patient's desires is therefore likely to be temporarily overridden until their psychological functioning has improved, leading to the justification of measures such as enforced feeding, use of chaperones while using the bathroom, and involuntary detention. Furthermore, measures of the patient's pleasure and subjective psychological states while being fed and watched over may not be taken to be a relevant indicator of their well-being at early stages in their recovery, while later on in the recovery process the indicators of well-being will likely change, and may include personal decision making, positive subjective experience, and personal freedom, in addition to clinical indicators and life expectancy.
- iii. *Profound cognitive impairment*: People with profound and multiple learning disabilities (PMLD) tend to have more than one serious disability, such as severe autism, Down's syndrome, cerebral palsy, epilepsy or microcephaly, resulting in

profound cognitive impairment and learning difficulties, as well as one or more sensory impairment and limited control over their physical actions (Goldbart and Caton 2010). They are also typically non-verbal, with limited use of formal linguistic or graphical code and no prospect of learning or understanding language throughout their lives. Significant difficulties in communication—including not only difficulty in interpreting communicative intentions and desires, but also uncertainty about whether their actions are in fact intentional—are thus almost inevitable. Assessment of well-being in people with PMLD will likely refer to their perceived interests, with reference to an account of their flourishing or functioning. What it is for a person with PMLD to flourish will inevitably be very different from what it is for a person without learning disabilities, or with mild disabilities, to flourish, given the differences in physical and psychological capacity. Assessment of well-being will also likely involve consideration of their preferences and subjective psychological states, but communication difficulties mean that these must be assessed based on careful consideration of responses to parents, carers and doctors, rather than via self-report alone, as would normally be the case in people without such difficulties in communication and comprehension. An appropriate account of well-being for use in people with PMLD may not refer to their autonomous desires, or informed desires, as it's often not possible to ascertain these, and in extreme cases it might not be appropriate to ascribe agency at all (Grove et al. 1999).

- iv. *Development economics*: Measures of well-being are used in development economics to evaluate public policy initiatives and interventions at a population level. Typically such metrics define well-being in terms of a set of objective variables, for example, indicators of private consumption, such as access to food, shelter, clothing, and basic legal aid; indicators of health, such as life expectancy at birth or population rates of STDs or other diseases; indicators of education such

as literacy rates and school attendance rates; and civil and political rights such as the rights (and opportunity) to vote, to freely assemble and to criticise aspects of government (Dasgupta 2001). These variables are taken to be irreducible to a master variable. They do not usually include, nor are usefully expressed in terms of, the happiness of citizens, or the satisfaction of their desires. Where subjective affect or desire satisfaction is included, either as one of the variables, or as a consideration which helps to shape an appropriate set of variables for a particular context, it is not taken to be a master value in terms of which all other variables can be explained.

In each of these four contexts, different aspects of people's lives are prioritised in order to evaluate their well-being. The chief determinants of well-being in one context appear irrelevant to well-being, or even inconsistent with it, in other contexts. For someone in the final stages of a terminal illness, the fact that they are literate, STD-free and have access to legal aid is not relevant to an assessment of their well-being in this sense. For someone with PMLD, evaluation of their well-being in terms of their self-determination and autonomous decision making might well characterise them as incapable of having well-being at all, or may fail to distinguish between their level of well-being in cases where they are attentively cared for and cases where they are abused and maltreated. For someone with life-threatening anorexia, ensuring they have freedom for autonomous decision making and self-determination may well be inconsistent with their well-being, if they will continue to put themselves in mortal danger through their self-destructive preferences.

Finding a single thread which runs through all of these contexts (and countless more) is undoubtedly a difficult task. Such an account would have to allow that well-being is sometimes best conceived of in terms of subjective psychological states, and at other times in terms of non-subjective consideration of needs, rights and appropriate treatment or care. Defining well-being in terms of autonomy or independent decision

making is likely to fail to recognise the positive subjective experience and the formation of loving relationships that are accessible to people with PMLD as markers of well-being. Similarly, to insist on using a model of well-being which focuses on securement of civil and political rights in assessing someone with a terminal illness may fail to recognise their remaining possibilities for self-actualisation and satisfaction. What the four examples make clear is not only that assessment and promotion of well-being requires different action in different cases, but moreover that the identifiers of well-being are quite distinct in different cases, both conceptually—whether it's autonomy, subjective psychological state or access to resources that is constitutive of well-being—and ontologically—which physical and psychological markers indicate well-being or its absence.

If invariantism is to capture this variance, probably the most plausible approach is to characterise well-being in terms of 'flourishing' or 'functioning' where the specifics of what these comprise is context dependent. What it is for a person at the end of life to flourish is different from what it is for a person with PMLD to flourish and, accordingly, different metrics and assessments are appropriate. For the terminally ill person, flourishing may mean allowing them autonomy in their personal decision making about their remaining life and death, while for the person with PMLD, flourishing means reduction in self-harming or angry behaviour, engagement and relationship-formation with carers, and manifest positive subjective experience. An invariantist definition of flourishing must be determinate enough to pick out a coherent and consistent concept, and yet flexible enough that it can be multiply realised across different contexts. Defining flourishing in terms of pleasure, or in terms of autonomous action, for example, seems to fail to account for the variation exhibited, as either would preclude a large subset of putative instances of well-being from being characterised as such. However, the concept of flourishing cannot be so broad as to

offer nothing more than a synonym for well-being: it must offer enough substantive conceptual content to rule out variantism.

One limitation the invariantist seems to face in his definition of well-being is its consistency across different contexts. He can, certainly, adopt a conjunctive concept of well-being, taking it to be multidimensional, comprising several criteria which must all be fulfilled in order for an attribution of well-being to be made. Furthermore, he may be able to endorse a conjunctive concept of well-being such that the fulfilment of all conjuncts is not necessary for an attribution of well-being to be made. So, some conjuncts may remain unfulfilled when they are irrelevant to the particular context under consideration, and only when certain conditions are fulfilled do all of the various dimensions of well-being become relevant. In this sense, the invariantist might admit minimal context-dependence. However he cannot allow any attribution of well-being to be *inconsistent* with any other attribution of well-being. To do so would be to endorse a disjunctive concept of well-being, where mutually inconsistent definitions of well-being are appropriate for different contexts. This would undermine the uniqueness claim.

An account of well-being which might fit the bill is Martha Nussbaum's capabilities model (1988, 1992). Nussbaum develops a broadly perfectionist account of well-being, taking well-being to consist in the achievement of 'truly human functioning.' She specifies ten capabilities, which constitute the central elements of human functioning, fulfilment of which is necessary to attain a decent social minimum.<sup>14</sup> All of the ten

---

<sup>14</sup> Nussbaum sometimes characterises her account as an 'objective list account' (2008), but makes it clear elsewhere (2001) that the list is intended to be open to revision, with the members of the list being determined by their relevance to human functioning. In chapter 3, I argue that perfectionist theories of well-being should be understood as types of objective list theory, so I don't take her characterisation to be at all inconsistent.

components of functioning are of central importance, and they cannot be traded off against one another; going below the threshold with respect to one of the capabilities is a tragic loss, regardless of the extent to which the others are satisfied. Assessing whether someone has achieved truly human functioning requires assessment of all of the capabilities; looking at only one or a subset will give an incomplete picture. Nussbaum takes her list to be open-ended and open to contest, and it is also minimally context dependent: the capabilities are multiply realisable, which means that the components can be differently specified in accordance with particular circumstances.<sup>15</sup> Furthermore, the threshold which defines the decent social minimum for each capability is also context dependent.

Does the capabilities model account for all of the measures of well-being used across the clinical and social sciences? It seems plausible that Nussbaum might take many metrics to capture components of well-being, but not to measure of well-being *per se*. This is because, for Nussbaum, attribution of well-being requires knowledge of the degree of satisfaction of all of the capabilities. Measurement of one capability, for example bodily health, or of a subset of a capability, for example, disease specific health-related quality of life in a patient with bowel cancer, gives us only a partial evaluation of well-being. While the variantist would certainly agree that measurement of quality of life in bowel cancer patients does not tell us about their overall well-being, it is nonetheless a measure of well-being, which can, but need not, be used alongside a more comprehensive or general measure. The variantist will deny that general measures are, all things considered, more accurate or more conclusive measures of well-being than specific measures—though they may be more appropriate in particular contexts.

---

<sup>15</sup> Having said that her list is open-ended and open to contest, Nussbaum's specification of it has remained pretty much the same since the early 1990s (cf. Nussbaum 1992, 2000, 2011).

This may sound like a mere terminological distinction—Nussbaum takes the variety of measures of well-being to be contributors to well-being, whereas the variantist takes them to be measures of well-being—but it has consequences that are worth mentioning. For Nussbaum, trade-offs between capabilities which push one component below a baseline standard are always tragic and unjust, and failure to meet the standard should be conceived of as a loss. In the case of a terminally ill person at the end of life, shortening their life by withdrawing treatment would be, for Nussbaum, a *pro tanto* loss, even if they gain increased control over their environment, dignity, and integrity.<sup>16</sup> Nussbaum thinks in this way because, for her, the unit of well-being is a human life. The variantist, on the other hand, need not necessarily think of the trade-off as tragic or unjust, because they are able to take a fragment of human life, rather than a human life as a whole, as the unit of well-being. They can say that, given the end of life context, it is not a loss that the terminally ill person trades off the length of their life against other goods, even if they are in their 30s or 40s; rather, it is simply what constitutes well-being in that context. Nussbaum's approach does seem to be able to accommodate many of the variantist's demands regarding the context dependency of well-being. However, she does not take non-global measures of well-being at face value as measures of well-being, nor can she accept that the appropriate unit of well-being is context dependent.

But is Nussbaum's capabilities model really invariantist? The uniqueness claim entails that what enhances well-being in one context must also enhance well-being, or at least must not detract from well-being, in other contexts. There are two possible ways in

---

<sup>16</sup> If the person in question developed the illness in their old age, Nussbaum might argue that they have already reached a threshold for the capability to lead a healthy life of ordinary length, and so withdrawing treatment would not constitute a loss. However, if the terminally ill person is young, or in their middle age, this line of argument seems difficult for her to make.

which the capabilities model might fail to be invariantist. First, depending on the degree to which Nussbaum intends the capabilities to be differently specifiable in accordance with particular circumstances, the view may not satisfy uniqueness. If different cultural or circumstantial contexts can employ not just different but mutually incompatible interpretations of the capabilities, the theory looks likely to fail to conform to the requirements of unity. Secondly, through the—albeit tragic—admission of trade-offs between capabilities, the capabilities approach may allow that something that enhances well-being in one context detracts from it in another. For example, in the case of the terminally ill person, Nussbaum might accept that promoting their bodily health and extending their life may, all things considered, diminish their well-being. The trade-off may be considered tragic because the best possible functioning available to the patient involves a significant loss with respect to some of the typical contributors to well-being. However this seems to allow that something which typically contributes to well-being can sometimes diminish it and, consequently, should not only be excluded from the indicators of well-being in such contexts, but should be included in the indicators of diminished well-being. This goes beyond the minimal context dependence that may be admitted by invariantism, where something that is relevant to well-being in one context might be irrelevant to it in another.

Where does this leave invariantism? It looks as though the best attempt at constructing an invariantist theory of well-being which incorporates the many measures of well-being across the social, clinical and political sciences is on shaky ground, both with respect to whether it really takes those measures to be measures of well-being, and with respect to whether it remains invariantist or not. In order to remain strictly invariantist, the measures of well-being must be taken to be only partial constituents of well-being, and the idea that the unit of well-being can be anything other than a single human life must be rejected. Upholding the diversity of measures of well-being



seems to require trade-offs that permit different mutually inconsistent conceptions of well-being. This suggests that avoiding inconsistency will require the rejection of at least some putative measures of well-being as in fact capturing well-being.

#### *1.4 Defending invariantism II: Circumscribing well-being*

Given the difficulties with the first strategy, the invariantist may argue that many of the measures of life satisfaction, subjective psychological states, revealed preferences and health-related quality of life, for example, are not appropriately described as measures of well-being, but in fact as measures of other substantive concepts. Well-being should be kept conceptually distinct from other related or similar concepts. The invariantist may thus argue that in the four cases described in the previous section, what is in fact being measured in the anorexic person is ‘health,’ in the person with PMLD ‘happiness,’ in the person with a terminal disease ‘autonomy,’ and in the development economics survey ‘welfare.’ Only an invariantist account—of course the particulars of this will differ depending on which invariantist you ask—captures well-being *per se*. All other concepts should be treated as separate goods, which may also be valuable, and which may even contribute to well-being, or be necessary for it, but which are not identical to it (Alexandrova 2012a, p.637).

Alexandrova (2012a, p.630) argues that it is Procrustean to claim that the variety of ways in which well-being is measured are not in fact instances of well-being. To do so artificially constrains the ordinary usage of the concept of well-being outside of the philosophical debate. Insisting that a global, all-things-considered philosophical concept of well-being is the only true notion of well-being fails to acknowledge the way in which these concepts inter-relate and overlap, both linguistically and conceptually. It is difficult, impossible even, to extricate concepts such as quality of life, happiness, functioning, flourishing and welfare from well-being in a manner that is not to some degree arbitrary, impractical or implausibly restrictive. Well-being

invariantism thus requires us to give distinctive substantive content to each member of this family of concepts in such a way as fails to capture our ordinary linguistic usage.

Consideration of the disagreement between different invariantist accounts of well-being seems to support this. Particular theories of well-being are typically criticised on the basis that there is more to well-being than they suggest. Critiques of hedonist accounts of well-being, for example, often focus on the fact that there are some things which appear to be good for us, or make our lives go well, which do not necessarily involve pleasure—achieving or pursuing valuable ends, for example—and things which *do* involve pleasure which do not necessarily make our lives go well—sadistic enjoyment, for example, or opiate-induced stupor. Critiques of objective list accounts of well-being, on the other hand, may focus on the fact that they implausibly insist that someone's life is going well just in case they have sufficiently satisfied all of the objective requirements of well-being, even when they do not personally desire any of these things, or when they are very unhappy. Of any particular invariantist definition of well-being, we can ask 'Is this really a good life?' and invariably come up with other plausible instances of well-being which are not captured by the theory, or lives which are clearly devoid of well-being but which seem to be erroneously captured by the definition in question. Indeed, the chief way of critically assessing a conception of well-being seems to be to ask whether this indeed captures what we mean when we talk about well-being, or whether it fails to cover some dimension. This suggests that 'well-being' is a multiply realisable concept, which cannot be given the kind of universally applicable substantive content that the invariantist requires.

Fletcher (2009) agrees, suggesting that when we criticise accounts of well-being, we do so using other competing conceptions of well-being. No single theory of well-being successfully explains every case. This indicates that there is more to well-being than is captured by particular invariantist accounts. It also implies that there is some resource of intuitions or ideas about well-being which exists alongside the particular claims that

a given theory of well-being makes; that is, there must be some standard against which given philosophical accounts are judged. Stephen Campbell (2015) argues that substantive theories of well-being tend to be based on a ‘standard philosophical picture’ of well-being. However, this picture captures multiple substantive concepts, and describes a set of requirements which cannot be satisfied by a single concept. Campbell (2015, p.411) suggests that philosophers need to be more explicit about precisely what they take well-being to be in their particular theories, picking out specific features of the background account. This background picture is not, then, a single coherent theory of well-being, nor is it internally consistent. Rather, it is an indeterminate patchwork of intuitions and applications of well-being which informs our evaluation of particular examples and conceptions.

A background account of well-being, which subsumes a number of inconsistent substantive concepts, is consonant with the variantist picture. The concepts that Campbell (2015, p.403) takes to be present in the ‘standard philosophical picture’ are fairly high-level theoretical concepts—welfare, self-interest, advantage, quality of life, flourishing. Some of these constructs are standardly used in psychometric measures of well-being, though they are often broken down further into more readily measurable variables. Alexandrova’s account suggests that even this complex, internally inconsistent philosophical picture of well-being does not capture all there is to the ordinary language concept of well-being, which might also include many more context-specific constructs. In health-related contexts for example, a broader notion of quality of life might be broken down into a number of strands: pain, anxiety, mobility, self-care; and perhaps further still into particular symptoms and signs: breathlessness, ability to walk 100 metres unaided, number of anxious episodes in the past week. While such specific measures are unlikely to be part of a very high-level description of well-being, they are the features upon which a reasoned evaluation will depend when assessing whether someone is doing well or not in a particular context.

While such features do not exhaustively define well-being, and may only apply within a specific context, they should not be taken to be conceptually distinct from well-being: they are how well-being is best understood for a group of people, at a particular time, for a defined purpose. The variantist might agree that these are *also* measures of something else: health, emotional function, physical function. But, importantly, they need not thereby suppose that they are not measures of well-being.

Given this, there seem to be two possibilities for the invariantist: either to argue that the standard, background picture of well-being contains a number of errors about the nature of well-being, or to agree that the background picture of well-being should stand. If the latter is true then invariantism appears to be under threat, because it entails that no philosophical theory of well-being that purports to apply across all contexts exhaustively and uniquely captures instances of well-being. That is, no invariantist theory will ever fully capture well-being. The invariantist cannot, it seems, go down this route and remain invariantist. But if the former is true, an error theory is needed to explain why most people—bar a few fortunate or brilliant philosophers—fail to recognise the true meaning of well-being or the true scope of the concept of well-being, and are so liable to misattribute other concepts as well-being. Neither everyday nor scientific usage typically defines well-being in the invariantist philosophical sense, so some good reason is needed to explain the failure of non-philosophers to appreciate the ‘real’ nature of well-being (Alexandrova 2013, p.314). The invariantist can go about constructing his error theory in a number of ways. I consider three: first, he can criticise ordinary assumptions about well-being; secondly he can generalise from his intuitions about well-being, and thirdly, he can make normatively substantive claims about the concept of well-being.

First, the invariantist might critique ordinary assumptions about well-being. Take, for example, a rational desire theorist who claims that events occurring after the death of an individual can change an overall evaluation of their well-being, insofar as they affect

whether their rational desires have been satisfied. Such an account might also be committed to the claim that a person's well-being can be altered without them having any awareness of it and without any subjective psychological impact, if, unbeknownst to them, their rational desires are satisfied or frustrated. Their well-being can even change after their death. Let's assume that this commitment is ordinarily taken to be an implausible consequence of the view, and the desire theorist contests this.<sup>17</sup> In order to argue for his view, and to show why an inclination to abandon any view which has such implications is misguided, the desire theorist cannot, presumably, simply reiterate that as well-being is a matter of the satisfaction of rational desires, we should accept these implausible-seeming consequences. To do so is to beg the question about the nature of well-being—the very matter under consideration.

However, it's not clear what other resources he has to draw upon. Offering positive reasons for thinking that well-being is best understood as the satisfaction of rational desires will only take him so far. Insofar as such a defence does not eliminate the implausible aspects of the view, it only goes to show that well-being is *sometimes*, or *partly*, constituted by the satisfaction of rational desires. Offering reasons for thinking that the implausible consequence is not in fact implausible will have similar consequences. Suppose we agree that it does in some sense leave us worse off if our most significant goals are sabotaged, despite us not knowing about it, due to our death or dissociation. This need not entail that rational desire satisfaction exhausts well-being, or that it constitutes well-being in all contexts. Such arguments again only suggest that a rational desire account of well-being is plausible, not that it is universal.

---

<sup>17</sup> Indeed, we need not assume it—a form of this criticism of desire-satisfaction theories of well-being is oft-rehearsed, and is reproduced by Griffin (1988, pp. 16–7), Parfit (1984, p.494), and T.M. Scanlon (1998, p.114). Thomas Nagel (1979) makes a case for the view that one's life *can* be affected by events that happen outside of its mortal boundaries.

Secondly, the invariantist might take an intuition he holds about well-being and argue that it can be generalised across other cases. So, for example, we may agree that the fact that a suicidal person is caused great distress by being forcibly detained and medicated does not negatively contribute to her well-being in that context, despite her distress being characterised by unpleasant feelings. This may be used, by the invariantist, as the basis for arguing that well-being is not characterised by pleasure and other relevantly similar instances in which people experience pleasure (or displeasure) should not therefore be understood as instances of well-being (or its absence). So, he might argue that pleasure and displeasure should not be taken to contribute to well-being in other cases. But what happens if further unintuitive commitments result from this generalisation? Consider the forced feeding of a person with terminal illness who wants to stop all nutrition and hydration in order to bring about his death. Here, his distress, and subjective psychological experience in general, *does* seem to be part of what contributes to an assessment of his well-being—to say it is not relevant is counterintuitive. In such a case, the invariantist may try to make one of two claims. First, he may reassert the uniqueness claim: that if something is a contributor to well-being in one instance, it will always be a contributor to well-being wherever it appears.<sup>18</sup> In doing so he would declare, but not prove, invariantism. Or he may claim that in this case, the two cases are not relevantly similar, such that the same feature (subjective psychological experience, for example) is a different contributor to well-being in each case.

The second response is not open to the invariantist. Given the invariantist's commitment to securing a universal definition of well-being, which captures all and every instance of well-being, he seems to be committed to the view that if something

---

<sup>18</sup> Or perhaps the more general claim, which Kagan (1988) calls the *ubiquity thesis*, that if variation in a given factor makes a difference anywhere, it makes a difference everywhere.

is a contributor to well-being in one instance, it will always be a contributor to well-being wherever it appears. If something which contributes to well-being in one case fails to contribute to it in another case, this appears to undermine the invariantist project. As all invariantist conceptions of well-being are supposed to apply across all human lives, the relevant feature of cases that determine whether or not they are appropriate cases about which to evaluate well-being is whether they refer to human lives or not. This makes it very difficult for the invariantist to deal with conflicting intuitions when those intuitions seem to point towards different substantive conceptions of well-being. This is especially concerning for the invariantist if he relies on intuition to ascertain instances and determinants of well-being in the first place.

Finally—and perhaps this is the most plausible explanation—the invariantist might be making a substantive normative claim about the definition of well-being. He may acknowledge that well-being is used in a variety of ways, including ones that are inconsistent with an invariantist picture. However, well-being *ought not* to be understood thus. In giving reasons for this claim, the invariantist might argue that many of the definitions and metrics of well-being that are in use fail to capture what makes a human life go well, and therefore ought not be understood to be measures of well-being. This seems to entail that the goal of the invariantist approach is not to set out to define well-being and to determine what falls under it, but instead to determine what makes a human life go well, out of which will drop a plausible account of well-being. However such an approach also begs the question, presupposing the very aspects of invariantism that the well-being variantist seeks to challenge. First, it starts off with the assumption that well-being is something which is assessed over the course of a life, rather than over the course of a disease or an event, or when in a particular context, such as while employed in a given job or while living in a certain city. Secondly it assumes that the relevant unit of well-being measurement is a human life, rather than context-specific classifications—train driver, person with emphysema, citizen of

Mozambique, child. This way of thinking about well-being does not, then, help the invariantist respond to the variantist challenge, it merely restates the invariantist assumptions about well-being.

Constructing an error theory that offers a plausible critique of the large majority of ways of using and measuring the concept of well-being is something of a minefield. It's difficult to find convincing reasons that the invariantist can offer for rejecting a subset of technical and ordinary usage without presupposing some of the major assumptions of the invariantist approach. One of the major difficulties the invariantist faces is that his own justification of an invariantist definition of well-being also needs to depend on an appeal to an ordinary language concept of well-being. He can't outright reject ordinary language as a means of supplying content to a philosophical definition. Typically attempts are made to justify and criticise theories of well-being on the basis that they accord, or fail to accord, with ordinary usage, and the criticisms exchanged between different philosophical theories of well-being discussed above seem to depend on, and stem from, a perceived failure to accord with aspects of ordinary usage. The variantist has a somewhat easier task in this respect. As the well-being variantist starts off with an inclusive attitude towards everyday language about well-being, and the complexity of usage found there, she has significantly less explanatory work to do. As she endorses multiple substantive conceptions of well-being, the variantist takes well-being science on its own terms, as measuring well-being, rather than correlates or causes thereof, and does not have to pull off any magic tricks to show how apparently inconsistent conceptions of well-being in fact reduce to a common or master value.

### *1.5 Stance independence and stance dependence*

These arguments provide some grounds for adopting a variantist account of well-being. But so far I've said relatively little about what a variantist account of well-being might look like—what defence I have given has been in virtue of the shortcomings of



invariantism. Alexandrova acknowledges that well-being variantism can also be understood not just as a methodological thesis, but as a substantive theory of prudential value (2012a, p.628). That is, well-being variantism can be construed as a critique of the notion of a unified theory of well-being, entailing that well-being should be understood to be conceptually plural. I seek to support this stronger, metaethical claim about well-being, by developing a plausible pluralist account of well-being.

Pluralism is not a concept that is used with much consistency across different areas of philosophy. Political pluralists take there to be multiple irreducible and perhaps inconsistent conceptions of the good, which are taken to be equally true, or which ought to be equally tolerated (Rawls 1972, 2005; Berlin 2013). Value pluralists think there to be multiple fundamental and authoritative values, which are in conflict, or apparently in conflict, and which need to be weighed up against one another in practical reasoning and decision-making contexts (Ross 1930; Williams 1981; Thomson 1997; Wiggins 1998). Some accounts of scientific pluralism take there to be a plurality of classifications of reality into kinds, and for reality to be non-reductive between different levels of description (Dupré 1993), whereas others take it to consist in the non-universality of the laws of physics (Cartwright 1999). Metaphysical pluralism is the view that that there can be more than one experiential perspective, and that there is no unique set of concepts we use to think about the world, coupled with the idea that our knowledge of the world is in some sense constructed by our conceptual scheme (Lynch 1998). There is relatively little to connect all of these views except for the fact that their being 'pluralist' refers to the idea that there are two or more of some entity in play concurrently—conceptions of the good, values, classifications of reality, systems of physical laws, conceptual schemes—which cannot be reduced to one another or to a common entity.

Well-being variantism clearly expresses a form of well-being pluralism in some sense, taking there to be multiple, irreducible accounts of well-being. However, there are a

number of different ways in which a theory of well-being might be pluralist, and not all of them are variantist. For instance, well-being might be made up of multiple, irreducible components or people might, additionally, give different weights to the various components of well-being (Qizilbash 1997). If well-being is made up of multiple, irreducible components, these might disjunctively or conjunctively constitute well-being. That is, well-being might comprise the simultaneous realisation of all of these components, or, it might comprise the realisation of one, or at least one, component. The components of well-being, or the core components thereof, may be taken to be a fixed and closed list, or not. For example, the perfectionist views of Griffin and Nussbaum both take ‘the ends of life’ (Griffin 1988, p.70), or ‘truly human’ functioning (Nussbaum 1988, p.181) to be captured by a discrete set of values. While it is the concept of functioning or flourishing which captures well-being for perfectionists, with the list of values being only instrumentally valuable with respect to this end, both Nussbaum and Griffin take the lists they respectively develop to be pretty much final. Whereas Nussbaum (1992, p.222) is reluctant to allow for any trade-offs between values on her list, taking them to be incommensurable, Griffin takes trade-offs between prudential goods to be an inevitable part of decision making about well-being.<sup>19</sup> For Griffin (1988, p.72), different goods are more or less important to achieving well-being in different contexts, and not all of the components of a valuable human life need be simultaneously satisfied in order for someone to have reached an acceptable standard of well-being. Some objective list theories of well-being (see, for example Finnis 1980 and Fletcher 2013) take a fixed set of entities to be non-instrumentally constitutive of well-being. Such views are paradigmatic examples of

---

<sup>19</sup> See Griffin (1988, part II) for his discussion of incommensurability. Griffin does accept limited incommensurability about well-being when it comes to diminishing marginal gains—that is, in some cases, if we have enough of one good, B, any amount of another good, A, will outrank any additional amount of B (1988, p.85).

pluralist accounts of well-being with fixed components. However, they too may allow for trade-offs between goods, and again, need not require all of the goods to be simultaneously fulfilled. On other views, the core components of well-being are not taken to be fixed. Amartya Sen (1987), for example, takes the components of well-being to be open-ended, requiring a valuation exercise to determine the importance and relative weight of the components of well-being for particular individuals and in particular contexts.

I will discuss Sen's nuanced view in some detail in chapter 3 but, for now, he makes a useful distinction which helps to illustrate the forms of pluralism that are relevant to the variantist (1987, pp.2–3). Sen discusses two forms of plurality with respect to the 'standard of living': constitutive plurality and competitive plurality. Constitutive plurality involves seeing the standard of living as a 'basket of multiple attributes' which are irreducible. This is compatible with an overall standard of living index being calculated for the basket as a whole. Competitive plurality involves a choice between multiple alternative baskets, which may or may not contain multiple attributes. This distinction also seems to apply to well-being, though Sen understands the standard of living to be a narrower concept than well-being.<sup>20</sup> In the context of well-being, constitutive plurality would be the idea that well-being is made up of multiple attributes and competitive plurality the idea that there are multiple alternative conceptions of well-being to choose between.

Clearly well-being invariantism is compatible with constitutive plurality, as an invariantist definition of well-being can comprise a set of individually necessary and jointly sufficient criteria. However invariantism is incompatible with competitive

---

<sup>20</sup> Sen takes well-being to be influenced by sources other than the nature of one's own life, for example, the achievements and failures of loved ones, whereas the standard of living reflects only features of one's own life (1987, pp.26–7).

plurality. Variantism, on the other hand, is committed to competitive plurality about well-being. However, it is possible that not all variantist accounts of well-being are constitutively plural—there could be some appropriate contexts for using constitutively singular accounts of well-being, though of course these will likely exist in competitive plurality with constitutively plural accounts of well-being. Well-being variantism implies not just that well-being is made up of a series different components, or that these components can be differently weighted by different people, though it most likely accepts both of these claims. Variantism makes the further claim that well-being comprises different sets of components in different contexts. Henceforth, when I talk about a ‘pluralist theory of well-being’ or ‘well-being pluralism,’ I refer to competitive, rather than constitutive plurality. Such a pluralist theory of well-being is needed by the well-being variantist to justify their claim that we should treat the variety of measures and uses of well-being in the clinical, social and political sciences as instances of well-being. Nonetheless, though variantism implies competitive plurality, there are different forms that this can take.

There are two kinds of pluralism that are ostensibly consistent with well-being variantism: *stance independent pluralism* and *stance dependent pluralism*. In metaethics, stance independence is the view that there are moral truths that are independent of any given perspective on them, and that are made true by something other than people’s attitudes towards them (Shafer-Landau 2005, p.15). This does not necessarily mean that moral truths are not in some sense anthropocentric, insofar as they are defined relative to human concerns and activity, but it does entail that moral truths are not response dependent, that is, that they do not exist only as dispositions to elicit particular responses (Dancy 1986). In contrast, stance dependence refers to the view that moral truth is fundamentally anthropocentric, and defined in terms of the responses and views of agents. While this does not preclude the existence of a moral reality, it takes it to be dependent on the minds of moral agents, and thus socially

or individually constructed. Just like financial value, contracts, and social status, stance dependent moral concepts and properties are real, but entirely dependent on not just human interests, but human responses and attitudes. Just as it doesn't make sense to talk of the value of the pound or a king outside of a very particular set of attitudes and attitude-dependent social institutions, for a stance dependent account of morality it doesn't make sense to talk about overall moral concepts like right and wrong, or more specific moral concepts like obscenity, kindness, corruption, and honesty, outside of a specific response-dependent context.

Stance *independent* pluralism about well-being would entail that what it is to do well in a particular instance is not merely dependent on what particular individuals believe it means to do well in that context. Well-being may be dependent on human interests in a broader sense—what kind of beings we are, our physical and psychological capacities, our life cycle and mortality—but it is not fixed by the responses of particular people. A stance independent account of well-being is likely to take accounts of well-being to be fixed separately from attempts to identify and apply them in particular cases. The perfectionist and pluralist objective list theories mentioned above are likely to be best classified as stance independent pluralist accounts. Many such theories take well-being to be multiply realisable. Even accounts like Nussbaum's, which are resistant to trade-offs between the capabilities, acknowledge that the capabilities can be differently realised, and so that different things can constitute well-being in different contexts. What determines the content of well-being in different contexts on such accounts is something that is independent of people's attitudes—for perfectionists it is defined in terms of the fulfilment of one's characteristically human capacities, or species-typical flourishing. For non-perfectionist objective list theorists, it is defined in terms of some specified attitude-independent set of goods. Assessments of well-being in particular cases are thus made relative to an external standard—is this

person fulfilling their characteristically human capacities, does this person have any of these goods in their life.

Stance independent pluralism about well-being would not preclude subjective conceptions of well-being, or rule out theories of well-being that require self-report or assessment of subjective psychological states. Rather, for the stance independent pluralist, the facts that determine the substantive conception of well-being that is appropriate in a given instance are not established by human interests and attitudes. So on this account it may be that, in a given context, well-being consists in a measure of affect or subjective psychological state. Of course this can only be measured by subjective assessment. But such a view remains stance independent if evaluating well-being consists in an assessment of subjective psychological states regardless of what anyone thinks. It is not the fact that evidence for the level and valence of well-being in a particular case must be determined independently from any subject that makes a theory stance independent, but the fact that the conception of well-being itself is determined independently. Stance independent pluralism about well-being will take what it is to do well in any context to be independent of human responses and attitudes, and so will require that possible determinants of well-being are fixed, or at least fixable, independently of the assessment of particular cases. Of course, as the world changes and new technologies and social structures create new contexts for well-being, new accounts of well-being might emerge, but at any point there will be a non-negotiable set of possible conceptions of well-being, where their appropriateness in given contexts is fixed.

Stance *dependent* pluralism about well-being would entail that what it is to do well in a given context is dependent not just on a broad account of human interests, but also on the specific responses of a relevant group of people. This person or group need not be the group to whom the conception of well-being pertains, but might be, for example, a concerned friend, parents, a clinician, policy makers, the voting population,

a linguistic community. Stance dependent pluralism about well-being would not, then, take the appropriate accounts of well-being to be fixed independently of any given perspective on them, and will take particular concepts of well-being to come into existence only in the context of attempts to define and evaluate it. On this approach, appropriate accounts of well-being are determined relative to human interests and attitudes. This does not mean that people can assign well-being to people or decide which theory of well-being is appropriate for a given context arbitrarily, but rather that the appropriate theory, measure and assessment of well-being is not determined by attitude-independent facts. This is not to say that facts about people play no role in assessments of well-being, but rather that such facts justify assessments of well-being only in light of particular people's attitudes and purposes. For the stance dependent pluralist, facts about well-being are only about well-being in virtue of particular people's attitudes, whereas for the stance independent pluralist facts about well-being are about well-being independently of anyone's attitude.

Again, stance dependence does not entail that all theories of well-being must be subjective: stance dependence is a feature of the conception of well-being, not its content. So an appropriate conception of well-being in a given case might be an objective goods account, where assessments of well-being are not reliant on subjective assessments. However such an account would remain stance dependent insofar as the fact that this is an appropriate conception of well-being for this context is dependent on the particular attitudes of some person or group of persons. There are not many existing accounts of well-being which are best described as stance dependent pluralist. I take Sen's capability theory to be more or less a stance dependent pluralist account of well-being. Sen himself claims that the variety of perspectival accounts of well-being that his view endorses are objective, a suggestion which I dispute in chapter 3. Alexandrova's well-being contextualism (2013), which I discuss in chapter 6, comes

pretty close to stance dependent pluralism, though she too takes the content of well-being claims to be grounded in objective facts about the context of their speaker.

Stance independent and stance dependent accounts of well-being will offer different explanations, and different *kinds* of explanations, of how prospective concepts of well-being are to be determined in a given context. For stance independent pluralists, the explanation will likely focus on how we can come to know context-specific accounts of well-being, and will require some account of the mechanisms for attaining epistemic access to facts about well-being. How do we know what constitutes human flourishing or functioning? How do we know that *these* goods constitute well-being and not some other set of goods? For stance dependent pluralists, the explanation will focus on who determines the accounts of well-being, and via what mechanism. There is a spectrum of possible responses to this question. A more objective approach might depend on an account of idealised human attitudes and responses to determine the appropriate way of conceiving of well-being in a given context. Various moral theories, usually in the Kantian tradition, take moral truth to be constructed from idealised human attitudes, that is, the views of people who seek to identify moral principles which no one could reasonably reject (Scanlon 1998), or of parties behind a veil of ignorance (Rawls 1972), or from some other informed and rational perspective (Shafer-Landau 2005, pp.39–40). On such idealised accounts, individual beliefs and actions can be criticised from the ideal standpoint, enabling characterisation of individuals' moral principles and actions as false, or wrong. In the context of well-being, such an ideal account might claim, for example, that *laundered* or *rational* preferences offer a more accurate picture of well-being than crude preferences.

By contrast, a more subjective approach might take appropriate conceptions of well-being truth to be fixed by people's actual attitudes. In a metaethical context, this might be understood as the view that there is no ideal standpoint from which to assess moral claims. Criticism of people's attitudes can only take place from the perspective of an



agreed moral code, where this might be represented by a codified set of laws, a historical tradition which remains unchallenged, or the majority view. Normative truths—or justified normative claims—are thus subject to change over time. In the context of well-being, a similar view might arise, that there is no ideal set of preferences, and no way of determining an ideal account of what it is to do well in any context. This must be decided by appeal to the relevant set of persons, or using an agreed decision-making procedure. In some cases this might be a matter of self-report and individual decision making, in other cases it might be a matter of community-wide agreement.

Between these two points there are a number of more conciliatory approaches. Some more extreme objective stance dependent theories might take the ideal human perspective to be a self-evident fact, or something which can be determined through a good understanding of what it is to be human. Such accounts might end up looking rather more stance independent than stance dependent—that is, they might end up providing something like a general account of human rationality which is taken to determine the meaning of well-being in particular contexts, independently of any actual human attitudes. However, a less radical approach might think that the ideal human perspective is a matter of community assent, and cannot be determined without at least some consideration of people’s actual attitudes, traits and practices. More objective accounts will likely be much more strict about how well-being can be characterised, and may have more stable application of conceptions of well-being over time. More subjective approaches will be liable to change over time, as the attitudes and goals of those measuring and evaluating well-being change, and will not set so many constraints on appropriate conceptions.

### *1.6 Invariantism, variantism, and the philosophical schism*

My thesis begins to build a case for a stance dependent pluralist conception of well-being. I argue that conceiving of well-being as stance dependent better makes sense of

attempts to define and measure well-being. First, though, a brief discussion about the schism between invariantism and variantism is warranted. I certainly don't expect any well-being invariantist to raise a white flag and admit defeat at this point. The philosophical disagreement between the invariantist and the variantist runs deep, and hangs on a profound disagreement about the nature and application of the concept of well-being, and, indeed, about the role of philosophers, and the relationship between philosophy and other disciplines. Understanding this schism will provide some indication as to how an argument for a pluralist account of well-being might best proceed.

The conceptual disagreement between the variantist and the invariantist about well-being raises a more general, and deeply contentious, question about whether logic reflects or regiments language—at least with respect to the concept of well-being. In 1.1, I suggested that invariantist theories of well-being lean on an *a priori* Platonic or Russellian view of concepts and their application. For the well-being variantist, well-being is not a classical concept in this sense. The very aim of capturing the *practice* of well-being in an account of the concept of well-being speaks to a different understanding of concepts altogether. One way of understanding the variantist concept of well-being is as an indeterminate concept which requires contextual specification for meaningful usage. Wilson (2006) critiques the classical picture, arguing that in fact examination of concrete circumstances typically demonstrates the incompleteness or contradictoriness of our concepts, and raises questions about how we should proceed with classification which cannot be resolved simply by reference to the concept itself. He suggests that understanding and applying concepts requires us to recognise that our understanding of phenomena is not fixed, and needs to be monitored and corrected as our standards of adequate understanding change. Furthermore, mastering a concept sometimes requires significant shifts according to context. Concepts, for Wilson, are rough around the edges, and in constant need of

policing, renegotiating and specifying. He argues that we commonly, and mistakenly, expect concepts to carry great evaluative burdens but not to buckle under weight. Successful use of concepts requires a modest expectation of the scope of any particular application thereof.

Another way of understanding the variantist concept of well-being is as a family resemblance concept (Wittgenstein 1989, §65ff). A family resemblance concept is one where there is no single characteristic or set of features shared across all of its instances, such that it cannot be given a single, universal definition (Forster 2010, p.69). Wittgenstein's central example of this is the concept *game*: different examples of games have different features, some of which overlap or 'criss-cross.' This means that two instances of games can share no characteristics, and yet still both be appropriately characterised as games. Nancy Cartwright and Rosa Runhardt (2014) argue that well-being is best understood—along with many other concepts in the social sciences—as a family resemblance or *Ballung* concept.

While Wilson's indeterminacy view chiefly contests the invariantist view that the definition of well-being provides necessary and sufficient conditions for its own application, the family resemblance view contests the idea that well-being must pick out a single shared characteristic across all instances. Of course, one can also take concepts to be both indeterminate *and* family resemblance concepts—indeed, Wittgenstein seems to think that at least some family resemblance concepts are not expressible in terms of necessary and sufficient conditions for application and are therefore indeterminate in extension (Wittgenstein 1989, §68–69; Foster 2010, pp.76–7). I will not argue for either, or both, of these views here. I only flag them up as a way of illustrating that the starting point of the variantist, which is broadly practice-inclusive, requires a different understanding of the relationship between the concept of well-being and its linguistic application than that implied by the starting point of the invariantist.

Crucially, the variantist and the invariantist see the different ways in which well-being is *used* as playing a different role vis-à-vis the concept itself. For the invariantist, instances of well-being can be adjudicated against the definition of the concept to determine whether they are appropriately designated 'well-being' or not. The definition regulates appropriate usage. For the variantist, on the other hand, the different instances of well-being tell us something about the concept of well-being. Usage informs the meaning of the concept. This is particularly true of stance dependent pluralism, but may also be true of stance independent pluralism if, though particular conceptions of well-being are determined by facts about a given context and case, there is a certain amount of convention which groups separate instances of well-being together *as* well-being. For the well-being invariantist, insofar as the concept of well-being is appropriately interpreted differently in different contexts, this is understood as simply a matter of pragmatics, where the utterance has context-specific significance, but the underlying concept remains stable. However, for the well-being variantist, if there is stable variation in the use of the concept of well-being across different contexts, this tells us not just something about the way in which context affects meaning, but something about the concept itself.

This disagreement about the nature of concepts reflects a further disagreement about the role of the philosopher, and the relationship between moral philosophy and other practices. For the invariantist, the moral philosopher is in a position to look at linguistic and social practices from an external perspective, and pass judgement on them. The philosopher critiques the way in which people use normative concepts and the evaluative claims that they make, in both formal disciplinary settings and in ordinary, informal contexts, and she does so with reference to objective moral principles. On the other hand, the variantist approach, and certainly stance dependent pluralist accounts, takes human practices on their own terms, as a starting point to which philosophy is answerable. The philosopher does not have a properly external

perspective on other practices, because she is herself embedded in the shared linguistic and social practices that she seeks to examine. So, what exactly is the job of moral philosophers on this account? Is there any work to be done if our normative practices are not to be policed and to have rules imposed upon them? My inclination is that moral philosophers serve as what Philip Kitcher (2012) calls *philosophical midwives*, that is, they help to identify and provide resources for thinking about tensions and problems that emerge in our current practices, and guide discussion which might lead to solutions. They are problem-solvers, rather than diviners of truth. Furthermore, any solutions that emerge ultimately have to arise at the level of practice, rather than be dictated by a cadre of scholars.

Those views that I have characterised as instances of stance *independent* pluralism occupy a space somewhere between these two positions. Typically, perfectionists and objective list theorists about well-being take the reality of human life and experience as their starting point, and are clear that our concept of well-being should reflect the kind of being that we are and the kind of things that we do (Nussbaum 1988, pp.150–1; Foot 2001, p.51; Fletcher 2015, p.152). But it is only some aspects of human experience and activity that are taken to tell us about well-being. For the perfectionist it is only those beings and doings that are ‘so important that we will not count a life as a human life without them’ that constitute what is good for us (Nussbaum 1992, p.208). Some kind of objectivity about the relationship between human lives and prudential value is sought, via the discernment of those things that are necessary to our lives *qua* human lives. And for the objective list theorist it is not all things that are valuable or that are taken to be valuable that constitute the prudential good, but only those goods that have objective value, or attitude independent value (Fletcher 2015). So there is a move here towards understanding prudential value in terms of actual human practices, but where not all actual human practices are taken to be determinative of well-being. The philosopher on such views adopts an attitude of what

Nussbaum calls *historically grounded empirical essentialism* or *internalist essentialism* (1992, p.208). Philosophy does not offer an external perspective on human practices and, indeed, to pursue such an end is a deeply confused endeavour. However, from an internal perspective it is nonetheless deemed possible to discover essential properties of human lives, and to provide a fairly determinate account of our good. The philosopher, then, for the stance independent pluralist, finds the necessary in the contingent.

Characterising the disagreement between variantism and invariantism about well-being thus has important implications for the way in which invariantism can be criticised. The philosophical chasm between the invariantist and the variantist about well-being is unlikely to be resolved by a few counter-examples and a tentative critique. Faced with the discussion in this chapter, the invariantist may just bite the bullet and argue that all other alleged applications and definitions of well-being really are wrong, and that the true concept of well-being can be determined by clear thinking alone, rather than by examination of its varied uses in ordinary language, and the clinical and social sciences. Even if it can be empirically shown—using factor analysis, for example (Alexandrova and Haybron 2016, pp.1099–100)—that well-being is generally understood as a family resemblance concept, the invariantist might nonetheless argue that this is the wrong way to think about well-being, and we should change our linguistic practice. The kinds of philosophical and normative claims to which invariantism and variantism are committed are not ones that are easily dislodged.<sup>21</sup>

Thus, while my thesis develops and defends a theory of well-being which embraces the variability that invariantism rejects, I do not hope to deliver a death-blow to the

---

<sup>21</sup> I think more can be said about the approach taken by perfectionists and objective list theorists. In chapter 3 I argue that the prospects for such views remaining objective, while at the same time offering credible and sufficiently rich accounts of well-being, are slight.

invariantist. I develop an account out of a close scrutiny of some of the idiosyncrasies and problems which typically characterise the measurement of well-being. I think, and argue, that my account makes better sense of the measurement and invocation of well-being than invariantist and stance independent pluralist views, but it can only do so in virtue of first adopting an ecumenical attitude towards those measures and invocations *qua* measures and invocations of *well-being*, that is, by adopting variantism as a methodological thesis. This approach is unlikely to convince committed invariantists, beginning as it does from a starting point that they can't accept—that an account of well-being should endeavour to explain and include the variety of conceptions and metrics of well-being in use outside of a strictly philosophical framework. What such an account can do, however, is to offer a plausible alternative to the many invariantist theories of well-being. Developing an alternative view not only indicates how philosophers might expand their horizons in thinking about prudential value, but also challenges the invariantist philosophers' established turf.

## CHAPTER 2

### Well-being and measurement

The variety of concepts of well-being that are used across the clinical and social sciences, and which appear to draw on a variety of theories of well-being, are almost always developed in order to *measure* well-being. The measurement of well-being provides grounds for decision making which takes the well-being of particular people or groups of people into account—choosing between prospective public policies, for instance, or determining whether a pilot study has been successful, or evaluating whether a novel pharmaceutical therapy delivers enough additional benefit to make it worth its higher cost. Measurement implies a certain degree of consistency in the thing that is being measured. So, if I measure your well-being at two different times, and get different measurement indications at each time point, this should indicate that some change has in fact occurred in your well-being, and not that there has been a change in the measuring instrument, or that I am measuring a different attribute. And similarly, if I measure the well-being of two different people, using the same measuring instrument, and again two different measurement indications result, this should indicate some relevant difference in their well-being, where the same thing is being measured in each of them. The measuring instrument is taken to measure the same thing across measuring objects and across time, within certain constraints. Variantism about well-being implies that in different contexts, different measures of well-being are appropriate, and so this kind of consistency should only be expected within contexts, and not across them.



In this chapter, I argue that the measurement and measurability of well-being is not something that is incidental to its conceptualisation and definition, nor something which is relevant to well-being only in particular applied contexts. Instead, I argue that the identification of well-being in people is central to any theory of well-being, and that this makes measurement central to theories of well-being. By the *identification of well-being* I simply mean determining whether someone is doing well or not. In 2.1, I argue that any theory of well-being must say something about how to identify well-being in people. Most theories of well-being will also have to say something about how to compare well-being across people and across time. In 2.2, I propose that the identification of well-being in individuals amounts to a form of measurement. This means that any theory of well-being must be a theory of the measurement of well-being. This may sound like a controversial claim. However, I find my way there not via a contracted definition of well-being, but via an inclusive view of measurement. This, I contend, is not so controversial after all. In 2.3, I outline the difference between *realism* and *instrumentalism* about measurement—two ways of understanding the relationship between measurement operations and the objects of measurement. In 2.4, I suggest that the invariantist must be a realist about well-being, that is, he must maintain that the property or state or object that is picked out by well-being exists, and has a certain quality or magnitude, independently of attempts to measure it. On the other hand, the well-being variantist will be either instrumentalist or realist about well-being, depending on whether they adopt a stance dependent or stance independent approach. Instrumentalism about well-being implies that the measurement of well-being is dependent on an object being represented in such a way as to enable its interpretation as an object of *measurement*. Finally, in 2.5, I argue that thinking of well-being as something which is measured foregrounds the importance of the ontology of well-being in our theorising about it, that is, the importance of the nature of the kind of thing that we pick out when we talk about well-being. It need not amount

to a claim about the nature of well-being, rather it is a way of framing well-being which enables us to think about its nature in illuminating ways.

### 2.1 *The scope of theories of well-being*

For something to be a theory of well-being it must not only describe what well-being consists in, but also say something about how to identify it. That is, it must be able to tell us how to determine whether someone is doing well or not. I'll call this the *identification thesis*. If a theory of well-being identifies well-being with pleasure, for example, it will inevitably have something to say about what pleasure is and what it isn't, and why it is that pleasure exhausts well-being—in this particular context, or in general—rather than something else. But unless it also says something about how to determine whether someone has pleasure in their life, it will remain incomplete. The well-being theorist might try to deny the identification thesis, claiming that all he must do is say something about what well-being is at the *general* level. Anything further is supererogatory.

There seem to me to be two possible routes that the well-being theorist might take in justifying this line of argument. First, he might agree that there is more philosophical work to be done in determining how to apply his theory, but insist that this is not part of *his* job—it's not the job of the well-being theorist, but a task for the *well-being identification theorist*, or the *well-being\** theorist, or some other philosopher. This seems pretty implausible. Insofar as there is philosophical work to be done concerning well-being, it's reasonable to think that it's within the well-being theorist's remit to do it, and furthermore to judge his theory to be incomplete to the extent that he fails to do so. Again, if someone argues that well-being is pleasure, but refrains from saying anything about how to characterise pleasure, then it's simply not clear what he means by this claim. His theory could be entirely subjective, such that pleasure is always and only characterised and identified from a first-person perspective, and that only the person in question can assess what brings them pleasure and whether they are

currently experiencing it. He might insist that structured surveys, designed by specialists, are needed to assess whether someone is really experiencing pleasure. He might take pleasure to be constrained by objective principles, such that, although it is typically characterised and identified from a first-person perspective, sadistic, perverted and wrongful pleasures don't contribute to well-being, even if the person in question claims to find them pleasurable. Or he might intend an entirely objective theory of pleasure, such that the presence of pleasure is determined by the measurement of brain states, rather than subjective reports. Finally, he might think that measuring pleasure directly is a fool's game, and we can only approximate it by eliciting preferences. These different characterisations of pleasure pull in quite different directions, not just in their detail, but with respect to their broad conceptual and theoretical substance. There is clearly much more work to be done in characterising well-being, and it isn't peripheral business. Explaining how to go about identifying well-being in individuals is part of the core work of a well-being theorist.

Alternatively, then, the well-being theorist might deny that there is any more philosophical work to be done at all once the general account of well-being has been provided—determining whether someone is doing well or not is a matter for psychologists, psychometricians, or neurologists, not philosophers. Again, this seems like a pretty difficult position for the well-being theorist to defend. Even those well-being theorists who think that the primary business of moral philosophers is to secure *a priori* principles are likely to think of the alternative characterisations of pleasure just mentioned as distinctively philosophical views, and identifying and justifying them is distinctively philosophical work. Of course, it may be right that identifying and measuring well-being isn't always a matter for philosophers—if, for example, well-being is best captured by measuring brain states directly, designing the tools for determining how well a particular individual is doing might well be within the remit of a neurologist. But the well-being theorist still needs to provide justification for his

favoured approach to identifying well-being. He can't go straight from the claim that well-being is pleasure to the identification of observations of electrical activity in the left frontal cortex with well-being. In-between, he needs to explain that—and why—pleasure is best understood in terms of brain activity, and not in terms of subjective reports, for instance. The well-being theorist thus has a role to play in the specification and justification of the theoretical framework which other well-being practitioners work with, and this framework includes specification of how to identify well-being in individuals.

So a well-being theorist needs to say *something* about how to identify well-being. A failure to do so is a failure to specify meaningful individuation conditions for well-being. Furthermore, a theory of well-being will typically tell us how to determine which of two people, or two states of the same person, has greater well-being. I say 'typically' because there may be some accounts of well-being which deny the possibility of inter- and intrapersonal comparisons of well-being. For example, take a theory which identifies well-being with the satisfaction of desires, where there is taken to be no interpersonal standard for determining the degree of satisfaction, and very limited scope for quantifying the strength of different desires and aggregating across them. Perhaps instead the standard for assessment is the subjective appraisal of the person in question—their desires are satisfied insofar as they judge this to be the case, and they are doing well to the extent that they so judge. Such a theory would explicitly preclude the possibility of interpersonal comparisons of well-being. However most theories of well-being *do* take interpersonal comparisons of well-being to be possible, at least in some cases, and insofar as this is part of a theory, something needs to be said about how it is done.

On some theories, interpersonal comparison will follow fairly easily from the identification of well-being in individuals. Where well-being is understood in terms of the achievement of functionings or goods that can be measured against a non-

subjective standard, such as knowledge, political participation, or bodily health, comparing the well-being of two people or two states of the same person will be fairly easily determined by considering their respective levels of these goods. The difficulty in these cases will sit with the identification of well-being in individuals—that is, in determining which functionings are constitutive of well-being, determining whether or not particular individuals have achieved these functionings or goods, or have the capability to achieve them, and working out the appropriate information bases for each functioning.<sup>22</sup> If the same standards can be identified and used to measure objective well-being in different individuals, then the task of comparing their well-being will be comparatively easy.

However, on other theories interpersonal comparison will pose problems as numerous and as great as posed by identification. If well-being is understood to comprise goods which are at least partly subjectively-determined, such as happiness, friendship and self-respect, or if well-being is understood in terms of desires or preferences, then even once some method has been devised for identifying these goods, more needs to be said about the mechanism and logic of interpersonal comparability. For example, suppose that well-being is understood in terms of self-reported life satisfaction, and that an acceptable life-satisfaction tool has been constructed. In order to compare well-being across individuals using the tool, there must be reason to think that everyone is using the tool in the same way—that when I report that my life satisfaction is 0.7 I mean the same thing as you do when you report that your life satisfaction is 0.7. Without grounds for this, interpersonal comparison remains a pipe dream. On such theories of well-being, insofar as they seek to retain the possibility of interpersonal and

---

<sup>22</sup> Chapter 3, which takes as its subject those theories which identify well-being with objective goods, includes an extended discussion of the difficulties involved in defining, identifying and measuring goods and functionings.

intertemporal comparisons of well-being, some explanation is needed to show how this proceeds and how it is possible. Lacking this, the theory remains at best incomplete, and at worst indefensible.

## 2.2 *Well-being and measurement*

If a theory of well-being needs to say something about how to identify well-being and how to compare well-being across people, time and hypothetical conditions, this involves saying something about how to measure well-being. Well-being is measurable and, moreover, even fairly informal frameworks for the identification of well-being should be understood to be measures of well-being. Theories of well-being, then, specify measurement operations. This may appear to be a pretty controversial claim—other philosophers of well-being argue that well-being isn't measurable *at all* (Hausman 2015, 2017), and the argument that well-being can be measured is hard-won (Alexandrova 2017a). To see why well-being ought to be taken to be a measurable quality or quantity, a little needs to be said about the nature of measurement. In this section, I argue that nominal and ordinal measurement scales should be taken to be genuine instances of measurement. This inclusive view of measurement does not limit measurement to quantifiable properties; it also allows that the identification of qualitative and unordered properties can constitute measurement. This means that even fairly minimal identification of well-being in persons counts as a form of measurement.

Contemporary measurement theory takes measurement to be a matter of expressing the relations among objects using relations among numbers or other mathematical entities (Tal 2013, p1163). A measurement scale maps relations between objects and mathematical entities, such that relations among the objects of measurement are reflected in empirically significant features of mathematical entities. So, for example, the 'greater than' relation is empirically significant in measuring both length and mass, and reflects, respectively, the 'longer than' and 'heavier than' relations among objects.

Different measurement theories take different positions on the nature of the objects of measurement, and on the appropriate empirical standards for establishing the measurability of attributes and the accuracy of measurements. Some theorists argue that there is more to measurement than the construction of measurement scales; measurement also involves the theoretical and statistical representation of the objects of measurement operations and their data (Mari 1996; Tal 2013, p.1164).

In a seminal paper, published in 1946, Stanley Smith Stevens characterised four types of measurement scale. His taxonomy remains widely used today. His classification groups scales of measurement according to their mathematical group structure, that is, according to what kinds of transformations leave the scale-form invariant (Stevens 1946, p.678). *Nominal* measurement scales classify objects into categories that have no particular order, such as animalia/plantae/fungi/protista/monera (the five standard biological kingdoms) or female/male/intersex, or the numbering of players in a football team. *Ordinal* measurement scales put objects into a ranking, without capturing any information about the intervals between objects. Ordinal scales include Likert scales, which capture people's attitudes via ordered responses such as strongly agree/agree/neither-agree-nor-disagree/disagree/strongly disagree. *Interval* measurement scales capture equality and inequality among intervals, using an arbitrary zero point. Temperature on Fahrenheit and Celsius scales, for example, are examples of interval measurement: the intervals between measurements are non-arbitrary, but the zero-point in both cases is merely conventional, and doesn't indicate the absence of temperature. Finally, *ratio* measurement scales capture equality and inequality among intervals, using a non-arbitrary zero point. Temperature in Kelvin and length are ratio measurement scales, because their zero-points are determined by some non-arbitrary fact—complete absence of molecular motion, in the case of Kelvin, and absence of extension, in the case of length.

A measurement scale is a *homomorphism* between objects of measurement and mathematical entities (Mari 2000). A homomorphism is a structure-preserving map between two entities. Some theorists think measurement scales are, or can also be, *isomorphisms*. Isomorphisms are a subset of homomorphisms, where each element of one set is paired with one element of the other set, and each element of the second set is paired with one element of the first set—essentially homomorphisms where no structural information is lost between the two entities. Whereas a homomorphism only partly preserves structural information, an isomorphism, or *bijection*, preserves full structural information. Nominal and ordinal scales will likely be homomorphic rather than isomorphic, because the numerals carry only limited structural information. However in some cases of nominal and ordinal measurement, if there is reason to think that there is no structural information to convey other than a difference relation (for a nominal measure) or a greater than/lesser than relation (for an ordinal measure), then a case may be made for thinking of them as isomorphisms. Some measurement theorists think that no instances of measurement should be understood as isomorphisms (Mari 1996). Isomorphic measurement implies that the system under consideration is already conceivable in terms of quantifiable attributes, that is, it implies that the object of measurement has a structure which exactly mirrors some mathematical structure. Isomorphic measurement involves mapping the quantifiable attributes onto the structure-preserving symbolic entity. However, if measurement also involves the non-pre-determined modelling of things in terms of measurable attributes prior to determining their magnitude then there is reason to think that measurement can never be truly isomorphic.

Nominal and ordinal scales do sometimes use numerals to characterise items, but they are strictly *numerals* and not *numbers*. I use ‘numbers’ to refer to the set of real numbers, and ‘numerals’ to refer to the symbols or figures which are used to denote those real numbers, but also, importantly, sometimes used to denote other entities.



Although most of the properties of real numbers have no empirical significance when numerals are used to characterise items in nominal and ordinal scales, in either case, some of their properties *are* significant (van Fraassen 2008, pp.160–1). So, in the case of nominal scales, different real numbers represent different properties or symbols, but don't represent ordering information or other algebraic structures. As Stevens puts it:

the use of numerals as names for classes is an example of the assignment of numerals according to rule. The rule is: Do not assign the same numeral to different classes or different numerals to the same class. Beyond that, anything goes with the nominal scale. (1946, p.679)

In the case of ordinal scales, different real numbers carry ordering information, but don't represent any other algebraic structure. In many nominal and ordinal scales the numerals will be equally well represented by other symbols, male/female rather than 0/1, or very poor/poor/average/good/very good rather than 1/2/3/4/5, with the numerals playing a symbolic role in both cases.

There is some disagreement as to whether nominal and ordinal operations should be called measurement; it is suggested that they don't provide a sufficiently complex structure to warrant the use of numbers, compared to interval and ratio scales. Norman Campbell, an early 20th century measurement-theorist, argues that the use of numerals in nominal measurement is 'completely arbitrary,' and that measurement requires measurable properties or *magnitudes* to be capable of order (1928, pp.1–8). Moreover, he claims that ordinality is insufficient for measurement, as it leaves open the possibility of infinitely many ways of assigning numerals to objects of measurement (1928, p.10). In fact, in an ordinal series, a 'greater than' relation between magnitudes of the object of measurement need not be represented by numerals which possess that same relation, so long as two systems which produce the

same state in a measuring instrument are always assigned the same numeral, and are always be found in the same place in the order of the magnitudes 1928, (1928, p.11). So an ordinal system very poor/poor/average/good/very good could assign the numerals 2/4/1/5/3 to the five magnitudes, and this wouldn't pose a problem in theory, as long as these numerals were used consistently. Of course the practicality of this approach is another matter entirely—its users would have to ensure they did not observe the conventional greater than and less than relations between the numbers represented by the numerals. In addition to ordinality, Campbell argues that the systems to be measured must be capable of *addition*, such that the sum of the magnitudes of two systems is equal to the magnitude of the combination of both systems (1928, pp.14–15). Campbell's restrictive view of measurement rules out nominal and ordinal measurement.<sup>23</sup> On his view, measurement involves counting concatenations of objects that equal the object with respect to some measurable property or magnitude, and where the concatenation has the same properties as addition in arithmetic—including commutativity and associativity (Savage and Ehrlich 1992, p.2).

Nominal scales certainly don't map objects onto numerical entities, and if the nominal categories are represented by numerals—think, for example, of the numerals on football players' shirts—these can be substituted by any symbol.<sup>24</sup> Where numerals are

---

<sup>23</sup> Campbell's view of measurement is sometimes called 'conservative' and its counterpart 'liberal.' These names sit uncomfortably with me, having too many unhelpful political and evaluative associations. I prefer to think of Campbell's approach as a *restrictive* view of measurement, and Stevens' characterisation as an *inclusive* view of measurement. I use these designations henceforth.

<sup>24</sup> Originally, the numerals on football players' shirts represented a rough spatial ordering of players from goal-keeper to forwards, based on a 5-3-2 formation, so might have been characterisable as an ordinal scale. That characterisation is difficult to uphold once play formations became less static, and numerals came to distinguish players from one another, rather than to mark their position on the pitch.

assigned to objects in ordinal scales, their numerical properties don't carry much significant empirical information. If someone rates their subjective well-being on a scale of 1 to 5, the intervals between 2 and 3 and between 4 and 5 don't necessarily correspond to equal increments of subjective experience. However, both of these operations *do* involve mapping objects into a symbolic relational structure, even if not a strictly numerical one. This involves the identification of particular salient features of objects, and categorisation of objects in terms of these, rather than other features. So the classification of humans into male/female/intersex nominal categories *doesn't* treat eye colour or heart rate as significant, but *does* treat primary and secondary sexual characteristics as significant. And the classification of books into an ordinal scale of best to worse will treat their literary and narrative qualities as significant, but not (typically) attributes such as their length, year of publication and typeface. Both kinds of scale involve a mapping from the objects onto a symbolic structure. Nominal and ordinal operations do involve the assessment and categorisation of objects based on a symbolic system of classification. It is only if measurement is defined as the expression of attributes in quantitative terms that numbers are a requisite feature of measurement—and this essentially assumes, rather than proves, the requirement of numerical valuation. Campbell's characterisation of ordinality as insufficient for measurement, on the grounds that it leaves open the possibility of infinitely many ways of assigning numerals to objects of measurement ultimately begs the question.

The limited empirical significance of the properties of real numbers doesn't provide good grounds for Campbell's restrictive view that nominal and ordinal scales are not instances of measurement. The conception of measurement as the assignment of numbers to things so as to represent facts and conventions about them represents a short-sighted view of mathematics, as being concerned entirely or chiefly with numbers (van Fraassen 2008, p.159; Frigerio et al. 2010, p.125; Mari 1996, p.130). Indeed, even Campbell cautions that nothing should be read into the numerals which

are used represent magnitudes that is not stated explicitly (1928, pp.15–16). While interval and ratio scales are typically taken to be properly quantitative measures, the relevant structural information of numerals is also limited in interval measurements. Although 100 is twice as large as 50, 100°C is not twice as hot as 50°C, because the zero point of temperature on a Celsius scale is conventional. This doesn't mean that it is questionable as to whether temperature is a measurement scale. Interval scales are taken to be instances of measurement, even by people who hold restrictive views, despite the fact that their numerals represent limited structural information. But limitations on the significance of numerical relations cannot alone foreclose the possibility of measurement without excluding interval scales too. Understanding the relevant mathematical entities as symbolic, rather than strictly numerical, seems to better account for the variable empirical significance of numbers in different measurement scales (Mari 1996). But it also brings into question whether there is any good reason for drawing the line between measurement and non-measurement between ordinal and interval scales. Adopting an inclusive view of measurement, as Stevens does, opens up the scope of measurement, as it takes everything to be capable of measurement in some sense (Savage and Ehrlich 1992, p.3).

One concern with treating nominal labelling as measurement is that it may allow unsystematic and non-mutually exclusive classification of entities to be treated as measurement scales. If Jorge Luis Borges' Chinese encyclopaedia counts as a measurement scale, the concept of measurement might reasonably be thought to be bloated beyond usefulness:

These ambiguities, redundancies, and deficiencies recall those attributed by Dr. Franz Kuhn to a certain Chinese encyclopedia entitled *Celestial Emporium of Benevolent Knowledge*. On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g)

stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance. (Borges 1964, p.103)

However, the treatment of nominal scales as measurement scales need not entail that all and any labelling systems are forms of measurement. Nominal and ordinal measurement involves putting things into categories *in a systematic way* (Cartwright and Runhardt 2014, p.267). This means that a measurement scale should characterise and represent the object of measurement in a methodical and conceptually coherent manner. This will typically mean that the mapping will leave neither large gaps, where aspects of the object are unclassified, nor substantial areas of overlap among labels. It also suggests that the measurement scale classifies objects in terms of the same property or magnitude, rather than capturing different things in different objects. Finally, it will likely also entail that the classification of the object of measurement is congruent with our other theoretical and conceptual commitments. Campbell (1928, p.1) proposes a somewhat more stringent but functionally similar limitation on measurement: that it must present properties in accordance with scientific laws. So, while the classification of animals according to the *Celestial Emporium of Benevolent Knowledge* might not count as a measurement scale, the Linnaean system of biological classification likely will, as will post-Linnaean taxonomies. The former is neither exhaustive nor are its categories exclusive nor uniform in scope. It does not match up with other theoretical biological commitments. Linnaean and post-Linnaean taxonomies, however, systematically classify biological organisms in such a way that no known organisms fall outside of their structures, and are designed such that newly 'discovered' organisms can be placed within them. There is no overlap between categories—no animal can be a member of two different species, for example—though there may be disagreement as to where the boundaries between them fall. They also

endeavour to reflect accepted evolutionary theory about biological development and the relationships between species.

I don't want to over-emphasise the strength of the systematicity constraint on measurement scales—it seems likely that some measurement scales will have small theoretical contradictions, gaps between categories or overlapping categories, and that this need not speak against them *qua* measurement scales. Indeed, Bas van Fraassen (2008, p.163) argues that because measurement outcomes are not infinitely precise, the real outcome—when we measure something's height or temperature, for instance—is an interval and not a point. This means that temperature and length scales which appear to be linear, with each numeral appearing once and only once, are in fact made up of a series of overlapping, ordered intervals. This does not indicate that they fail as measuring scales—rather, it is a feature of the inevitable limitations on the precision of measurement. That a measurement scale must be systematic does not entail a particularly clear line between those classifications which are measurements and those which are not, as the adequacy of a system of classification may in part be down to pragmatic considerations, highly linked to both the purpose of classification and the effectiveness of the classification in furthering or realising that purpose. This may not be a binary matter, and a classification will achieve these ends to a greater or lesser extent. Furthermore, the restriction of systematicity doesn't entail that the classification is the *only* or the *best* way of mapping the object or objects under consideration. There are several mutually exclusive, competing systems of biological classification, each of which attempts to systematically map living organisms, and all of which are plausibly measurement scales. So, that a measurement scale is systematic does not entail that the mapping of an entity onto a symbolic structure is entirely without gaps, overlap or error.

This discussion suggests that there is reason to think that nominal and ordinal scales *can* be understood as measurement scales, and that we should adopt an inclusive view

of measurement. If this is right, then even minimal identification and comparison of well-being states will count as a form of measurement. Picking out whether someone has well-being, whether directly, or via the identification of some set of constitutive goods, requires a mapping of the relational structure of the object of measurement onto some symbolic relational structure. In the simplest cases, this will be a yes/no structure, capturing the presence or absence of some good, such as knowledge or happiness, in the same way that you might identify someone's sex by picking out sexual characteristics, or identify the phylum of a plant or animal on the basis of its biological features. Nominal measurement will allow only for minimal comparison between individuals on the basis of the presence or absence of the good, and will not enable any more fine-grained comparison. However, most theories of well-being take well-being to be something which can be possessed in degrees—people can be better or worse off with respect to well-being, and often interpersonal and intertemporal comparisons can be made. Slightly more complex symbolic structures, then, will involve the ranking of states of people in terms of their well-being or its constitutive goods, where this involves identifying degrees of achievement using a less than/greater than structure. Using ordinal measurement will not determine the exact interval between ranked states, and many states may not be comparable in terms of the prudential good, and may as a result remain unranked or be assigned the same ranking. To the extent that comparisons are not possible, a partial ordinal ranking will result. Nonetheless, even in such cases, something is being measured, insofar as relations among objects are mapped onto symbolic entities. Of course, there are also many interval measures of well-being, which no-one would dispute to be attempts at measurement—although their success in actually capturing well-being may well be queried. But if nominal and ordinal scales can also be included as varieties of measurement, it's difficult to see how the well-being theorist can avoid having to say *something* about the measurement of well-being—how it is picked out, how we know whether someone has it, and to what extent.

The controversial-sounding claim that well-being is not just sometimes measurable but that the identification of well-being is identical with its measurement is thus somewhat defused by the invocation of an inclusive view of measurement. Being the kind of thing that is measurable, on this account is a fairly ubiquitous thing to be.<sup>25</sup> But if the bar is set so low, what's the benefit of conceiving of well-being in terms of measurement at all? For readers who remain unconvinced, my thesis does not rest on this argument. The argument in chapters 3–6 depends only on the conclusion of the previous section—that well-being theorists must at minimum say something about how to identify well-being in persons, and cannot just stop at a general definition of well-being. Nonetheless, I think that measurement theory provides a set of conceptual tools which help to elucidate the problems that emerge when attempts are made to identify well-being, and the implications of this for the nature of well-being. In the final section of this chapter, I will suggest why it is helpful to conceptualise well-being in terms of measurement. But first—and in order to motivate these considerations—more needs to be said about the ontology of measurement.

### 2.3 *What is measurement?*

If the specification of well-being involves something which amounts to measurement, the well-being theorist will have to say something about what measurement is. There are, broadly speaking, two ways of thinking about the relationship between measurement operations and the objects of measurement: *realism* and *nominalism* (Chang and Cartwright 2008, pp.367–8).

Realism about measurement takes measurement to be a method of empirically estimating the magnitude of properties and relations of objects (Tal 2017a). Typically

---

<sup>25</sup> My discussion of the Borges example above suggests, however, that being a *measurement* might not be as ubiquitous as being *measurable*.



this involves a commitment to realism about properties and relations. On such a view, there is a fact of the matter about particular properties and relations, which is independent of attempts to measure them (Chang and Cartwright 2008, pp.368–9; Trout 1998, Ch. 4). Insofar as the measurement concept refers to real features of the objects of measurement, representing the concept is a matter of representing the relational structure of the objects, and measurement processes assign values to tokens corresponding to the existence or magnitude of those features that the token displays (Bradburn et al. 2017). Measurement *approximates* the real properties and natures of the objects of measurement, and in many cases measurement can give us information about objective features of phenomena (Swoyer 1987, p.237). Realism about measurement implies that measurement can be more or less accurate, depending on the extent to which the measurements agree with the true values (Chang and Cartwright 2008, p.370). A key argument in favour of realism about measurement is an argument from best explanation. We generally behave as if the objects of measurement have real properties with real magnitudes, which exist independently of our measuring them, and the simplest explanation of this is that they really do exist (Swoyer 1987, pp.238–9; Ellis 1992, p.175). Realism therefore claims that the best explanation is a causal relationship between the objects of measurement and their magnitudes, and the measurement thereof.

In the case of well-being, realism about its measurement would entail that measuring well-being is a matter of obtaining knowledge about states of people and the world. Usually this would mean that people have properties, such as having mental states and affects (“being happy,” “feeling depressed”); having preferences and desires and having them satisfied (“preferring to stay at home, but having to go out for dinner”, “wanting the cake, and having it”); or having prudential goods in their life (“having a great group of friends,” “being well-nourished”). Furthermore, these properties are—really are—constitutive of people’s well-being. This means that such properties are constitutive of

well-being, and each person has a definitive well-being-status, *regardless of whether anyone tries to identify and measure it*. A realist measure of well-being can be more or less accurate, depending on whether the represented values agree with the actual states being measured. For a realist about the measurement of well-being, measuring well-being is a matter of representing the properties of the object of measurement according to some measuring scale, and assigning values corresponding to their existence and magnitude. An accurate well-being measurement operation *reveals* the magnitude of well-being in a person. If and insofar as well-being is taken to be a psychological property of persons, as it is on most subjectivist and some hedonist theories of well-being, realism about well-being implies realism about psychological properties.

Nominalism about measurement, on the other hand, takes measurement operations to define or partly define the concept under consideration (Chang and Cartwright 2008, pp.367–8). The more radical branch of nominalism, *operationalism*, takes the meaning of a concept to be entirely specified by its method of measurement. This means that each measurement operation defines a new concept. So, for example, for the operationalist the measurement of the passage of time via the movement of the sun, on the one hand, and the measurement of the passage of time using electron transition frequency (as is used in atomic clocks), on the other, capture two distinct concepts of time—solar time and atomic time. These may, of course, overlap in their domain of application, but it should not be assumed that two overlapping concepts are in total concurrence, particularly in those contexts where they cannot be checked against one another. At its most radical, operationalism might take each *token* measurement operation to define a new measurement concept, but typically it is understood as taking each *type* of measurement operation as the relevant unit of concern. The view was first promulgated by Percy Bridgman (1927), though he later distanced himself from it (Chang and Cartwright 2008, pp.367–8). Bridgman came to recognise, along with his critics, that operationalism in its most extreme form implies

the meaninglessness of theoretical terms (Mari 1996, p.134). Operationalism implies that a given measurement operation is by definition the correct way of measuring its corresponding concept, insofar as the concept is specified by the operation (Tal 2013, p.1160). It thus becomes very difficult for the operationalist to classify particular operations as inaccurate or unreliable, because there are no non-operational concepts against which to assess the measurement operation. For the operationalist there is no genuine generic concept of length, time, mass, and so on, which covers multiple measurement operations.

A less radical and altogether more plausible approach to nominalism is *conventionalism* (Ellis 1966, 1992). Conventionalism about measurement is the view that some non-trivial aspects of measurement are fixed by convention. To say that something is a *convention* or *conventional* is to say that it is in some sense up to us, and not determined by intrinsic features of the world. So conventionalism about measurement takes some aspects of the way that quantity terms like length and mass apply to objects of measurement to be up to us. Conventionalism implies that no measurement operation is uniquely true, and that there are alternative conventions that could take its place (Chang and Cartwright 2008, p.368). This means that there is no fact of the matter about how long, heavy, hot, or acidic something is that is independent of the particular operations that are used to measure it. Conventions take the form of non-empirically verifiable statements, which link attributes of the objects of measurement with measurement operations. So, for example, the measurement of temperature using a mercury thermometer depends on the assumption that mercury expands uniformly with temperature. This is not an empirically verifiable assumption: evaluating the relationship between the volume of mercury and real temperature requires some way of measuring real temperature that doesn't use a mercury thermometer (Chang and Cartwright 2008, p.369). Of course, some other measurement operation can be used to measure temperature, such as an alcohol

thermometer, which might seem to provide empirical verification. But the very same problems of justification apply here too, as this operation will rest on a similar assumption, for example that alcohol expands uniformly with temperature. This leaves us with a verification regress. Early conventionalist Ernst Mach argued that there is no fact of the matter as to which thermometric fluid expands more uniformly with temperature, as in order to have a determinate concept of equality among temperature intervals, we already have to have chosen a standard thermometric fluid (1986, p.52). Another example of an untestable assumption or a priori statement, which conventionalists took to ground measurement of length, is the claim that measuring rods retain their length when transported (Tal 2013, p.1161).

Unlike operationalism, conventionalism does not take it to be tautologous that a particular measurement operation is the correct way of representing a corresponding concept, though it does take our measurement concepts to be partly grounded on *a priori* statements. Conventionalism is *not* the view that truth is based on convention, though it is prone to be mischaracterised as such; rather, conventionalism 'seeks to expose conventions likely to be mistaken for truths, and calls our attention to the fact that we do have discretion even in contexts where we appear to have none' (Ben-Menahem 2006, pp.1–2). Conventionalists seek to highlight the degree to which our measurement practices depend upon empirical assumptions and *a priori* conventions, and to show the extent to which these are fixed by empirical facts, theories, and considerations of convenience, and the extent to which they are arbitrary (Ellis 1992, p.170). Different measurement operations can be assessed on the grounds of their usefulness, relative to particular purposes, and given other commitments and theoretical frameworks (Ellis 1992, p.169). We can fix on a definition with the intention of regulating the use of a particular measurement concept, but such restriction is only justified insofar as it actually garners agreement among relevant parties (Chang and Cartwright 2008, p.368). This means that the conventionalist can

identify a particular measuring system as the correct one to use in a given context, where the appropriateness of a given concept will depend on a description of the goal and context of measurement.

Conventionalism in its original form declined alongside logical positivism in the mid-20th century, but contemporary theories of measurement have unmistakably conventionalist roots. Measurement theorists such as Eran Tal (2017a, 2017b), Bas van Fraassen (2008), and Luca Mari (2000) endorse a *model theory of measurement*. Model theories take measurement to involve two stages: first, there is an interaction between an object of measurement, an instrument, and the measuring context, of which the measuring subject (the person making the measurement) is a part (Tal 2017a, 2017b). This process produces *measurement indications*, which may be the final resting point of a pointer on a dial, the level of a thermometric fluid against a calibrated scale, or the numbers displayed on a digital screen. Secondly, an abstract representation is constructed to model the interaction between these *relata*, and values (sometimes numerical) are assigned to some of the variables in that abstraction. This abstraction produces *measurement outcomes*, which are claims about the object of measurement. The model simplifies the interactions between the *relata*, and is used to predict and explain aspects of the phenomenon of interest: the object of measurement. Models are constructed out of theoretical and statistical assumptions (Tal 2017b), and will reflect the purpose of the experimenter (van der Bos 1994). Obtaining measurement outcomes requires that we engage in theoretical and statistical modelling of the measurement process—without this we just have digits on a screen or numbers on a calibrated scale.

The implication of this view is that measuring something always involves describing it in such a way so as to classify it as a system that is characterised by whatever quantity or attribute you seek to measure. The model places the object of measurement, which has already been classified as within the scope of a given theory, in what Bas van

Fraassen calls a *logical space* and Tal calls a *parameter space*, which represents a range of possible states in which the object could be or characteristics that it could have (van Fraassen 2008, p.164; Tal 2013 p.1165).<sup>26</sup> The outcomes of measurement are not, then, numbers, but regions in a parameter space. A measurement outcome is considered an adequate representation of the object of measurement if the theory provides a coherent explanation of how the possible set of indications on the apparatus reflect possible states of the object. Crucially, on such an account, measurement is understood as a means of gathering information about the object, but this information is not understood to exist in an extra-theoretical sense—theoretical representation of the object of measurement is not only necessary for interpreting indications, but for measurement *per se* (Tal 2013, p.1165). In this sense, this approach is *instrumentalist*—it takes measurement to serve a practical purpose rather than an absolute one—be a means of gathering information, rather than a means of finding out about objective quantities (van Fraassen 2008, p.143).<sup>27</sup>

Coherence and epistemological holism are central to the instrumentalist project—while the instrumentalist *may* take certain theoretical, empirical and conventional claims as fixed points on which to ground measuring systems, in theory these are open to change, depending on their fitness for purpose (Ellis 1992, p.179). The instrumentalist must, to some extent, endeavour to distinguish facts from

---

<sup>26</sup> Mari (2000, p.75), in his characteristically opaque style, says that a ‘standard set’ or ‘reference relational structure’ is ‘conventionally chosen by the operator with the aim of materializing his knowledge on the *measurand*.’ This, I take it, amounts to much the same thing as placement in a parameter space.

<sup>27</sup> van Fraassen (2001) rejects the instrumentalist label and refers to his view as *constructive empiricism*. I retain the label, as I think it captures something about the way that, on his view and others like it, measurement is understood as a means to generating information about an object of interest. But I recognise that instrumentalism is a fairly broad church and includes more and less permissive views.

conventions, that is, between empirically fixed points in our theories and theorising on the one hand, and claims over which we have discretion on the other (Ben-Menahem 2006, pp.3–4). However, she need not think that facts are strictly independent of conventional assumptions, nor that conventions are unable to be displaced by facts. A convention may, for instance, turn out to be unfounded on the grounds that it is incompatible with some broader theoretical claims, which are themselves grounded in empirical observation. And the collection of empirical data may be possible only because of certain assumptions about the possibility and nature of measurement. This is a far more nuanced concept of convention than the static positivist account of a fixed and arbitrary a priori assumption.

For the instrumentalist, there may be multiple ways of measuring a given concept, and these methods may not be in perfect agreement over all contexts. Moreover, there may be some contexts that are out of the scope of particular measurement operations. So, for example, the length of a day measured using the astronomical movement of the sun might not exactly accord with the length of a day measured using an atomic clock, and it will not be possible to measure very short periods of time using the former method. However, in order for two measurement operations to represent the same concept, there must be some overlap between them—some extent to which they concur on the magnitude of the object of measurement. We can continue to hold a general, cross-operational concept of time concurrently with the concepts of its various divergent measurement operations. In any particular measurement context one operation has to be chosen, and doing this will inevitably change or limit the meaning of the concept relative to the overall general concept and to the conjunction of all of the other operations. The realist notion of accuracy does not figure in the instrumentalist account of measurement, as there is no attitude independent fact of the matter about the properties of the states being measured. Instead, the adequacy of instrumentalist measurement rests on a number of factors: *precision*, that is, how

specific measurement results are, or how close measurement values are to one another; *coherence*, that is, whether the measurement outcomes, their relation to the object of measurement, and their function as sources of information for a particular purpose cohere with the theoretical characterisation of the measurement scenario (van Fraassen 2008, p.145).

Instrumentalism about the measurement of well-being would entail that there is no fact of the matter as to what is being measured when we measure well-being. There is no truth as to how well someone is doing that is independent of conventions about its definition and measurement. These will include linguistic conventions, but also practical limitations on the possibility and ethics of collecting information about the people who are the objects of measurement, and their environments (Alexandrova 2008). Instrumentalist well-being claims and measures would be limited by theoretical constraints, empirical facts and conventions—for example, facts about people’s access to goods and resources; facts about the results of psychometric surveys; claims that no-one who says that they are not doing well can be declared to in fact be doing well; claims that any dominant states (where someone has exactly the same properties as another person plus more along any relevant dimension) must be indicative of higher-levels well-being. However, these too are up for debate, and may be applicable only in certain contexts. And importantly, it is not only the clearly normative assumptions and claims that are up for debate, but also the facts: the relevance, definition and identification of the goods which constitute well-being can be called into question, as can the validity and relevance of surveys and tools. Certain conceptions of well-being would be ruled out by conventional policing of the concept—that is, they wouldn’t count as ways of identifying and measuring well-being in absence of agreement among the relevant parties.



## 2.4 *Realism, instrumentalism, and well-being*

In this section I argue that invariantism about well-being implies some form of realism about the objects of measurement. For variantism, the picture is slightly more complex: stance dependent pluralism about well-being likely goes hand in hand with some form of instrumentalism, but realism about well-being and its measurement remains a possibility for the stance independent pluralist.

The way that philosophical theories of well-being describe the concept of well-being often implies a form of realism. Invariantist theories of well-being, discussed at length in chapter 1, seem particularly prone to such a reading. This claim is quite difficult to definitively show, as the vast majority of well-being theorists do not address the nature of well-being in quite these terms. However, the straightforward reference to well-being in terms of states of persons and the world appears indicative of a realist view. For example, Chris Heathwood writes:

An instance of "subjective desire satisfaction" is a state of affairs in which a subject (i) has an intrinsic desire at some time for some state of affairs and (ii) believes at that time that the state of affairs obtains. An instance of "subjective desire frustration" occurs when (i) above holds but the subject believes that the desired state of affairs does not obtain. The value for the subject of (or the amount of welfare in) a subjective desire satisfaction is equal to the intensity of the desire satisfied. Likewise for frustrations, except that the number is negative. The theory is summative so that the total amount of welfare in a life is equal to the sum of the values of all the subjective desire satisfactions and frustrations in that life. (Heathwood 2006, p.548)

This passage, and in particular the final sentence, seems to suggest that the subjective desire satisfactions and frustrations exist prior to an independently of the attempts to

measure them and calculate the total amount of welfare. Or—another example—Christopher Rice writes:

I will defend the objective list theory of well-being, according to which all instances of a plurality of basic objective goods directly benefit people. This is a pluralistic view, since it does not identify a single feature that is common to all states of well-being and ultimately explains their value. It is also an objective view, since it holds that states of affairs benefit people independently of their positive reactive attitudes toward them. (Rice 2013, p.197)

This is, perhaps, a little less clear, but again the language suggests that the goods in question ‘directly benefit’ people independently of whether attempts are made to assess this. In particular, the idea of people having ‘states of well-being’ suggests that well-being or its constituents are properties that individuals have. Occasionally, well-being theorists state something like a realist conception of well-being fairly explicitly. Ben Bradley (2009, p.6), for example, writes: ‘If two people's lives differ in value, it must be because one has a property the other lacks; that property must show up somewhere in our axiological theory, or our theory is incomplete.’ This unequivocally identifies the magnitude of the value of people’s lives with the presence or absence of some property or set of properties.

So, on such theories, well-being is identical with some property of persons or the world. For hedonist theories, this will be pleasure, satisfaction, enjoyment, or other forms of positive experience which are all typically understood be states or properties of a person. For desire fulfilment theorists this will be either a person’s subjective perception of the satisfaction of their desires—again, a property of the person themselves—or the actual satisfaction of their desires—a state or property of the world. Those desire theorists who restrict the desires that are relevant to well-being, or otherwise idealise or launder the desires and preferences in question, may identify

well-being with a property of a person or of the world, depending on how the desires are restricted. For example, if well-being amounts to the satisfaction of the rational desires that a person in fact has, then well-being picks out a property of the person. If, on the other hand, well-being amounts to the satisfaction of the desires that a person would have were they rational, well-being appears to pick out a property of the world. For objectivists about well-being, well-being is identical with the achievement of a set of prudential goods, this also either amounts to a property of persons or a property of the world, depending on the nature of the goods. Does this identification of well-being with a property or set of properties of persons or the world entail realism?

It's difficult to make sense of theories of well-being that seek to identify the necessary and sufficient conditions for well-being to the exclusion of other accounts on anything other than a realist model. If an invariantist about well-being is also an instrumentalist, this would amount to the view that there is no single real underlying property 'well-being' that is measured, but nonetheless there is only one plausible way of conceiving of well-being, which is consistent across all contexts. That is, the referent of well-being is not real, but it is constant. This is not in itself in tension with an instrumentalist framework—it could well be the case that there is a single, static referent of a given concept, where the referent is not a real entity. Take, for example, an instrumentalist who thinks that 'electron' refers to patterns in cloud chambers, and we need not infer the existence of some real underlying entity. There might be very little variation in the concept electron, perhaps chiefly because the concept is a theoretical postulate, and so acquired meaning only given a theoretical framework, as a way of describing a particular phenomenon. This suggests that instrumentalism *supports* but doesn't *ensure* pluralism. However any instrumentalist monism is a contingent monism—it is dependent on theoretical stability, ongoing empirical support (or at least no substantial empirical contradiction) and continued acceptance of the relevant conventions. Take the electron example again. The instrumentalist about electrons

would not believe electrons to be a natural kind, and it could well turn out that making a distinction between two types of patterns in cloud chambers, between  $\text{electron}_1$  and  $\text{electron}_2$ , better predicts and explains certain empirical phenomena. The instrumentalist might therefore determine that 'electron' is better understood as a plural concept. Similarly, the instrumentalist invariantist about well-being would have to be open to the possibility that other conceptions of well-being *could* emerge with alternative, theoretically consistent ways of describing and measuring well-being. Such alternative well-being operations are, on this account, genuine conceptions of well-being so long as they are consistent with our broader understanding of well-being and facts about well-being, which are themselves fit for purpose.

This is a weaker claim than the typical invariantist wants to make. The invariantist does not deny that people use and theorise the concept of well-being in ways other than those specified by their theory, but they argue that they are wrong to do so, or wrong to treat them as distinct conceptions of well-being. The invariantist seeks either to assimilate purported theories of well-being under their own theory, or to deny that they in fact describe well-being, and so to reject them as theories of well-being. The instrumentalist invariantist could, then, attempt to assimilate or reject alternative well-being operations on the grounds that they do not cohere with our broader concept of well-being, or with fixed facts about well-being. But it doesn't look as though such an approach can be successful without a certain amount of question-begging. In order to motivate the claims of incoherence, the 'broader concept of well-being' and the 'fixed facts about well-being' have to be defined with reference to the invariantist well-being operation in question. In a context of multiple purported well-being operations, there isn't a good basis for doing this unless there are some stronger grounds for thinking that the invariantist operation is the right one. Without such grounds, other purported well-being operations *themselves contribute to the broader concept of well-being*, insofar as they are taken to describe well-being. If we assume invariantism about well-

being, then it makes sense to rule out alternative well-being operations on the grounds that they are inconsistent. But if invariantism about well-being is not taken as a fixed point of reference, then discounting other operations because they are inconsistent with the broader concept of well-being is unjustified.

If the invariantist is to reject purported theories of well-being on the grounds that they do not in fact capture well-being, *even though people think that they do*, they will have to make a stronger claim about the nature of well-being than instrumentalism allows. Realism about well-being measurement seems well-placed to explain this—other conceptions of well-being should be rejected because they do not capture the truth about well-being. People have some real property which is constitutive of well-being—or some set of real properties which are jointly constitutive thereof—and the correct theory of well-being is the theory which tells us how to identify and measure these, where measurement of them involves representing them and the relations between them *as they really are*.

While realism and invariantism are certainly natural bedfellows, there remains the possibility of a realist variantism about well-being. Suppose we accept a variantist account of well-being, and thus think that different well-being operations are appropriate in different contexts, and that these are specified by different theories of well-being. But furthermore, suppose that we take there to be some real property of persons or the world which each of these operations measure. So, in one case, well-being is measured by measuring subjective psychological states, in another case well-being is measured by measuring desire satisfaction, in a third case well-being is measured by measuring capabilities, and so on, but all of these operations measure the same underlying property—the person's *well-being*—and this exists and has a magnitude independently of our attempts to measure it. How do we know that all of the various well-being operations are in fact measures of well-being, and not measures of a diversity of different properties? There seem to be two possible options: first, we

can identify them all as measures of well-being because we have some wider theory of well-being which connects all of the different concepts of well-being, and shows how they all define and measure the same property. The realist about measurement can acknowledge a number of different operations for measuring, for example, length, and can accept that some of them are only appropriate in particular concepts. So, for example, measuring length via the concatenation of metal rods is not appropriate for measuring interplanetary distances. What makes these all measures of length, for the realist, is the fact that they all produce estimates of the same underlying property, and this can be shown by proving that they are consistent with one another at those points where they have overlap in scope. In the context of well-being, this would require some way of assessing the consistency of the variety of concepts and measures of well-being. But this takes us out of variantist territory. If some overarching theory and concept of well-being can be identified, which connects all of the context-specific theories and concepts of well-being, this would constitute an invariantist explanation. That is, it would provide a central definition of well-being, which captures all and every instance of well-being, and would relegate the variety of measures of well-being, casting them as estimates of the overarching concept, merely picking out causes and correlates of well-being (Alexandrova 2017b, p.47). So such a response will not do for the realist variantist.

Secondly, we can identify the variety of well-being concepts as instances of well-being on the grounds that they play a similar functional role in each of their contexts, even though they capture different properties in different cases. On this view, the variety of concepts and measures of well-being would each measure some distinct real property or set of properties, which captures what is good for the person in question in that context. Measuring well-being in any given case would involve mapping real properties onto a mathematical structure, but in different contexts, different properties are mapped onto different mathematical structures. It is unclear whether

such an account should be taken to be instrumentalist or realist. On the one hand, an instrumentalist variantist might take the identification of a variety of properties as instances of well-being might be a matter of convention, on the grounds that conceiving of them as such best accounts for the roles they play in their respective contexts. So it makes sense to conceive of them all as measures of well-being, and as being relevantly similar, because this makes good sense of the way that we conceive of human lives and purposes across contexts.

This instrumentalist variantism is aligned with the stance dependent pluralism outlined in chapter 1. On the other hand, a realist-inclined variantist might insist that the variant, real properties which constitute well-being in each case are facts about the world, not a matter of convention at all. So the fact that  $x$  plays the function of well-being in this context and  $y$  in that context is true in virtue of attitude independent facts about the world, rather than in virtue of any conventions. This account would remain realist in an important sense—in each case there is some underlying property of well-being, which particular operations of well-being estimate, and which is constitutive of well-being independently of any attempts to measure it. This realist variantism is aligned with stance independent pluralism. So depending on how she conceives of the notion of function—whether in terms of human interests, ends and pursuits, or in more naturalist terms—and how she understands the relationship between different context-dependent concepts of well-being, the variantist is led in different directions.

My arguments in the three chapters of part II will suggest further reasons against being a realist, and against being a stance dependent pluralist, about well-being. I will argue that we should understand well-being to be a relation between the *object of measurement*—the individual whose well-being is being measured—the *measuring tool*, and the *measuring environment*, which includes the *measuring subject*—the

person or persons attempting to measure well-being.<sup>28</sup> Assessments of well-being are ‘up to us’ in a distinctly non-trivial way—there is no way to conceptualise or measure well-being without such input from the measuring subject. The discussion in part II thus provides *prima facie* evidence against realism about well-being. If well-being is best understood relationally, and depends upon the attitudes and purposes of the measuring subjects, then it cannot plausibly be described as a property of persons or the world in the realist sense. If the measurement of well-being depends on the attitudes and aims of the measuring subjects, then clearly some aspects of it are ‘up to us’ in the conventionalist sense.

### 2.5 *Why measurement?*

In this chapter I have argued that a theory of well-being must say something about how to identify well-being; that attempts to identify well-being according to systematic characterisations thereof should be understood as attempts to measure it; and that one can only be an invariantist about well-being if one is a realist about well-being. You might well be wondering why it is helpful to think of well-being in terms of its measurement at all, especially given that the definition of measurement to which I have aligned myself is an expansive and inclusive one, and does not imply, for instance, that well-being is always quantifiable. The preceding sections have, I hope, given some indication of the contribution that measurement theory can make towards theorising about well-being, but I will end the chapter with some explicit remarks on this subject.

---

<sup>28</sup> For now, I continue to use the term ‘object of measurement’ to describe the individual whose well-being is being measured, despite the fact that in the case of well-being measurement, the object is, of course, a subject, and often their subjectivity is crucial to ascertaining their well-being. Understanding them as an object of measurement does not, I think, preclude them from being simultaneously a subject. Sometimes the measuring subject and object of measurement might be one and the same—when someone introspectively measures their well-being. I discuss such cases explicitly in chapter 6.



Most importantly, if my arguments for the identification thesis and the inclusive view of measurement are sound, such that (i) theories of well-being must say something about the identification of well-being, and (ii) nominal and ordinal measurement scales are in fact best understood as measurement scales, then well-being *is* best understood as a measurable quality or quantity. This must provide *some* good reason for thinking that measurement should feature in our philosophical thinking about well-being. However, you might remain unconvinced by these arguments, or might think that insofar as well-being is measurable, it is a very limited form of measurement, and the fact of its measurability is barely worth mentioning at all. After all, if well-being is measurable then most philosophical concepts—justice, freedom, equality, knowledge, belief, justification—likely turn out to be measurable too. In this case, a brief consideration of the role I see measurement as playing may help to convince the sceptical reader of the benefit of taking well-being to be measurable, and of thinking of theories of well-being as specifying measurement operations.

Thinking of theories of well-being as specifying measurement operations provides a helpful framework for considering the nature of well-being, and highlights aspects of the nature of well-being that are less obvious when theorising about well-being is taken to be a matter of conceptual or definitional analysis. Well-being theorists often invoke Roger Crisp’s distinction between *enumerative* theories of well-being—those that describe or enumerate the things that make someone’s life go well for them—and *explanatory* theories of well-being—those that explain *why* those things make their life go well (Crisp 2006; Woodard 2012; Fletcher 2013; Rice 2013). Crisp thinks any credible theory of well-being must attempt to do both. My suggestion that theories of well-being must also involve an identificatory aspect brings a third question to the floor—not just the enumerative “what makes life go well for someone?” and the explanatory “what it is about these things that makes life go well for someone?” but

the identificatory “how do I determine the extent to which life is going well for someone?”

Knowing what makes life go well for someone—be it pleasure, or satisfaction of their rational desires, or the achievement of ten basic capabilities—does not in itself tell us how to determine whether or not someone’s life contains those things. Measurement theory provides tools for understanding what is involved in determining the extent to which life does well for someone—beyond knowing what makes life go well for them, and why it does so—and so for understanding the scope of the third question. It suggests, for example, that answering this question requires us to provide some kind of a mapping from ‘someone’s life’ to ‘well-being’ in order to evaluate the former in terms of the latter. This mapping might be a very simple yes/no framework—they either have well-being or they do not—or, more likely, it might involve further discrimination between degrees of well-being. Measurement theory also tells us that we need to determine the nature of the relationship between ‘someone’s life’ and ‘well-being.’ Is the latter a property of the former, or is it a relation between some other properties of the former? Does well-being exist in or have relevance to someone’s life prior to, or only within, the context of attempts to determine its presence and extent? This brings into clear focus a question about the ontology of well-being, which is not necessarily obvious on consideration of the enumerative question—that is, not just how to describe the things that makes life go well for someone, but how to understand the nature of those things. Answers to these questions inevitably shape and constrain the possible answers to the enumerative and evaluative questions.

By suggesting that well-being is understood in terms of measurement—and, importantly, given my inclusive and broadly coherentist view of measurement—I don’t mean to suggest that this in itself tells us something significant about the nature of well-being. It’s not that insofar as well-being is something to be measured, it is fundamentally different from what we might have previously thought it to be. Rather,

measurement provides a conceptual framework which brings certain kinds of questions to the fore—questions which have always been relevant to theorising about well-being, but which are often side-lined in pursuit of answers to the enumerative and evaluative questions. The inclusive view of measurement that I defend supports this characterisation: measurement and measurability are not concepts which apply only to particular kinds of objects or concepts. Rather, they particular are ways of characterising objects—that is, characterising them *systematically* and *coherently*, and *for a particular purpose*. Conceiving of the identification of well-being as the measurement of well-being is therefore a gateway into a discussion of the nature of well-being, which aspires to consider not just how to define well-being and how to justify the characterisation of well-being at the most general level, but how a more fine-grained consideration of the identification of well-being in individuals feeds back into, shapes and constrains the way in which well-being can be defined in the more general sense.



## Part II



## Introduction to part II

The three chapters of part II merit a brief introduction. In some ways, they appear to disrupt the narrative of the thesis; in another sense, they are its backbone. So far I have made a case for taking the variety of theories, concepts and measures of well-being in use in policy and practice at face value, and for understanding theorising about well-being to involve theorising about its measurement. The following three chapters each take as their subject one approach to conceptualising well-being. In exploring the identification and measurement of well-being in detail, in relation to each of these three approaches, I build up a case for a more general defence of my stance dependent account of well-being.

Chapter 3 considers objectivist approaches, which equate well-being with the presence of some set of specified goods in people's lives, and accordingly take the measurement of well-being to involve measuring their presence and magnitude in people or in the world. Chapter 4 considers preference-based approaches, which equate well-being with the satisfaction of preferences, and so take the measurement of well-being to involve the elicitation of preferences and the measurement of 'utility.' Chapter 5 considers hedonist approaches, which equate well-being with the presence of certain subjective psychological states, and so take the measurement of well-being to involve the measurement of the presence, quality and magnitude of subjective states. The examples that I explore in detail in each chapter—the theories of Nussbaum and Sen in the case of objective list theories; economic theories of utility in the case of preferences; and Daniel Kahneman's expected utility theory in the case of subjective

experience—are all fairly well operationalised theories of well-being. That is, they provide not just high-level definitions, but worked out accounts of how to identify and measure well-being.

In selecting these three approaches I do not intend to suggest that they exhaust the possible ways of theorising about well-being. The tripartite characterisation is a nod towards the convention in the philosophy of well-being of dividing up the well-being landscape into desire or preference-satisfaction theories, hedonism, and objective list theories. Although I am not convinced of the merits of the tripartite framework in general, consideration of three quite different approaches furnishes me with a broad evidence base for my more synoptic conclusions in the final chapter.<sup>29</sup> Nor do I seek to critique or defend the conceptualisation of well-being in terms of preferences, subjective experience, objective goods, or any other enumerative or explanatory theory of well-being. Instead I explore what the identification and measurement of well-being under each of these descriptions entails, and consider what this might tell us about the nature of well-being. I suggest that in each case, structurally similar considerations emerge, which indicate that well-being does not pick out a real property of persons, which can be measured more or less accurately, but that well-being is a relation between the object of measurement, a measuring instrument, and the environment in which measurement takes place.

---

<sup>29</sup> I am by no means the first to take issue with the three-way distinction. Andrew MacLeod (2015) and Christopher Woodard (2013) each propose scales or typologies of well-being that eschew the orthodox dissection. Crisp's aforementioned enumerative/evaluative distinction also offers another way of thinking about the well-being theory landscape. Alexandrova (2017b, appendix A) neatly shows the distinctions between the three canonical approaches to be fluid, with very few theories best described as 'pure' examples of any one approach.



Many philosophical theories of well-being take a ‘top down’ approach to deliberating about well-being, starting from broad conceptual definitions and distinctions and attempting to apply these to particular cases. This is consistent with an invariantist framework, where finding the correct definition of well-being provides the tools for assessing and critiquing the application of it in particular instances. As I discussed in chapter 1, the invariantist might invoke intuitions about particular cases in order to determine broad principles about well-being, but for him it is the principles which explain the cases, rather than the cases which explain the principles. The variantist framework suggests a ‘bottom up’ approach, starting from the fact and act of measuring well-being in particular cases, and determining what this implies about the concept more generally. For the variantist, a more general account of well-being is constituted by judgements and practices in particular cases—it does not necessarily give rise to such judgements and practices. Moreover, the centrality of measurement to well-being indicates that understanding well-being requires close consideration of the operationalisation of concepts of well-being more broadly defined. It is this which leads me to seek evidence for my more general account of well-being in the particular.



## CHAPTER 3

### Objective goods

Objective list theories of well-being take what is good for us to be a matter that is independent of our attitudes. In this chapter, however, I argue that even if we accept that the goods on the objective list can be identified as good for us independently of our attitudes towards them, it is not possible to determine how well off someone is, or to compare the well-being of two people or groups in an attitude-independent manner. This means that the objectivity or attitude independence of objective list theories is significantly limited, and any plausibly comprehensive objective list theory requires a specification of well-being that is at least partly attitude- and context-dependent. In particular, specification of the indicators of prudential goods and of how to measure them are dependent on the purposes and context of the person or group attempting to measure well-being, and so unavoidably attitude dependent in some sense. While I begin the chapter by considering a wide array of objective list theories, I zero in on the capabilities theories of Sen and Nussbaum. These accounts take seriously the requirement of a theory of well-being to say something about how to identify and measure it in people, and so they provide good examples of theories which go beyond a high-level definition, and attempt to specify well-being and its indicators in some detail.

The chapter proceeds as follows: in 3.1, I outline objective list approaches to well-being. In 3.2, I explore the nature of the relation between the prudential goods on the

list and well-being more generally. In 3.3 I argue that the objective list theorist will have difficulty identifying the goods, even at the highest level, in an attitude independent manner. In 3.4, I argue that an objective list theory of well-being needs to do more than describe the prudential goods at a high level of description: it needs to say something about the interaction between the goods, the identification of the goods in particular individuals, and how to go about measuring them. In 3.5, I look at Nussbaum and Sen's perfectionist theories, suggesting that they *may* provide sufficient detail, but that their objectivity is questionable. In 3.6, I argue that other objectivist theories cannot avoid following a similar route to Nussbaum and Sen insofar as they develop a sufficiently specified theory of well-being. The identification and measurement of prudential goods will always involve a certain degree of context relativity, where the context in question includes the attitudes and goals of those people who are measuring well-being.

### *3.1 Objective list theories*

Objective list theories of well-being take what is good for us to be constituted by a series of objective goods. Such prudential goods might include knowledge, rational activity, friendship, practical reasonableness, religion and achievement (Fletcher 2015, p.149). The chief attribute of these theories is the objectivity of the goods, that is, that they are good for us regardless of our positive attitudes towards them. This means that, at least on some objective list theories, something can contribute to an individual's well-being even if they do not think that it is good for them, and would rather not have it. So, for example, some objective list theories take knowledge to be an attitude-independent prudential good.<sup>30</sup> On such views, knowledge is good for someone even

---

<sup>30</sup> John Finnis (1980), for example.

if they have no desire to seek it. It matters not that the lotus-eater wants to live their life in a state of dreamy, apathetic bliss—it would be better for them were they to pursue and acquire knowledge.<sup>31</sup> I will follow Fletcher (2015) in calling this characteristic of objective prudential goods ‘attitude independence.’

Another attribute sometimes ascribed to objective list theories is pluralism about prudential value (Rice 2013). Many objective list theories have more than one good on the list, and take these to be irreducible to some master value. Fletcher (2015, pp.148–9) takes such accounts to be paradigmatic objective list theories in this respect. Pluralism about prudential value entails that there are several things that have non-instrumental value for us. They are non-instrumentally prudentially valuable in so far as they are prudentially valuable as ends in themselves, and not merely as means to other ends. This means that the plurality of prudential goods are not reducible to some overarching master value, or to one another. Pluralist objective lists might take there to be no features that items on the list share, or insofar as they do share features, these don’t exhaust their prudential value or fully explain their presence on the list. However, pluralism is not taken to be a necessary constraint on objective list theories, which can also be monistic (Fletcher 2013, p.210). Certain hedonist and perfectionist theories of well-being count as objective list theories of well-being—with lists that contain only one item.

The suggestion that a theory of well-being which says that well-being is constituted by pleasure alone is an objective list theory of well-being is perhaps initially counter-intuitive, but calls attention to a feature of the objectivity of objective list theories worth emphasising. The items on any objective list are not necessarily objective in the

---

<sup>31</sup> Finnis *does* in fact think that individuals can reasonably chose to treat one or some of the goods as more important than others, so construct a subjective ranking of goods (1980, pp.93–4), but he does not take the identification of the goods themselves to be a subjective matter (1980, pp.64–5).

sense that they capture non-subjective states of a person or their environment. To the contrary, many objective list theories include goods such as happiness, pleasure, and self-respect, which capture subjective attitudinal states. All of the goods on Fletcher's list—achievement, friendship, happiness, pleasure, self-respect, virtue—necessarily require engagement in the form of pro-attitudes such as desire and endorsement (2013). He takes this to be necessary for ensuring that people are not alienated from their own prudential good, and to avoid the uncomfortable position of insisting that something is good for someone, despite him not engaging with it at all (2013, pp.215–17). The items on the list are objective in the sense that their presence on the list and their contribution to well-being is attitude independent, rather than their content. Furthermore, objectivity or attitude independence need not require that the goods exist independently of human thought. The goods may well be distinctively *human* goods—of knowledge or practical reason, perhaps—which don't, or at least might well not, exist outside of human society. However they are objective or attitude independent insofar they are good for us independently of whether we think that they are.

I take there to be no good reason to insist on pluralism as a constraint of objective list theories. It may well be the case that pluralist objective lists tend to provide more plausible accounts of well-being than many monistic lists—restricting our understanding of well-being to pleasure, for example, may fail to capture lots of the things that we take to be good for us. However, insisting on pluralism rules out objective theories of well-being which meet the attitude-independence requirement, but which take there to be a single prudential value in terms of which all other values should be understood. In absence of the pluralism constraint, those theories of well-being which don't count as objective list theories include only those theories which take what is good for someone to be determined by their desires or preferences. Such

theories are attitude dependent because whether or not something contributes to well-being depends on the attitudes of the subject under consideration.

Perfectionist or essentialist accounts of well-being attempt to provide a unifying justification for the prudential value of objective goods like friendship, rational activity, and happiness, arguing that they are good for us because, and insofar as, they constitute our flourishing, or manifest the development of our characteristically human capacities (Bradford 2015). Offering an explanation for the value of objective goods, that is, by expressing their value in terms of another, final, value, seems to be something of an advantage for the perfectionist over pluralist objective list theories, which can struggle to offer a satisfying explanation for the presence or absence of goods on their lists. Perfectionist accounts of well-being typically *do* enumerate a number of objective goods which constitute our well-being, but they seek to give a further explanation for their prudential value. Philippa Foot, for example, takes there to be several potential objects of well-being: she mentions having a home, a family, friendships and meaningful work (2001, p.88). Explanation as to why these are good for us does not stop at the goods themselves, rather they are good because they form part of a flourishing human life. Nussbaum (1992) identifies ten *capabilities*—that is, things that people are able to do and to be—which she takes to constitute the central elements of ‘truly human functioning,’ fulfilment of which is necessary to attain well-being. These include bodily health, practical reason, political and material control over one’s environment, and play. While Nussbaum specifies what appears to be a fixed list of goods, presence on the list is explained by the role that the goods play in truly human functioning, and she sees the list as ultimately open ended (2001, p.77).

While efforts are sometimes made to distinguish objective list accounts and perfectionist accounts of well-being, perfectionism can be treated as a variety of objective list theory (Varelius 2004, p.81). There are, of course differences between pluralist objective list theories and perfectionism, mainly relating to the kind of

explanation that is given by each in justification for inclusion on the list. However, I will treat perfectionist theories of well-being as objective list theories. Perfectionism, like all other objective list theories, takes the fact that what is good for us is the development of our characteristically human capabilities to be true independently of our attitudes. It is good to pursue flourishing, or to cultivate one's characteristically human characteristics independently of whether you want to do so. It therefore fulfils the attitude independence constraint.

### 3.2 *Prudential goods and well-being*

Whether pluralistic or monistic, objective list theories must say something about how the item or items on the list relate to well-being. I mentioned above that objective list theories typically take the goods to be *non-instrumentally* good for us. But what does it mean to say that some value or set of values is non-instrumentally good for us? Something having non-instrumental value typically implies that it is not good for us because it brings about some other prudential good, but rather, it is good for us for its own sake. Some objective goods theorists, however, also describe the goods as having *intrinsic* value. John Finnis, for example, claims that knowledge is an intrinsic good, by which he means that it is:

considered to be desirable for its own sake and not merely as something sought after under some such description as “what will enable me to impress my audience” or “what will confirm my instinctive beliefs” or “what will contribute to my survival.”(1980, p.62)

This sounds a lot like the definition I have given of non-instrumental goodness, which suggests that ‘intrinsic’ and ‘non-instrumental’ are synonymous. Indeed, some philosophers do contrast intrinsic value with instrumental value (Shafer-Landau 2007, p.281; Schroeder 2016). However, Christine Korsgaard (1983) entreats us not to conflate the distinction between *intrinsic* and *extrinsic* value with the distinction



between *instrumental* and *final* value. She argues that something that is *intrinsically* good is *not* something that is valued for its own sake, but rather something that ‘has its goodness in itself,’ in virtue of its non-relational properties (Korsgaard 1983, p.170). Conversely, something that is *extrinsically* good has value in virtue of its relational properties. Instead we should understand *final* goods to be those goods which are valued for their own sake, as ends, and, in contrast, *instrumental* goods as those goods which are valued for the sake of something else, as means. So perhaps Finnis is mistaken in his invocation of ‘intrinsic’ value here, but regardless, it’s possible that objective goods could also be taken to have intrinsic, *as well as* non-instrumental or final prudential value.

However, just in case the items on the list are non-instrumentally prudentially valuable does not mean they necessarily have final prudential value in Korsgaard’s sense. They could be instead understood to be *constitutively* prudentially valuable. Something has constitutive value when it derives its value from the role it plays in constituting something else that is valued for its own sake (Shafer-Landau 2007, p.281–2). Whereas something with instrumental value is valuable insofar as it brings about or causes something of final value, something with constitutive value is valuable insofar as it is definitive of some larger whole which is in itself valued (Dworkin 1988, pp.80-81). The notion of constitutiveness is more commonly known in the context of speech-act theory. John Searle (1969, pp.33–4) introduces the notion of a constitutive rule, which is a rule that defines new forms of behaviour, rather than merely regulating existing forms of behaviour. The movement rules of chess or the offside rule in football are examples of constitutive rules: the rules constitute the activity in such a way that the activity would not exist without the rules; there is no game of chess or football without the rules. Correspondingly, in the context of prudential value theory, those things that are constitutively prudentially valuable constitute what is good for us, such that there is no such thing as what is good for us separate from the constitutive goods.

Of course, constitutive prudential value is only one kind of constitutive value. The invocation and definition of constitutive value is somewhat inconsistent. Some take constitutive value to be contrasted with intrinsic value, but this seems partly down to a failure to distinguish intrinsic from final value. Russ Shafer-Landau (2007, p.281), for example, distinguishes constitutive value from instrumental value, and contrasts both with intrinsic value. Mark Schroeder (2016, §2.1.1) explicitly treats constitutive value as a form of instrumental value, in contrast to intrinsic value, thus precluding the possibility of constitutive value being a form of final value. Allan Hazlett, who *does* adopt Korsgaard's distinction between intrinsic/extrinsic value and instrumental/final value, takes constitutive value to be a form of final value. Hazlett (2013, pp.14–16) argues that knowledge is partially constitutive of well-being, so it is a constitutive prudential value. It doesn't constitute well-being insofar as it causes well-being, or leads to well-being, rather, the pursuit of knowledge is a basic form of good, and so is itself good for us. However, it's not the only thing that is good for us, nor is it always good for us. Clearly constitutive value cannot be a form of intrinsic value—something that is constitutively valuable is valuable not in virtue of its intrinsic properties, but in virtue of its relation to the whole, an extrinsic property. But constitutive valuable *might* be a form of final value. That is, something that is constitutively valuable it might be valuable for its own sake, because pursuing and achieving it is part of what it is to pursue and achieve the good. Even though there is *some* conception of the supervenient good in question, it is not separate or separable from its constitutive parts.

Understanding objective goods in terms of constitutive prudential value seems to be the best way of making sense of perfectionist objective list theories. Constitutive prudential value captures the idea that objective goods (such as knowledge, friendship and play) are 'an aspect of' or 'a basic form of' the good, which is conceived in terms of authentic human flourishing (Finnis 1980). If we understand constitutive

prudential value as a form of final value, then it also captures the sense that objective goods are ultimate values, or ends (Griffin 1988, p.68). The idea here is not that pursuing or achieving objective goods *leads to* flourishing, or *causes* us to exercise our characteristically human capacities; pursuit and achievement of these goods *just is* what it is to flourish and to exercise our human capacities. For the perfectionist, human flourishing is not instrumental to or constitutive of prudential value, it *simply is* prudential value. The relationship between objective prudential goods and well-being is less clear for other objective list theorists, although understanding the items on the list as constitutively valuable with respect to well-being seems to make best sense of them. Fletcher, for example, sometimes describes the items on the list as *promoting* well-being (2015, p.148) and at other times as *contributing* to well-being (2013, p.206). He makes it clear that the items on the list are not instrumentally valuable, but the idea of *promotion* in particular seems somewhat in tension with this. This tension is, I think, best explained by taking the objective prudential goods to promote or contribute to well-being insofar as they are individually only part of well-being, but nonetheless understanding this contribution to consist in their partly *constituting* well-being.

To cap off this discussion of constitutive value, and before moving on to talk about the identification and measurement of the objective prudential goods, I want to say a few words about explanation. Another way of understanding what it means to say that objective prudential goods are of final prudential value is that there is no further explanation to be given as to why they constitute well-being. Finnis, for example, claims that to say that knowledge is a basic good is simply to say that ‘reference to the pursuit of knowledge makes intelligible...any particular instance of the human activity and commitments involved in such a pursuit’ (1980, p.62). No further explanation can be given as to why knowledge is good for us: reference to knowledge *is itself an explanation* for why some activity or pursuit is good for us. For the non-perfectionist

objective list theorist, it is a brute fact that friendship, pleasure, knowledge, are good for us (Kitcher 1999, p.59). There is no further explanation. Of course, the objective goods might also be instrumentally valuable towards some other ends, and it might be possible to explain their value in reference to these, and to convince others of their instrumental value insofar as they appreciate the final value of these other ends. However, an explanation of instrumental value cannot be an explanation of final value. That I believe friendship to be valuable because it leads to happiness, stability and health cannot alone be a justification for believing friendship to be of final value. This is not to say that an objective list theorist can't successfully convince people to agree with her characterisation of the prudential good by entreating them to examine a life without one or more of the items on her list, for example, or by giving a particularly evocative account of the good life, which includes the items on her list. But the success of such an account will always depend on their interlocutor endorsing the final value of the prudential goods.

Fletcher (2013) suggests that objective list theories of well-being are merely enumerative, and not explanatory, that is, they tell us what is good for us, but not why it is good for us. The prudential goods *just are* good for us, and their prudential value cannot be further explained. The perfectionist, of course, *does* have an explanation as to why various objective prudential goods are good for us—their pursuit is constitutive of human flourishing. However, for the perfectionist too, explanation has to stop somewhere. Ultimately, human flourishing just is, by definition, what's good for us. To flourish is to instantiate the life form of one's species, and the life form of one's species fully determines one's good. As I'll discuss in the following section, the perfectionist might give some kind of a background story to help explain why we should accept a naturalist account of well-being, such that well-being is understood in terms of the kind of creatures that we are. However this doesn't obviate the explanatory stopping point. A rich naturalist background picture only provides support for the

neo-Aristotelian account if one is already convinced by the asserted relationship between the life form of one's species and the good. It is a brute fact, for the perfectionist, that human flourishing is good for us and, therefore, that well-being must be understood in terms of human flourishing. Just as the non-perfectionist objective list theorist can't get behind the question of *why* the prudential goods are good for us, the perfectionist can't get behind the question of *why* flourishing is good for us.

It is the *finality* of the prudential value, both for the perfectionist and the pluralist objective list theorist, which means that explanation bottoms out here, and not elsewhere. The fact that the objective goods are prudential ends means that further explanation of their contribution to well-being won't add any content which isn't provided by the claim that the goods have constitutive prudential value, and is not likely to convince anyone who isn't already sympathetic with the characterisation in question.<sup>32</sup>

---

<sup>32</sup> That explanation must stop at a certain point does not necessarily open up objective list theorists to claims of arbitrariness. Railton (2003) argues that non-hypothetical elements in reasoning can be non-arbitrarily identified using constitutive arguments. A constitutive argument seeks to justify normative requirements by showing that we are already committed to them insofar as we are believers or agents (Tubert 2010). If a normative requirement can be shown to be constitutive of agency, then there may be certain questions that are unaskable insofar as we are, and take ourselves to be, agents. Railton focusses on questions like 'Why should I be rational?' or 'Why should I aim at the good?' but there is plausibly something similar to be said in response to analogous questions relating to prudential value: 'Why should I seek to act in accordance with my characteristically human capabilities?' or 'Why should I seek to acquire knowledge?' If a constitutive explanation can be given in answer to these questions—if acting in these ways is constitutive of agency, and one wouldn't be an agent should one fail to act in these ways—then explanation may non-arbitrarily bottom out at this point. Nonetheless, constitutive

### 3.3 *Identifying objective goods I: Individuating the goods*

Most objective list theories focus on the individuation of the relevant goods. However, in order to identify whether someone is doing well, and to say what it would be for them to do well, much more needs to be said about the nature, interaction and measurement of these goods. Individuating a set of objective prudential goods does not alone amount to a theory of well-being; much of the philosophical substance—not to mention complexity—lies in the detail.<sup>33</sup> In this and the following section I will set out four elements that I take to be involved in the identification of objective goods: first, individuating the objective prudential goods; second, specifying how these goods relate to one another; third identifying the indicators for the goods; and fourth, identifying the metrics for the indicators. Each of these steps raises distinctive philosophical questions.<sup>34</sup> In this section, I focus on the highest-level identification of the objective goods.

---

arguments may be less convincing in the case of some prospective prudential goods than others—merely claiming that something is constitutive of agency doesn't make it so. Furthermore, perfectionist theories may be advantaged in this respect, due to the close relationship they invoke between rationality and the good (see Foot 2001, for example).

<sup>33</sup> Some objective list theorists deny or fail to recognise that an adequate objective list theory requires anything other than identifying the basic goods at the broadest level. For example, Christopher Rice (2013, p.201) suggests that a 'complete explanatory theory of well-being' is one which '[specifies] a full list of basic goods.' Rice *may* intend that a full list includes specification of interaction between the goods, indicators of the goods, metrics for the indicators, but he gives no indication of this. Others acknowledge the need to include detail about the weighting and interaction of goods, but do not provide it (Fletcher 2013, p.214).

<sup>34</sup> The identification thesis that I outlined in chapter 2 asserted that any theory of well-being needs to say something about the identification of well-being in people. While I develop and discuss this four

First, then, the objective list theorist must individuate the objective goods which constitute well-being. Selecting the goods will require some kind of reflection on the nature of well-being, of lives which we consider to have well-being, and those things which we take to improve or diminish our well-being. One way of justifying the inclusion and exclusion of attitude-independent goods is by appealing to pre-theoretic judgements or intuitions about what is good for us (Fletcher 2015, pp.152–3). Fletcher (2013, p.218–19) suggests a couple of heuristics for determining whether particular items should be on the list or not. First, he says, we can imagine two people with identical bundles of goods, who have equal well-being. Then imagine that one of them has a purported well-being enhancer added to their bundle. If their well-being is thereby made unequal, this provides evidence that the well-being enhancer should be added to the list of goods. If their well-being remains identical, this should be taken as evidence that the good is not in fact an attitude-independent prudential good, and it should not be added to the list. Furthermore, the proposed attitude-independent goods need to be assessed in relation to one another to ensure they have final, and not just instrumental, value. If it can be shown that one of the potential goods is covered by another good on the list, or is only valuable insofar as it leads to another good on

---

part identification matrix in the context of objective list theories, I also take similar steps to be necessary for developing any theory of well-being. Of course, the specification of how the prudential goods relate to one another won't be a relevant consideration for monistic theories of well-being, but I do take the high-level identification of well-being, the identification of indicators of well-being, and the identification of the metrics for these indicators to all be necessary components of any theory of well-being.

the list, then it can be removed. Such exercises lean heavily on intuitive judgements about well-being and its constituents.<sup>35</sup>

However, these exercises don't tell us much about why the goods in question are on the list, or give us any justification for their inclusion. Even if there is no explanation for why the goods are good for us, a convincing objective list will require some kind of explanation of how the goods are determined, that is, how it is known that *they* constitute the good, rather than some other set of goods. Fletcher appeals to our *intuition* about whether particular prudential goods really are of final prudential value as a mechanism for determining what stays on the list. Given widespread disagreement about the nature of well-being, even amongst objective list theorists, a little more needs to be said to explain how an appeal to intuition can justify the selection of a particular set of goods. Jukka Varelius (2004, p.75) argues that objective list theorists must either take the items on the list to be intersubjectively regarded as good, or to be good in a stronger realist sense. Intersubjective agreement about value looks, at first glance, to fail to meet the attitude independence condition, which characterises objective list theories. If what makes a good intrinsically prudentially valuable is the very fact that a group of people intersubjectively agree that it is intrinsically valuable, while its prudential value may not depend on any particular individual's attitudes towards it, it *does* depend on a set of people's attitudes towards it.

However, intersubjective agreement may be taken to play an epistemic, rather than a constitutive role: while the goods are identified by appeal to intersubjective agreement, their value does not derive from this agreement. Indeed perhaps there is intersubjective agreement as to their prudential value precisely because they really are

---

<sup>35</sup> In the following two chapters I discuss cognitive biases such as framing and focussing effects at some length. It seems as though intuitive thought experiments of this kind are liable to be affected by framing and focussing effects too, but I won't spend time exploring this here.



prudentially valuable. Identifying prudential goods as objectively prudentially valuable in this stronger, realist sense requires some kind of explanation or justification that does not depend on intersubjective agreement. Some philosophers appeal to the 'self-evident' value of goods by way of explanation. Finnis, for example, takes it to be self-evidently true that knowledge is a prudential good. He specifies what he means by this: '[i]t cannot be demonstrated, but equally it needs no demonstration' (1980, pp.64–5). This doesn't mean that everyone recognises the value of knowledge—although Finnis also seems to think that its value is 'obvious,' which perhaps suggests that it is easy to come to such a recognition. He does not think that the value of knowledge—indeed any value—can be inferred from any facts about the world (Finnis 1980, p.66). This seems to entail a form of normative realism about the basic value of knowledge—the prudential value of knowledge cannot be denied, and the truth of the proposition that knowledge is good for us is not a matter of convention or opinion (Finnis 1980, p.69).

A claim is self-evidently true when adequate understanding and consideration of it is sufficient to justify believing it (Shafer-Landau 2005, p.247). Given disagreement about what is good for us, and about what items are on the list, what does it mean to say that something is 'self-evidently' good for us? Robert Audi (2015) points out that the self-evidence of a proposition or judgement does not imply its intuitiveness. The proposition that a child can be borne by its grandmother, for example, is a self-evident but potentially counterintuitive claim (Audi 2015, p.65). Furthermore Audi argues that self-evident judgements and propositions, even when intuitive, are not beyond rational disagreement. Someone may rationally disagree with a self-evident proposition or judgement, even though they adequately understand it, for example if they adhere to a thorough-going scepticism, or if they are committed to a theory which implies the negation of the self-evident proposition (Audi 2015, p.68). So, for the objective list theorist, the self-evidence of the presence of the goods on the list need not imply absence of disagreement.

However, it is nonetheless difficult to justify claims of *normative* self-evidence without resort to intersubjective agreement. For Audi's example of the proposition that a child can be borne by its grandmother, its self-evidence is definitional. Your grandmother is the mother of your parent, and it is possible for someone to be both your mother and the mother of your parent. The standard of assessment for determining the meaning of the word 'grandmother'—for evaluating whether a particular use of it is right or wrong—is fairly well fixed. We might point to dictionary definitions, or to very high levels of consistency in the use of the concept, for example. However, in the case of many normative judgements or propositions—such as Finnis' claim that knowledge is good for us—their contentiousness partly lies in the fact that the standards of assessment for determining their truth are under dispute. This often comes down to disagreement about the definitions of the terms involved, which may result in inconsistent use of key terms—'knowledge,' 'good'—or in there being no single clear authority to appeal to for resolution.

The self-evidence of syllogisms, tautological propositions and (at least some) mathematical claims is grounded in either the presence of a very stable and widely recognised standard for assessing their meaning, or the propositions themselves being a standard for assessing the truth of other propositions. Their self-evidence *is* grounded in intersubjective agreement, but at the level of the standard of assessment, rather than at the level of the claim itself. For normative propositions, on the other hand, without such a shared standard of assessment of what counts as evidence for self-evidence, intersubjective agreement at the level of the proposition becomes important. Some of the rhetorical force of claims of self-evidence in logical and linguistic contexts in which there *is* a stable standard of assessment inevitably carries over into the normative sphere, but there is something lacking in the normative case, namely, a standard against which to assess the adequacy of people's understanding and comprehension of purportedly self-evident moral claims. This makes intersubjective

agreement about the truth of normative claims central to justifying claims of normative self-evidence, and it seems plausible that intractable disagreement about the truth of a normative claim should provide evidence against its self-evidence. Given this, it would be philosophically irresponsible to use claims of self-evidence to play a justificatory role in assertions about the truth of normative claims. This calls into question the validity of claims of self-evidence as justification for the identification of prudential goods as objectively prudentially valuable in a realist sense.

Some theories of well-being, notably neo-Aristotelian perfectionist theories, provide what Varelius (2004, p.77) calls a 'background story' to help motivate claims of self-evidence. This will be some kind of story about the nature of prudential value, which will prime readers to accept claims of self-evidence. Philippa Foot (2001), for example, argues that what is good for human beings is determined by their nature, and this is determined by facts about the kinds of beings that we are. Evaluative claims about what is good, or what is good for us, can be identified by considering the facts about human lives (Varelius 2004, pp.78–9). Accepting this naturalist story about the nature of prudential value (and value in general) justifies the perfectionist's appeal to characteristically human capacities or human flourishing as intrinsically valuable objective goods. However, the presence of a background story doesn't fully jettison the difficulties involved in making claims about the self-evident value of prudential goods, though it may add some rhetorical force to such claims. In order to accept human flourishing as an objective prudential value, in a strong, realist sense, the perfectionist explanation of why this is good has to stop somewhere. To go any further requires agreement that pursuing human flourishing, or pursuing things that it is in our nature to pursue, just is intrinsically good for us.

This discussion suggests that it will be difficult for the objective list theorist to justify the identification of the goods without some dependence on intersubjective agreement. However, for the remainder of this chapter, I will shelve these concerns

and assume that the objective list theorist *can* identify a set of goods which are widely agreed to be intrinsically good for us, and whose prudential value is not taken to be dependent on intersubjective agreement, but to be real in some attitude-independent sense. Even if the objective list theorist *can* secure attitude independent specification of the prudential goods, I will argue, they *cannot* secure attitude independence in the specification and measurement of these goods. I return to the discussion of normative realism and intersubjective agreement in chapter 6.

### 3.4 *Identifying objective goods II: Specifying the goods*

Knowing the goods does not alone amount to a full theory of well-being, as it doesn't suffice for knowing whether someone is doing well or not. Knowing that knowledge, friendship, and achievement, for example, are good for someone doesn't alone enable me to work out whether they are doing well, or how well they are doing compared to other people—an assessment is needed to determine whether their lives possess or embody the prudential goods or not. This will involve a detailed account of the good: not just the headline statement that knowledge is good for us, for example, but a specification of what counts as knowledge and what doesn't. It also involves some account of how to identify the features of people's lives that instantiate the prudential good—what I call *indicators of the good*. Determining how well off someone is compared to other people will involve ranking states of prudential goodness. There is no reason to think that this need be a complete ranking. While a theory of well-being might be unsatisfactory if it was unable to compare any states of well-being, it would be excessively demanding to require that any two states of well-being be comparable in every theory of well-being (Sen 1987, p.198). Indeed, a theory's claim to be able to do so might set alarm bells ringing, indicating a failure to appreciate the prevalence of hard cases in well-being valuation and comparison. Nonetheless, a partial ranking seems an appropriate demand—a theory of well-being needs to be able to tell us some

of the time, and perhaps much of the time, which of two people, or two states of the same person, has greater well-being.

Merely knowing the set of goods may take us some distance with respect to a partial ranking, given a fairly innocuous assumption. A simple dominance relation would assert that if lifestyle  $x$  has more of one of the prudential goods and no less of any of them, in comparison to lifestyle  $y$ ,  $x$  has higher well-being than  $y$  (Sen 1987, pp.4–5). Given the fairly uncontroversial acceptance of such a rule, the statement of the set of prudential goods generates a very skeletal ordering of well-being states, which Sen calls a ‘partial dominance ordering’ (1987, p.5). However, such a ranking would be extremely minimal—too minimal to be of much use. For example, if a high-flying executive has more achievement in her life than a barista, but less friendship, the two lives will not be ranked on the partial dominance ordering, and instead will be treated as prudentially incomparable or equal. However, even the possibility of a partial dominance ordering depends on being able to identify the presence and extent of the goods in people’s lives. This means that some detail about the indicators of the good, and a detailed specification of the good, is needed to get even this far. In order to determine whether someone is doing well, then, the objective list theorist will need to add some detail to determine how the goods relate to one another and to identify when a life should be counted as having one or more of the goods. The detail that the objective list theorist has to provide can be divided into three distinct activities, which I outline in this section: first, determining the relationship between the goods, if more than one has been identified; secondly, identifying indicators for the goods; and finally, identifying metrics for the indicators and the goods.

First, then, the objective list theorist must determine the relationship between the goods on the list. Take, for example, Fletcher’s objective list (2013), which contains achievement, friendship, happiness, pleasure, self-respect, virtue. The objective list theorist must detail how the goods are to be weighed against one another in order for

the list to function as a means of assessing well-being. For example, he will have to explain whether all of the items are equally good for us. Do friendship and virtue contribute to our well-being equally to pleasure, for example, or does virtue contribute to well-being more than the others? Is there room for individuals to specify the relative importance of the goods, or is this fixed? Furthermore, he will have to determine whether someone has well-being only if they have some amount of all of the goods in their life, or whether they can be doing well if they have only five out of the six, or only four out of the six. Perhaps there is one good which is alone sufficient for well-being, whereas in its absence several of the other goods must be present in combination in order for someone to be said to be doing well. The objective list theorist must moreover address how much of each good is needed for someone to be doing well—will any amount suffice, or is there a threshold level of each good which must be met? Are any of the goods lexically prior to the others, such that even a small amount of one is better than a large amount of another good? Are the marginal returns for all or any of the goods diminishing? Given a single list of objective goods, there are an indefinite number of ways of specifying their interaction, and the way in which they relate to assessments of well-being.

The perfectionist objective list theorist has questions of her own to answer. She must explain whether any exercise of characteristic human capacities is good for us, or whether the capacities must be exercised well (Bradford 2015, pp.127–8). She must also tell us if it matters whether we achieve outputs which manifest the capacities—that is, whether we actually acquire knowledge, friends and virtues—or if it only matters that we strive towards them. Furthermore, questions arise as to whether assessment of the achievement of outputs or the exercise of capacities are calibrated on an absolute scale, or one that is relative to an individual's natural capacities. I do not intend to suggest that these are issues and questions that are unanswerable by objective list theorists, or even particularly difficult to address. They are, however,

issues which must be addressed in order for an objective list theory to be able to tell us whether someone is doing well or not, and which cannot be answered by merely stating the items on the list. A theory that tells us what is good but doesn't tell us how the goods relate to one another doesn't give us a complete theory of well-being (Bradley 2009, p.16). Moreover, different answers to these questions will result in quite different theories of well-being.

Once these initial specifications for the contribution of objective goods to well-being have been determined, the objective list theorist needs to identify indicators of the good. An example will help to spell out exactly what I mean by an indicator. Suppose friendship is an objective prudential good. How do I go about identifying whether someone has friendship in their life, and so satisfies (some of) the conditions for well-being? Knowing this will require some specification of the objects, events, activities and states of being that are indicative of friendship. These are indicators of the good, the presence of which in someone's life denotes the presence of the good. They might also be thought to be the means by which the good is realised in someone's life. Indicators of friendship will likely include personal relationships, contact, correspondence and communication with others, mutual affection, relationships which involve loyalty, trust, and freedom from judgement, and so on. The number of personal relationships someone has may be important for assessing the extent to which their lives contain the prudential good 'friendship,' up to a point, as may amount or frequency of contact. However just measuring the number of friends a person has (or says they have) and the amount of time they spend with them is clearly inadequate for assessing friendship. Some accounts of friendship might not take the number of friends to be an indicator of friendship at all—one true friend may be all that is needed. Measuring the degree of loyalty, trustworthiness and affection shared between people seems essential for determining whether they are friends, and yet these qualities are not easily identifiable and quantifiable, unlike counting the number of friends and

measuring the amount of time spent with them. Specifying the indicators of a prudential good might be seen as part of what it is to identify the prudential good itself. However, the specification of the indicators is likely to be controversial and ignite disagreement, even among those who agree on the prudential value of the good in question, and different indicators might be relevant in different contexts. It is easy to agree that friendship is good for us, far more difficult to agree precisely what friendship entails.

The identification of indicators of the good also involves some specification of their relation to the good. Suppose I have a friend who has behaved in ways that I consider to be emotionally manipulative and exploitative. How do I assess whether this is indeed an example of friendship, and whether the relationship contributes to my well-being or not? One way of doing this might be to explore whether the relationship has any of the typical characteristics of friendship—mutual affection, honesty, loyalty, compassion, freedom from judgement, and so on. It may well be that my manipulative relationship exhibits some of these indicators, some of the time, but not all of the time. Determining whether this is indeed an example of friendship requires knowledge of whether enough of the indicators of friendship are present, or are present in appropriate combinations, or whether there enough of each of them present, to in fact indicate that this is a friend. We all fail to meet ideal standards of friendship occasionally, but this presumably need not invalidate all of our friendships. On the other hand, occasional glimpses of loyalty and affection do not alone make a friendship. Furthermore, not all cases of loyalty are indicators of friendship—consider loyalty to an employer or to a football team, for instance. Other indicators of friendship may need to be present—the presence of an interpersonal relationship, for example, for loyalty to be activated as an indicator of friendship. Similar questions arise for other potential prudential goods. Identifying the indicators of the prudential good thus



requires some specification of the relationship between the indicators and the good, and the way that the indicators relate to one another to indicate the prudential good.<sup>36</sup>

Once the indicators of the good have been decided, there remains the task of determining whether a particular entity, event, or pursuit is in fact an example or indicator of the prudential good, that is, deciding whether a life contains these indicators or not, and to what extent. This will require some kind of measurement or assessment of people's lives, to determine whether they contain the indicators of the good or not. Specification of a normative principle, such as 'knowledge is good for us' or 'achievement is good for us,' imposes informational constraints, that is, evaluating it entails the use of some types of information, and the exclusion of other types. Sen argues (1985b, pp.169–70) that if two objects belong to the same *isoinformation set*, they must be treated in the same way. This means that if they are the same in all relevant ways, they must be treated the same. In the case of well-being, this means that if two lives—or two different time-points within the same life—have the same levels of all prudential goods, they must be understood to have the same level of well-being. Sen calls this an *invariance requirement*. Of course, as two factual circumstances are never identical, only those aspects of lives that are relevant to the principle in question contribute to the assessment of invariance (1985b, p.170). Making such an assessment requires a characterisation of the objects in question, in order to determine their informational content.

---

<sup>36</sup> One way of assessing whether a particular relationship is an example of friendship would be to assess whether that relationship is a net contributor to or detractor from my well-being. However this seems to make its membership of the category 'friendship' explanatorily impotent with regards to my well-being, and to presuppose some way of assessing or measuring well-being independently of the goods. This would be inconsistent with a constitutive reading of the relationship between the prudential goods and well-being.

Determining whether an event or entity is an example of the prudential good, or an indicator thereof, may thus require the development and use of metrics. So, for example, if I am trying to determine whether my studying for a PhD has contributed to my knowledge, I might first try to assess whether it embodies any of the typical indicators of knowledge: increased understanding, perhaps, or the attainment of certain skills, such as being able to explain things. Determining whether my study of philosophy has these indicators requires some kind of standardised form of assessment to determine whether I have understanding, and whether I am able to explain things. The assessment must be standardised in order to be able to describe my abilities and state at different points in time, and to be able to compare the results. If it isn't standardised, then while I may in some sense be able to assess my well-being at one point in time, the result will be completely isolated, as it will be incomparable with well-being at any other point in time or in any other people. It's therefore unclear what exactly it would reveal about how I am doing in any meaningful sense.

So it might be that assessment of increased understanding in philosophy is determined by the agreement or assent of the set of relevant agents: my students, peers, supervisors, and examiners, for example. In this case, assessing whether my study of philosophy brings me increased understanding, and thus determining whether it contributes to my knowledge, involves identifying who the relevant set of agents is, and what constitutes agreement between them. Perhaps the assent of just one person is not enough, and the assent of all of the relevant agents is too much to ask for, but some degree of agreement between relevant parties is enough to verify my state of understanding. The invocation of 'metrics' here does not necessarily imply numerical quantifiability, but it does imply comparable quantifiability, that is, it must be determinable whether one object has more or less of the relevant good than another, even if there is no way of giving an exact amount by which one is better than the other. This won't necessarily require that indicators of well-being are measured on a

numerical scale, or that all indicators of well-being can be numerically represented. However, it will require some means of assessing whether or not they are present, and, at least in some cases, providing an ordinal ranking of different people, or the same person at different times, with respect to these indicators.<sup>37</sup>

Identifying a list of high-level objective goods only gets an objective list theorist so far in developing a theory of well-being. In order for such a theory to be able to identify when a particular person is doing well, to compare the well-being of two people, or to say which out of various options would be better for someone, far more detailed identification of the goods and specification of their measurement is necessary. Without such detail, an objective list theory will be nothing more than a list of concepts, substances or qualities, without any way of saying whether any person's life includes or fulfils these items in any way.

### *3.5 Specification and attitude independence*

For the remainder of this chapter, I will continue to assume that the objective list theorist can identify a set of goods—or a single good, in the case of the perfectionist—whose prudential value is real in some attitude-independent sense. Given this, I will assess whether the specification of the identification and measurement of the good can be similarly attitude-independent. For this to be true it would have to not only be the case that, for example, knowledge, friendship, and achievement are good for us regardless of whether anyone thinks that they are, but also that the indicators of the prudential goods indicate what is good for us regardless of whether anyone thinks that they do, and that the metrics for such indicators measure those things that indicate

---

<sup>37</sup> Some objective list theories—certain forms of utilitarianism, for example—may demand a cardinal ranking, which would imply the quantifiability and commensurability of all prudential goods. I view such a requirement as overly demanding, and see no reason to think that cardinal ranking is a necessary requirement of a theory of well-being.

what is good for us regardless of whether anyone thinks that they do. In this and the following section, I argue that even if we accept that attitude independent prudential goods can be identified, determining whether someone is doing well or not requires a series of decisions to be made about the specification of well-being, which inevitably depend on the context of well-being assessment and the purpose and attitudes of those people making the assessment. That is, even if the identification of objective goods is attitude independent, the specification of them is not.

I slightly shift my focus in this section to those perfectionist theories that identify not only the high-level goods, but also mid-level indicators of the good, focussing on the capability theories of Nussbaum and Sen. Unlike many objective list theories of well-being, these theories acknowledge and explore the need for a theory of well-being to include discussion of application and operationalisation, and not just specification of the good or goods in the broadest sense. These theories accordingly give a good indication of what the objective list theorist's attempts to specify and measure well-being will look like. In the following section, I return to other objective list theories.

Nussbaum's neo-Aristotelian perfectionist theory of well-being takes well-being to be a matter of securing for people the conditions in which each of them, as individuals, are able to choose to live a flourishing life (1988). Securing well-being is not about securing commodities, but about enabling people to function in ways that are distinctively human. Nussbaum seeks to identify a list of human functionings—ways of living and acting—that people should be able to choose. She stipulates that an objective valuational procedure should be used to assess and criticise proposed functionings, to ensure that they do not reflect people's unreflective preferences, which may be distorted by discriminatory and unequal circumstances. The list of objective functionings is derived by considering those conditions of human life that are common to all societies: that we are mortal, that we have bodily desires that are difficult to control and to satisfy; and that there is scarcity of material resources, for

example. The list endeavours to capture those functions without which we would not be human at all.

Nussbaum (1992) ends up specifying a list of ten functional capabilities. Capabilities are the functionings that people are able to choose to pursue and achieve, rather than the ones that they in fact achieve; Nussbaum takes the *choosing* of the functionings to be crucial to their value, and wants to leave space for people to choose not to pursue or achieve some functionings. She argues that such a list allows for multiple realisability, for example across different cultural contexts, with recognition that competing conceptions represent the same underlying functioning. Thus, functionings are attitude independent insofar as the selection of the functionings is not a matter of preference, but tied to what it is to be a human being. Nussbaum calls her account a ‘historically grounded empirical essentialism’ or ‘internalist essentialism’ (1992, p.208) capturing the sense in which it is historically sensitive and grounded in the physical and cognitive reality of what it is to be human, while at the same time not making well-being straightforwardly dependent on human desire and preference.

Sen’s theory is similar to Nussbaum’s at first glance, but harbours important differences. Like Nussbaum, he takes well-being to comprise the achievement of functionings, that is, activities and states of existence which capture how a person can function (Sen 1985b, pp.197–8). Sen claims that how well a person is must be a matter of what kind of a life he is living and on what he is succeeding in doing or being. Functionings capture these doings and beings. Nussbaum’s list of functional capabilities comprises high-level goods—life, bodily health, freedom from pain, being able to form attachments to other humans, having political and material control over one’s environment. For Sen, however, the relevant functionings are a large and unconstrained set of components, which may be as varied as eating, being well-nourished, reading, going to the park, being free from disease, accessing cultural goods

and choosing one's own education (Sen 1985b, pp.197–8). They can be broad or narrow, and can overlap. The functioning state 'being well-nourished,' for example, largely coincides with the functioning activity 'eating,' though eating is broader, covering social and cultural achievements—breaking bread with family and friends, participating in religious festivals—as well as nutritional ones (Sen 1985a, pp.25–6). There may also be some states of being well-nourished which don't overlap with eating—for those sustained via a drip or feeding tube, for example. Sen, notably, does not attempt to produce a list of functionings. Nussbaum (1992, p.222) recognises the complex dependencies between her capabilities, but they are decisively separate. Sen's set of functionings, on the other hand, is much more tangled, permitting overlapping concepts and partial re-descriptions.

Sen explicitly considers how to go about evaluating someone's well-being, whereas Nussbaum is more concerned with defining indicators of well-being (Sen 1985a, pp.46–52). For Sen, assessing someone's well-being requires a valuation of the set of functionings that they have actually achieved—their 'functioning vector.' This has two parts: first, it requires the specification of the functioning set. This involves a description of the relevant functionings, and the selection of metrics which capture the functionings. So, for example, a reading functioning might be captured by a measure of literacy, a measure of quantity of material read in a given time period, a measure of variety of material read, and so on. There is no fixed specification of the kind of data used to value functionings; different standards of measurement and types of data will be appropriate for assessing well-being in different contexts. Sen argues that information about functionings will typically come from direct observation of individuals, rather than, for example, market purchase data or information about the commodities to which they have access. This is because market purchase data is an indirect guide to functionings. For example, knowing how much food someone has bought does not tell us much about whether they are well-nourished—for this we

would need to know about food distribution within the family unit, the presence or absence of parasites, height, weight, sex, whether pregnant or lactating, metabolic rate, and so on (Sen 1985a, pp.25–6). However, sometimes market purchase data can be a rough and indirect guide to functionings in the absence of more direct data.

Secondly, once the set of functionings has been identified, its relative value, with respect to other sets of functionings needs to be determined. Different functioning sets are ranked against one another either by the person whose well-being is in question, or with respect to some shared social standard (1985b, p.198). One functioning set will often dominate another, in such comparisons, but this valuation exercise necessitates only a partial ordering of functioning sets, and not a complete one (1985a, pp.31–32). It seems that Sen thinks there will be quite a large degree of agreement about the key constituents of well-being, at least enough to find some agreement in the valuation functions of individuals, and enough to construct meaningful social valuation functions. However, he allows the possibility of personal valuation functions which differ from social valuation. As well as direct observation, market purchase data and utility data can also be used to inform personal and individual valuation functions (1985a, p.48).

The obvious question, perhaps, is whether these theories really meet the attitude independence constraint: do they really provide an objective account of well-being? Nussbaum, although her theory admits cultural differences, has a *prima facie* defence against claims of non-objectivity. Our most basic functionings and limits as human beings are not a matter of personal preference—our mortality, our need for food, shelter and social interaction, our sexual desire, and the affiliation that we feel towards other humans are basic features of the kinds of being that we are. Without these needs and capabilities our very nature would be called into question. Undoubtedly, though, Nussbaum's admission of multiple realisability raises some questions for her as a proponent of objective well-being. The experience and expression of many of her

capabilities is shaped by social and cultural context. This will be most obviously true of the explicitly social capabilities—the capability to laugh, play and enjoy recreational activities; the capability to live for and with others; the capability to engage in critical reflection about the planning of one's own life. However, Nussbaum also grants that some bodily capabilities—to avoid pain and to have pleasurable experiences; to have good health—will be culturally shaped as well (1992, pp.217–8).

It is clear, then, that on Nussbaum's view it would be true to say that what it is to do well in one cultural context is different from what it is to do well in others, and these differences are dependent on the attitudes of its members. The notion of 'attitudes' here is broadly construed, and includes beliefs and conceptual frameworks, as well as desires. Non-objective list theories of well-being are typically conceived of as desire-based theories, but the accepted definition of objective well-being is that what is good for us is good for us independently of our attitudes towards it. This seems to preclude mental attitudes other than desires—in particular beliefs, but also certain emotional states, such as hopes and fears—as the basis for well-being. For Nussbaum it is only under high-level descriptions that the contribution of particular capabilities to well-being is not attitude dependent. Perhaps the most she can say is that *under some descriptions* what it means to do well is independent of our attitudes towards it. It should be noted that Nussbaum doesn't argue for multiple realisability at the personal level, just at the cultural level, so an individual's personal attitudes are not, for her, determinative of well-being. The attitude-dependence of the capabilities, and thus of well-being, must be in some sense larger than the individual—it must refer at the very least to collective attitudes and ways of life which aren't just a matter of personal endorsement. However, Nussbaum doesn't do much to define the notion of a culture, so the extent of the multiple realisability is unclear—if the concept of the last surviving member of a culture is meaningful, then the possibility of a culture with one member



is live. Nonetheless, she does in general allow for interpersonal or collective attitudes and endorsement to shape the capabilities.

Sen also provides a defence of the objectivity of his view. He argues that objective moral principles are compatible with *positional variance*. That is, that the moral evaluations of different persons can vary with the different positions that they respectively occupy in the given state of affairs (1985b, p.174). So, for example, if I am responsible for a particular action, my moral evaluation of it will—must—take into account my agency. The moral evaluation of a third party need not take into account their own agency, but they might say “I can see that if I were in your position I would think differently.” Positional variance is different from subject variance or authorship variance—in order to remain objective, it must be the case that anyone in the same position would come to the same conclusion (Sen 1985b, pp.183–4). Therefore, even though Sen’s view admits many competing social and personal valuation functions, he thinks this represents only positional variance. This means that the different resultant accounts of well-being are perspectively objective. Sen (1985b, p.196) rejects T.M. Scanlon’s definition of an objective criterion—that it provides a basis for appraisal of a person's well-being which is independent of the person's tastes and interests—and instead argues that insofar as individual’s tastes and interests make an objective difference, they should be built into a well-being evaluation function.

The objectivity of this account is debatable. Indeed if this is the definition of objectivity, it’s difficult to see what a subjective theory of well-being would look like. Sen develops his account of positional variance in the context of moral valuation more generally (1985b, pp.183–4). In this context the position of an agent in relation to a case makes a difference in their moral valuation of the case—it is not their authorship that is relevant, but only their position. This, says Sen, means that the ‘positionality of moral valuation is perfectly consistent with objectivity of moral values’ (1985b, p.184). However, the reasoning here doesn’t necessarily carry over to well-being. Even if Sen

is right that positional variance with respect to moral valuation is consistent with the objectivity of moral values, it is not clear that positional variance with respect to well-being valuation is consistent with the objectivity of well-being criteria. This is because Sen allows that someone's evaluation of well-being is dependent on their tastes and interests: their religious views, personal goals and projects, preferences, beliefs, dispositions and desires, for example. He claims that well-being is objective because if another agent were in this position—if they had these tastes and interests—they would come to the same opinion about well-being. However this blurs the line between positional variance and subject variance. Our beliefs, goals and dispositions don't just represent the position each of us is in, but the subject that each of us is. Sen ends up saying something like "if I were you, I would have the same conception of well-being as you." It's difficult to see how this is objective in any meaningful sense at all. Unlike in the moral case that Sen describes, it is internal attitudes which are determinative of individual valuation of functioning vectors, and not just external factors.

So there are question marks over the objectivity of both Nussbaum and Sen's theories of well-being. The question for other objective list theories, then, is whether they can specify the prudential goods in such a way as to avoid these concerns. I think they largely cannot. Even if the objective list theorist rejects Nussbaum's culturally relative specification of the prudential goods and Sen's positional variance, the specification and measurement of objective goods that is required for an adequate theory of well-being leaves them highly context dependent, where the context in question includes at least some of the attitudes of those people engaged in the measurement of well-being. This puts significant pressure on the attitude independence of objective list theories.

### *3.6 The attitude dependence of 'objective' goods*

How might the objective list theorist attempt to avoid these concerns? They might first deny the multiple realisability of the objective goods. To do so would be to claim that the prudential goods cannot be significantly re-interpreted in line with personal and

cultural attitudes, and instead, they have the same meaning in every context. This approach is more plausible with respect to some potential prudential goods than others. It might be coherent, for instance, to insist on the cross-cultural uniformity of knowledge. Knowledge, and its acquisition, means the same thing in every context, and regardless of your attitudes to it. Astrology and augury aren't means of knowledge acquisition just because of the claims and attitudes of some people. Rational activity—one's of Parfit's suggested prudential goods (1984, p.499)—might similarly be argued to be context independent, although of course this is widely contested.

Other contenders, however, are less plausibly described as invariant across persons and context. Achievement and virtue, for example, seem liable to inter-cultural differences. What counts as an achievement in a given cultural context will depend greatly on technological development, available resources, existing knowledge, but also on the expectations and beliefs of one's fellows. Happiness and pleasure, more problematically, seem liable to inter-*personal* differences. What counts as happiness or pleasure for a particular individual will be shaped by their goals, expectations, beliefs, and desires. Of course, the objective list theorist might admit that there is variation in the realisation of the prudential goods in different contexts, but argue that insofar as the variation represents different versions of the same prudential good, it is still possible to say that the same thing is good for each of us independently of attitudes towards it. This is similar to Nussbaum's approach, and it's thus likely that the same objection applies to it. That is, this seems to amount to the view that *under some descriptions* what it means to do well is independent of our attitudes towards it. At best this amounts to a significant downgrading of claims of attitude independence.

Furthermore, those descriptions that *are* dependent on the attitudes of individuals and communities seem to be those which are more pertinent to evaluations of whether someone is actually doing well or not, and to producing comparative assessments or orderings of well-being states. That is, it might be possible to evaluate whether

someone has a prudential good in their life at all, in some very crude sense, without looking at the specific indicators and determinants of the good. This might permit the creation of a very partial ordering of lives in terms of the presence or absence of the various prudential goods. However constructing a more complete ordering involves understanding how well people are doing compared to other points in their life, for instance, or compared to other people in their society, and this depends on a more detailed evaluation of the prudential goods. If different individuals and societies have different standards for prudential goods, then clearly there will be multiple accounts of what it is to do well. Of course, these might not be strictly *competing* accounts, if they are tied to particular perspectives and cultural contexts, and thus objectivity might be preserved in some sense. However, given my objection to Sen in the previous section, preserving objectivity would depend on the exclusion of any attitude-dependent standards. So, for example, a society's distinctive account of achievement might be able to vary with respect to the state of technological development and resources, but not with respect to the expectations of other individuals. In the case of at least some prudential goods, this would lead to a pretty implausible restriction on their content. Inclusion of attitude dependent standards, on the other hand, would make variation in the evaluation of well-being dependent on the attitudes of some individuals.

Identifying metrics for the measurement of prudential goods presents a different set of problems for objective list theories, and may leave them with competing, rather than just context-dependent accounts of well-being. Measuring prudential goods involves the identification of an information base—the information that tells us whether someone has knowledge, happiness, bodily health, etc., and how much. However, for any state of existence or activity which might constitute well-being there are many different ways of bringing information about the world to bear on it. The more general the concept in question, the more evident the problem. Knowledge, for example, can

be measured by any number of different tests of aptitude and intelligence, formal examination, peer review, community acceptance, and so on. Each of these will give a different account of knowledge, based on different standards of assessment, and so will lead to a different ordering of people and states, and thus a different evaluation of well-being. Using a composite measure of all of the information sources won't help either. The existing measures of knowledge are not, and do not typically claim to be, the only or the definitive measures thereof. Nor is there any reason to think that collectively they are exhaustive of knowledge, each of them being developed separately and for slightly different purposes.

More localised objective goods exhibit similar effects. One of Sen's key examples is the functioning state of being well-nourished (1985b, pp.197–8). It may appear easier to attain a definitive measurement of specific goods, such as nourishment, than of broader goods. However there are a number of ways of assessing states of nourishment—measures of BMI and weight- or height-for-age, measurement of chronic energy deficiency, measurement of clinical indicators of micronutrient deficiencies, measurement of biochemical indicators of micronutrient deficiencies, and so on (London School of Hygiene and Tropical Medicine 2009, §§). Each of these has a distinct information base and uses requires different measuring instruments. Depending on the quality, type and amount of information available in a particular context, certain measures of the good will be possible and others impossible, and the valuation may be more or less precise, and more or less useful. Sen, for example, suggests that market purchase data might sometimes have to be used as a proxy measure of nourishment in absence of more detailed information about individuals, and where it is difficult, time consuming and expensive to make direct observations of nutritional achievements (1985a, pp.47–8). However, even when the data is not only the best available but also the best known source of information about nourishment, it need not offer an incontestable metric. Different information bases can be severally

relevant to, but not individually or collectively exhaustive of, the measurement of nourishment. Different measures of nourishment may each capture, for example, slightly different things that are all relevant to nourishment, which overlap, but none of which straightforwardly supplants any other. The development of new tools and techniques in the future will inevitably lead to new metrics, using different information bases.

The multiple realisability of any prudential good across information bases suggests a different kind of context-relativity. If different metrics of the good are developed for different purposes, such that each offers the best or most useful measure in different scenarios, then the measurement of the good will be linked to the goals of measurement. So, for example, depending on the particular reasons for measuring how well-nourished someone is, there will be a different information base which provides the best measurement of it. Sometimes there may be several competing information bases and tools which stake a claim to providing the best measurement thereof. This is a problem for the objective list theorist if the goals and aims of measuring well-being are understood in terms of the attitudes of those people making the measurement. Typically non-objective theories of well-being are described in terms of the attitudes of the person whose well-being is being assessed, that is, they are characterised as those views on which whether someone is doing well or not depends on their own desires and preferences. However, if the goals of measuring well-being determine, in part, the appropriate tools for measuring it, and if different goals thus generate different accounts of well-being, then well-being may depend, in some crucial way, on the attitudes of the agent who is measuring well-being.

There is, then, reason to think that all objective list theories come up against similar problems to those faced by Nussbaum and Sen, once a detailed specification and measurement of prudential goods is attempted. Even if we assume that the goods can be identified in an attitude independent manner at a high level, the identification and

measurement of the goods in particular people, in particular contexts, leads to context-dependent accounts of well-being which are ultimately attitude dependent in certain ways. This discussion presents a problem for the objective list theorist. If he stops at the identification of high-level objective goods, we are left with a partial theory of well-being which can't tell us whether someone is doing well, nor how people are doing relative to one another. The theory will be impoverished to the extent that he refrains from specifying the goods. However, if he attempts to provide the level of detail, as Sen, and to some extent Nussbaum, does, he will end up leaning on the attitudes of agents in one way or another. Adequate objectivist theories of well-being cannot, then, avoid being dependent on the attitudes of some agents. Identifying and measuring whether people have prudential goods in their lives, and determining how to construct the information bases of prudential goods, is dependent on the attitudes and goals of those people making the measurement. There is no neutral or context-dependent way of performing these exercises. This suggests that measuring objective prudential goods is not a matter of measuring states of persons, or states of the world directly, but measures some relation between the object of measurement and the person or people attempting to measure well-being.

Of course, the objectivist may bite the bullet here, and claim that attitude independence is only required at the highest level, and the theory thus remains objectivist in an important sense. My arguments in 3.3.i gestured towards some difficulties for those attempting to secure attitude independence of the goods at the highest level. I pick up these considerations in chapter 6, and develop my critique of stance independence. However, even to accept the attitude dependence of the goods in their specification would constitute a fairly major concession on the part of the objective list theorist. If the manifestation of objective goods is necessarily attitude dependent, then there is no attitude-independent fact of the matter as to whether a particular person has the good in their life and so no attitude independent fact as to

whether they are doing well or not. The sense in which well-being remains objective is decidedly limited.



## CHAPTER 4

# Preferences

Those theories that define well-being in terms of the satisfaction of desires and preferences are seen as the main competitors to objectivist theories of well-being (Fletcher 2013). Desire- and preference-satisfaction theories of well-being take what is good for someone to be a subjective matter—things are good for us insofar as we desire them or prefer them to other options, and our lives go well to the extent that our desires or preferences are fulfilled. These theories satisfy the uniqueness criterion: they specify that, across all people and contexts, well-being is enhanced by the fulfilment of one’s desires or preferences. However, the actual constituents of well-being in different contexts will be very varied, because different people desire different things. In a sense, then, well-being is context dependent for the desire- or preference-satisfaction theorist, although it remains invariantist under a high-level description. My discussion in this chapter, however, will identify a different sense in which such theories are context dependent. I will argue that people’s preferences are contingent on the context in which they are elicited and the description of the outcome over which they range. Preferences shouldn’t be taken to be stable properties of people, and don’t range over outcomes *qua* states of the world. Instead, preferences range over descriptions, and reflect a relation between a person, the survey tool used to elicit stated preferences, and the environment in which measurement takes place.

The chapter proceeds as follows: in 4.1, I briefly outline the relationship between desire- and preference-satisfaction theories of well-being, and explain why my

discussion will focus on preference-satisfaction theories, preferred by economists, rather than more traditionally philosophical desire-satisfaction theories. In 4.2, I outline axiomatic utility theory, a preference-based theory of well-being developed and used by economists. In 4.3 I give an extended example from the healthcare context to illustrate how cardinal utility values are calculated for health outcomes. In 4.4, I introduce the concepts of systematic bias and framing effects, and begin to explore the problems that these present for preference-based theories. In 4.5, I suggest one way that preference-based theories could guard against these concerns, by excluding irrational preferences from rankings. I argue that this fails to appreciate the problem that the existence of framing effects poses. Framing effects indicate that preferences range over descriptions of states, and not over states themselves. This points towards the failure of the axioms that form the basis of utility theory, and a reconceptualisation of preferences. I suggest that, insofar as well-being is understood in terms of preferences, these are not best taken to capture states of persons, but instead to represent a relation between an individual and their environment.

#### *4.1 Desire- and preference-satisfaction theories of well-being*

Desire- and preference-satisfaction theories of well-being take people to be doing well to the extent that their desires or preferences are satisfied, that is, the extent to which the things that they desire are fulfilled or they get the outcomes that they prefer. Such theories are subjectivist: they not only take well-being to depend on the internal states and attitudes of individuals, but enable individuals to determine their own standard of well-being. For a desire theorist, in contrast to the objective list theorist, knowledge, pleasure, achievement, virtue and other prudential goods are only good for me if I desire them. It is the fact that something is desired, rather than the nature of what is desired, that is important for determining whether someone is doing well or not (MacLeod 2015, p.1079). While desiring something does seem to entail having some kind of positive affect towards it—what Peter Railton (2012) calls ‘liking it’—

fulfilment of desires and preferences need not entail positive mental states such as pleasure or enjoyment. Consider the stylite who desires, and achieves, an ascetic life that is devoid of pleasure—we need not think that his achievement of a pleasure-free life should necessarily lead him to feel pleasure, thus undermining his desire for it. Railton also takes desiring something to involve having an attitude of ‘wanting’ towards some object—that is having some motivation to achieve it. On the preference satisfaction that I consider in this chapter, preferring something is not identified with any particular mental state or attitude prefer something, but in terms of choice behaviour. To prefer something is to choose it over other options, or to be liable to choose it.

Desire- and preference-based theories need not set any external constraints on the kinds of attitudes that determine well-being, although some do. Some desire-based theories of well-being are largely unrestrictive in the kinds of desires that can contribute to well-being (Heathwood 2005, 2006). Others stipulate that only those desires which pass some process of critical reflection can be said to contribute to our well-being (Brandt 1979; Railton 1986; Bruckner 2016). Still other accounts take a person’s desires and to contribute to her well-being only to the extent that they conform with her deeply held values (Tiberius 2011). Similarly, some preference-based theories of well-being take unrestricted actual preferences to determine what would make life go well for us (Sugden 2006). Others require preferences to be considered, coherent and stable under experience and reflection if they are to be determinative of our well-being (Gauthier 1986). For others, if preferences are to be relevant to well-being, they must conform to an externally defined standard of rationality (Savage 1954). So despite an appearance of subjectivism, some desire- and preference-based theories of well-being end up leaning on externally defined standards to appraise the contribution of subjective attitudes to well-being.

On the whole, philosophers who defend this kind of theory of well-being tend to talk about desires, while economists tend to talk about preferences. The main benefit of the move from desires to preferences is that it circumvents the tricky matter of identifying, categorising and measuring subjective psychological states. Preferences can be defined behaviourally—in terms of the things that people choose or would choose—rather than psychologically—the things they have a ‘wanting’ attitude towards. It is misleading, perhaps, to speak of a move from desires to preferences. Preference-based theories in economics predate and provide the theoretical grounding for contemporary desire-based theories of well-being in philosophy, rather than the other way around (Heathwood 2015, pp.136–7). Economists made the move to identifying well-being with preferences, as opposed to subjective psychological states, in the first half of the 20th century, in part due to the difficulties faced by utilitarianism in identifying, characterising and rendering commensurable pleasure and pain. Despite recognising the need for a measurable concept of well-being, and seeing value in the utilitarian’s identification of well-being with states of consciousness, there was growing scepticism as to the possibility of reliably measuring the pleasure and pain associated with particular outcomes (Marshall 1920, Bk.III Ch.III Fn.57; Pigou 1932, Pt.1 §5). Lacking the means to measure subjective psychological states, early welfare economists took choice behaviour to be a reliable indicator of well-being. And later, preferences ceased to be seen as an indicator of underlying subjective states, and instead came to be viewed as constitutive of well-being (Harsanyi 1977, p.645; 1998, p.285).

In some ways contemporary desire-fulfilment theories mark a regression from this, relying, as they do, on the identification and measurement of particular subjective attitudes to determine how well someone’s life is going. This makes identifying and measuring contributors to well-being far more difficult—there is, ostensibly, something more definitive and identifiable about people’s choice behaviour,

compared to their desire states.<sup>38</sup> Furthermore, in the mid-20th century, developments in economic theory produced and refined tools for determining the relative strength of preferences and assigning numerical values to them. Determining and comparing the relative strength or weight of desires remains a particular problem for desire theories—desires don't seem to be much easier to identify and measure than pleasures and pains, if at all. However, the desire-theorist's reluctance to divorce assessments of well-being from mental states also demonstrates, I think, calculated resistance against the economists' behavioural picture. There is something explanatorily inadequate about the view that equates well-being with that which is chosen out of all of the possible options—on theories of well-being which equate preference with rational choice, agents cannot help but maximise their well-being so long as they choose rationally. Moreover, someone may choose outcomes that are not in their interests, and towards which they have no positive attitude towards at all. This might well be taken to fail to capture the relationship between well-being and positive attitudes—something which seems to be a primary motivation for preferring a subjectivist theory of well-being in the first place. Of course people may also *desire* things that are not in their interests and that are defective in a number of ways.<sup>39</sup> Nonetheless, desiring does entail some form of positive attitude towards the objects of desire, which minimally connects someone's desires with their interests.

My discussion in this chapter concerns preference-satisfaction theories of well-being, and in particular I focus on utility theory, which was developed by economists as a means of measuring well-being via descriptions and predictions of people's choice behaviour. I take utility theory as my subject, rather than any of the philosophical

---

<sup>38</sup> I should note here that the arguments presented in this chapter will call this suggestion into question, though via the problematisation of choice behaviour, rather than via the elucidation of desire.

<sup>39</sup> See Heathwood 2005 for discussion of the variety of ways in which desires may be deemed defective.

desire-satisfaction theories of well-being, for two main reasons. First, utility theory directly addresses the measurement and identification of well-being via the measurement and identification of actual and hypothetical preferences. Desire-satisfaction theories, on the other hand, tend to avoid giving an operational account of well-being. Many desire-satisfaction theories focus on problematic or defective desires—base desires, desires for worthless things, desires based on false beliefs or lack of information, irrational desires, and so on, but have very little to say about how to determine whether someone has a desire for something or not, nor about how to measure or compare desires (Bruckner 2016; Heathwood 2005, 2006; Murphy 1999). Railton (2012) spends time thinking about the exact mechanism of desire and its manifestation in our subjective attitudes, breaking it down into a ‘liking’ and ‘wanting’ attitude and a feedback mechanism for modulating affects based on the comparison between expectations and actual experience. However, he too says very little about the identification of desire (or liking, wanting) in people, and so, despite giving a more detailed high-level account of desire, doesn’t operationalise his account. Utility theory, on the other hand, has the measurement and quantification of well-being at its centre. It is therefore a good place to start in examining the mechanism of a subjective theory of well-being.

My second reason for focussing on preference-satisfaction theories in this chapter is that much of my discussion of the direct measurement of subjective psychological states in the next chapter will also apply to desire satisfaction theories. My subject in chapter 5 will chiefly be hedonic theories of well-being, which I take to be examples of objectivist theories. Nonetheless, I take much of my treatment of hedonism to apply to any attempts to measure subjective psychological states directly, and thus it will apply to desire theories, insofar as they take particular mental states to be determinative of well-being.

## 4.2 *Utility theory*

Since the mid-20th century, economic orthodoxy has maintained that an approximation of individual well-being can be captured by eliciting preferences, that is, by asking or observing what individuals would choose when faced with a finite number of options (Sen 1973, p.241; Kahneman and Tversky 1979, p.263; Dolan and Kahneman 2008, p.215). Contemporary utility theory stems from John von Neumann and Oskar Morgenstern's influential *Theory of Games and Economic Behaviour* (1944), in which the authors develop a mathematical model which describes the 'endeavor of the individual to obtain a maximum of utility.' Individuals' preferences for particular outcomes or prospects are taken to reveal the utility attached to those prospects, given the assumption that people choose those options that will maximise their own utility. Utility theory is understood, by von Neumann and Morgenstern and their contemporary Jacob Marschak, at least, as an explanatory or predictive model of how individuals actually make choices (Marschak 1950, p.112; von Neumann and Morgenstern 1944, pp.1–5). So on this account, understanding the choices that individuals make as utility-maximising is the best way of explaining why those individuals chose the particular options they did, and accurately predicting the choices they will make.

Along with other preference-based theories of well-being, utility theory does not try to measure internal states—motives, interests, reasons—to determine which outcomes people value, but instead aims to assess the value of outcomes by creating mathematical representations of the choices people make. Utility theory takes subjective value to be determined by ranking outcomes in virtue of which an individual is more likely to choose, or which they would choose first. *Expected utility theory* extends utility theory to deal with uncertain outcomes—it takes subjective value to be determined by the *expected* valuation of uncertain outcomes. So, on such theories, an outcome has higher utility to the extent that it is, or is expected to be,

placed higher in an overall preference ranking. Where the ranking of preferences only includes information about the fact that one outcome is preferred to another, and not the degree to which it is preferred, this is known as *ordinal* ranking—only the *order* of outcomes is preserved. Expected utility theory enables the strength of preferences to be determined in addition to the mere fact of their preference, because it involves eliciting preferences over certain outcomes and lotteries among alternative outcomes. The terms on which an individual is indifferent between the certain state and the lottery determine the strength of their preference over the outcomes. This enables the degree to which one outcome was preferred over another to be settled, and precise utility values to be assigned to preferences, to designate how much they are valued by individuals. This is known as *cardinal* ranking.

Central to utility theory is the claim that preference behaviour that maximises expected utility is *rational* and conforms to a number of axioms of rational choice (Bermúdez 2009, p.2). These axioms include *transitivity*, *completeness*, and the *independence of irrelevant alternatives* (von Neumann and Morgenstern 1944, p.26). Transitivity states that for any three outcomes  $u$ ,  $v$ , and  $w$  if  $u$  is preferred to  $v$ , and  $v$  is preferred to  $w$ , then  $u$  must also be preferred to  $w$ . So it cannot be the case that I prefer pecan pie to cherry pie, and cherry pie to apple pie, but prefer apple pie to pecan pie. Completeness states that for any two stated outcomes, each individual must either prefer  $u$  to  $v$ , prefer  $v$  to  $u$ , or be indifferent between  $u$  and  $v$ —that is, decision-makers cannot fail to have preferences over stated prospects. When presented with a choice between cherry and pecan pie, I cannot find the two to be incomparable. Insofar as I am unable to choose between the two, it must be the case that I am indifferent between them. Finally, when there is uncertainty over outcomes—that is, where participants choose over probabilistic, rather than certain, outcomes—preferences must treat irrelevant alternatives as independent. This means that, if I prefer  $u$  to  $v$  out of the set  $\{u, v, w\}$ , neither removing  $w$  from the set nor adding  $x$  to it should change my ranking



of  $u$  and  $v$ . If I prefer cherry pie to apple pie out of a choice between the two, it cannot be the case that adding the option of pecan pie causes me to rank cherry below apple, though of course pecan may appear above, below or in-between them in the ranking.<sup>40</sup>

As utility theory is an explanatory model, these constraints are intended as descriptions of the relations between preferences that agents purport to have, or those that they reveal through their choices, as long as they are not misinformed, or missing information, when they make their choice. This final point is key—utility theory requires that individuals have awareness of all of the relevant options, and full understanding of what those options involve, in order for their choices to be truly indicative of their preferences and actually utility maximising (von Neumann and Morgenstern 1944, p.30). Given the constraints of transitivity and completeness on the ordering of preferences, it should be clear that preference is understood, on this model, as an all-things-considered-better-than relation, where ‘better than’ refers only to a higher position in the overall preference ranking.

Well-being *qua* utility does not, then, refer to a subjective psychological or experiential state; the claim is not that individuals endeavour to maximise their own happiness or pleasure, although plausibly this may figure as a desirable outcome in preference deliberations. Rather, a prospect’s utility or expected utility for an agent is just a function of how much they prefer it to alternative options (Hampton 1994, p.205). Utility is taken to represent the value of a prospect for an agent. This is a far cry from Jeremy Bentham’s definition of utility:

---

<sup>40</sup> The list of axioms also includes *continuity*, which doesn’t figure in my discussion in this chapter. Continuity is the principle that, if I prefer  $u$  to  $v$ , and  $v$  to  $w$ , there must be some probability  $p$  such that I would be indifferent between having  $v$  occur for certain, and having  $u$  occur with probability  $p$  or having  $w$  occur with probability  $1-p$ .

that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered.  
(1843, pp.118–9)

Utility moves from being used to refer to a tendency of an object to produce good to referring to the good or value that an object produces (Broome 1991, pp.1–2). For the utility theorist, ‘obtaining a maximum of utility’ essentially means choosing options which produce maximum value. But producing maximum value simply amounts to choosing the options which are preferred, regardless of the reasons for which they are preferred. It does not entail that utility maximisers choose options which are in their interests, where these are defined externally to their choice behaviour, or which will make them happy. This lack of restriction on the content of preferences leads to utility theory sometimes being taken to be ‘psychologically neutral’ about the motivations for individuals making particular choices, and preferring one option over another. Any reason for preferring a prospect will do, as long as revealed preferences conform to the axioms of rational choice. Utility theory keeps quiet, to a large extent, about the reasons for which an individual prefers those things that they actually do prefer. It eschews hedonistic accounts of value, and acknowledges the broad range of reasons for which people make decisions, including non-selfish reasons, and reasons unrelated to their psychological states. One implication of defining utility in behavioural terms is that, so long as they are acting in accordance with the axioms of rational choice, individuals can’t help but maximise their utility. Those outcomes that they choose necessarily maximise utility, insofar as they are rational.

### *4.3 Measuring health state utility*

An example from the healthcare context illustrates how preference elicitation techniques generate utility values, which are used to construct a social choice function for application in public policy decision making.

The elicitation of individual preferences about health states is widely used as a method of establishing how ‘health states’ contribute to health-related quality of life (Hausman 2015, chapter 8). The term health state is often employed to refer to either a condition or disease (deafness, paraplegia, diabetes, for example), or a description of the impact of that condition on physical and psychological functioning, such as its effect on mobility, mental state, pain or discomfort or health more generally. Health states are often given as functional descriptions, which describe the physical, mental and social state a person is in in fairly generic terms, normally in terms of what a person in that state is able to do. This means that people with disabilities or diseases which are quite distinct clinically can come under the same functional description, and that resulting health state utilities can be used to describe and compare the quality of life of people with different conditions.

The first step in measuring health state utility is to determine the complete set of health states over which utility values will range. So, for example, one of the most widely used tools for measuring quality of life, is the EQ-5D-5L (Herdman et al. 2011). It measures five domains: mobility, self-care, usual activities (e.g. work, study, leisure), pain/discomfort and anxiety/depression. Each of these is measured on five levels, from no problems (1) to extreme problems (5) in the domain. There are 3,125 possible health states given by the EQ-5D-5L, covering all of the permutations of levels across the five domains (Mulhern et al. 2018, p.700). So a possible health state might be 22131—slight problems walking, slight problems with self care, no problems performing usual activities, moderate pain, and no depression; or 11111—no problems in any domain. There are thousands of tools for measuring health outcome,

each of which use different sets of health states, with different functional state descriptions, most much more detailed than EQ-5D-5L.<sup>41</sup>

Once all of the different possible health states have been identified, the value of these health states is determined by eliciting preferences for them. The set of all health state utility values is called a *value set*. As preference-based measures of quality of life are used in cost utility analysis, and used to make decisions about public resource distribution, values are ascribed by members of the public, not necessarily people who have been, or will be, or know people who have been in any of the health states in question. They are, then, *social* values, and thus reflect the health state preferences of society, rather than of specific groups of individuals. As there is no actual market of health states, an individual's stated preferences about hypothetical choices between health states are taken to determine the utility of those states. Very rarely actual comparisons involving health states they have been in, or are currently in, are used. Individuals' revealed preferences between health states are used to assign 'health state utilities' to them, that is, to determine the level of utility each health state has for them. These utilities are used to create a ranking of health states, typically using 'full health' and 'death' as limits or reference points, which allows comparison between health states on the basis of their utility (Bowling 2005). Often health state utilities are assigned numerical values to denote their place in the overall ranking, so, for example, if full health is taken to be an upper limit of 1.0, a health state such as diabetes might be assigned a utility of 0.87, compared with asthma, which might be given a utility of 0.97, and emphysema a utility of 0.52. Often utilities are subsequently aggregated across individuals to generate an account of social preferences about health states

---

<sup>41</sup> The MOS 36-Item Short Form Health Survey (SF-6D), for example, corresponds to 18,000 health states (Brazier et al 2002); the Health Utilities Index (HUI) 3 corresponds to 972,000 health states (Hausman 2015, p.37).

(Hausman 2015, Ch. 8.). Because it is non-specific about the reasons underpinning preferences, health state utility, on this account, tells us something about the relative value of the options being evaluated, but is not supposed to make any claims about the reasons for which they are valued.

Health state utility is usually measured in highly structured interview settings, where participants are asked a series of questions, often using props and accompanying text, relating to a variety of health states (Dolan and Kahneman 2008). One commonly used approach is the *time trade-off* (TTO) model, a cluster of survey methods where an individual is asked to choose between two alternatives in the following manner:

*Alternative 1* is a treatment with a certain outcome of a chronic state for 10 years, followed by death.

*Alternative 2* is a treatment with a certain outcome of full health for  $x$  number of years (where  $x \leq 10$ ) followed by death. Time  $x$  is varied until the respondent is indifferent between the two alternatives. The point of indifference is used to produce a health state utility value  $u$  between 0 and 1.0, where  $u = x/10$ . So if Anne is indifferent between loss of a leg for 10 years and full health for 8 years, then the health state utility she attaches to loss of one leg would be 0.8, where 0 represents death and 1.0 represents full health.<sup>42</sup>

Another approach to the measurement of individual preferences between health states is the *standard gamble* (SG) model, which again asks participants to choose between two alternatives:

*Alternative 1* is the certainty of a given chronic health state for ten years, followed by death.

---

<sup>42</sup> Adapted from Gudex 1994a.

*Alternative 2* is a treatment with two possible outcomes: cure (i.e. perfect health) for ten years or immediate death. The relative probabilities attached to perfect health and death are varied (e.g. 10%/90%, 20%/80%, 30%/70%), until they arrive at a point where they are indifferent between the health state and treatment. The probability level attached to the cure at this point is taken to be the utility attached to the health state. So, for example, if Anne is indifferent between blindness in one eye and a treatment with a 90% chance of cure and a 10% chance of death, the health state utility she attaches to blindness-in-one-eye is 0.9, where 0 represents death and 1.0 represent full health.<sup>43</sup>

Note that in neither of these approaches is utility necessarily defined in terms of subjective psychological states, rather it is defined entirely in terms of preference, which is understood in behavioural terms. That health state  $x$  has a higher level of utility than health state  $y$  for Anne does not imply that she anticipates that she would be better in  $x$  than in  $y$ , or would be happier in  $x$  than in  $y$ , merely that she prefers  $x$  over  $y$ , or would (hypothetically, or actually) choose  $x$  over  $y$ . The reasons for which she prefers  $x$  could be nothing to do with her own subjective states, or anything to do with her health *per se*—for example, it could be because she considers choosing  $x$  to be her duty, or because choosing  $x$  would give certain opportunities to her children, or because choosing  $x$  would save money for the National Health Service. Although the states in question are health related insofar as they are health states—that is, they have been pre-specified as states of health by those designing the survey—there is no specification that they be evaluated with reference to their impact on the health of the agent in question.

When using the time trade off approach to develop a measure of quality of life, such as EQ-5D-5L, participants will answer questions about a series of different health

---

<sup>43</sup> Adapted from Gudex 1994b.

states, and this process will be repeated with different participants. The final utility value for a given health state will be determined by aggregating scores across participants. These preference elicitation scenarios generate a value set covering all of the health states, designating social preferences about each state. It would be too time-consuming and expensive to measure the utility of every health state in the EQ-5D-5L using the preference-elicitation techniques detailed in this section, so economists measure the utility value of a subset of health states—86 out of 3,125, in the case of the EQ-5D-5L (Hausman 2015, pp.85–6; Mulhern et al. 2018, p.701). They then use these values to impute the values of the remaining health states. For example, if one health state dominates another—if it is higher on no dimension and lower on at least one—it will be assumed that its utility value is lower. The degree to which it is lower will be determined by a multi-attribute function based on the utility values that have been directly measured (Hausman 2015, p.86).

In order to measure the quality of life associated with a particular disease or in a specific population group, a group of people matching that description will fill out the survey. This will identify each as being in one of the possible health states. Their utility values in those states are determined by looking up the social utility value for that state. The utility associated with a disease or condition, or for a particular demographic group, is determined by averaging out the individual utility values of all of the surveyed people in that group. Similarly, in order to measure the quality of life of an individual using the value set, they will first determine their health state by filling in the survey that corresponds to the value set in question. Their personal utility level is determined by looking up the social utility value for that state.

#### *4.4 Systematic bias and framing effects*

Two central problems have been identified with utility theory: first that individuals do not always act so as to maximise their utility, as defined by the axioms of rational choice. Instead, people systematically use decision-making heuristics, which do not

lead to maximisation of utility or expected utility. This suggests a disjunct between choice and utility. Secondly, the way in which outcomes are represented or *framed* affects the way people make choices. This suggests that our choices don't track our preferences, and that instead they are highly dependent on seemingly irrelevant features of the measurement context, such as the particular words used to describe outcomes, or the ordering of questions.

Kahneman and Amos Tversky (1979) identify a number of systematic decision-making biases which seem to be in tension with the utility-maximising supposition of utility theory. One observed inclination is that *certain* outcomes are over-weighted relative to *probable* outcomes; people systematically choose, for example, a certain gain of £450 over 50% chance of a £1000 gain and 50% chance of no gain at all. A second, related, heuristic is that people's choices systematically exhibit *risk aversion* in the positive domain, that is, a preference for certain gains over probable ones, and *risk seeking* in the negative domain, that is, a preference for probable losses over certain ones. So, for example, out of a sample of 95 people, 80% preferred a certain gain of 3000 over an 80% chance of a gain of 4000, whereas 92% of the same people preferred a 80% chance of losing 4000 over a certain loss of 3000 (Kahneman and Tversky 1979, p268).<sup>44</sup> In these cases, decision-makers systematically tend to choose lower expected utility (or higher expected disutility) in favour of some other good, such as 'certain gain,' or 'chance of no loss.'

Note that in these examples, 'utility' is defined as some specific outcome, such as financial gain, or increased life expectancy. Given this, it seems easy enough for the utility theorist to sidestep this objection. In cases such as these, agents can still be understood to be maximising their utility, even though they are not maximising the

---

<sup>44</sup> The utility in question was Israeli currency, in this case.



financial value of the outcomes in question. Instead there may be other things that they take to be valuable and which inform their preferences, such as the value of certain gains over possible ones, even when the expected outcome is probabilistically speaking higher in the latter case. Given this, when people choose a lower certain gain instead of a higher probable one in these examples, they can in fact be said to maximise their utility, because a full account of utility includes an account of the benefit attached to certain gain as well as the utility of expected gain. In this sense, utility theory can define 'utility' not in terms of pre-defined outcomes (such as income, welfare, or other particular goods) but as whatever it is that people happen to prefer, whatever it is that happens to make prospects all-things-considered-better for them. The only thing that matters is that the prospect is *all-things-considered* better, rather than better with respect to one feature or another in some regard or other, so that transitivity is preserved across all comparisons.

On this account, people can't help but maximise their own utility, unless their decision making doesn't conform to the axioms of rational choice theory (Sen 1977, pp.322-3). One concern with the equation of utility with preference is that it leaves us with a distinctly impotent conception of utility, which denies that people can make rational choices that diminish their utility. Instead, insofar as they are rational, choices maximise utility, even if they appear to make someone worse off in important ways; insofar as choices fail to maximise utility, they are irrational or misinformed. Choices are understood to maximise utility even if they result in someone feeling much worse or being much more unhappy, or if their choices fail to satisfy their desires or to serve their interests; regardless of the consequences of rational choices on subjective psychological states, by definition they maximise utility. Of course an agent can miscalculate the utility or expected utility attached to a particular prospect, or fail to understand what the alternative options are, in which case their decision no longer

meets the constraint of full information. However, insofar as an agent fulfils this requirement, their choices must be understood as maximising utility for them.

Clearly this deviates from common usage of ‘utility’ to mean something like happiness. Measures of decision utility tell us not what is of benefit to people, what is in their interests, what is enjoyable to them, but simply what they choose, which can fail to fulfil any of these descriptions. Equating utility with what is in fact chosen also makes it very difficult to say anything about the reasons for which the chosen outcomes are valued, as every choice is understood as being ‘all-things-considered’ the best outcome. However, this is not necessarily a problem for the utility theorist. Indeed, it might be thought to be a benefit of preference elicitation that understanding what is valuable for particular individuals doesn’t require knowledge of their subjective psychological states or their interests.

A second widely observed feature of decision making using preference elicitation is that stated preferences exhibit *framing effects*: the way in which outcomes are represented or framed affects the way people make choices (Arrow 1982; Kahneman 2003; Kahneman and Thaler 2006; Tversky and Kahneman 1981, 1986). Even if certain decision-making heuristics can be incorporated by an updated form of utility theory, there are some systematic biases observed in decision making which result in inconsistencies, and which seem to be in contradiction with the central claims of utility theory. Some examples will help to describe this phenomenon. In one study physicians and graduate medical students were asked to imagine they had lung cancer, and to choose between two different therapies, surgery and radiotherapy (Arrow 1982, p.7). Surgery presented a higher risk of death during the intervention, but a better life expectancy thereafter. Of those presented with the probabilities of *survival* for each therapy 84% preferred the surgery, and only 16% the radiotherapy. Of those presented with the probabilities of *death* attached to each therapy, only 50% chose the surgery. Both choice scenarios are clearly extensionally the same—the probability of survival is

equivalent to 1 minus the probability of death—but the respondents had a clear preference for the option where the higher upfront risk was expressed in terms of survival over that expressed in terms of death. Paul Anand (1987 p.196) suggests that the reason for these contrary preferences may be because the phrasing of the question in terms of ‘survival’ and ‘death’ makes these options respectively more salient, and so places these outcomes in the forefront of the mind of the decision-maker, or because clinicians tend to use positive language when recommending a treatment, and negative language when counselling against it. Whatever the precise explanation, this study clearly illustrates that the way a question is framed can radically alter the response, even when, on reflection, the outcomes are the same.

Another experiment with similar conclusions asked college students “How happy are you with your life in general?” and then “How many dates did you have last month?”, and found a small negative correlation between the answers to the two questions (Hausman 2015, p.110). However when the dating question was asked first, there was a positive correlation of 0.66 between answers (Kahneman 2003, p.710). Kahneman conjectures that the dating question provided a context within which the college students framed their answers to the happiness question. In the healthcare context, survey participants rate impairments such as paraplegia as worse if particular negative effects of the disease—sores, impotence, incontinence, limited access to many private and public buildings and spaces, restricted work opportunities, discrimination—are highlighted, and as less bad when these are not mentioned, and the condition is described in less unfavourable terms or just in less detail (Hausman 2015, p.56).

The way in which a question is framed—the language used, the order of questions, the information included and left out—leads survey participants to focus on particular things, and so influences their answer. This means that the way a choice is presented can impact on the preferences expressed by the respondent, resulting in different preferences being expressed over outcomes that are, on reflection, the same in an

important sense. This suggests that the preference rankings that result from these choices don't reliably reflect the considered preferences of the respondent, and thus don't necessarily maximise utility. Instead the resulting preferences are highly dependent on features of the measurement context which are usually taken to be irrelevant to the content of the outcomes, and their associated utility. This clearly presents a significant problem for the utility theorist. It appears to show that the ranking of outcomes reflects something other than their utility, something apparently irrelevant to the measurement of utility: the way in which the outcome is described. This means that the numerical utility value attached to each outcome may not represent or approximate preferences over the outcome, where this is understood to be separate from any given description of it, at all.

The utility theorist might argue that, in contexts where elicited preferences are determined or in danger of being determined, by framing effects, participants do not count as fully informed about outcomes, and so their preferences are not rational. If the participants fully understood what the outcomes entailed, their preferences would be consistent. This response assumes that a fuller description of the outcome will allow the respondent to understand the implications of the outcome in a way that they were not able to do when presented with less information. This implies that framing effects merely distort preferences, or that features of measurement scenarios get in the way of full comprehension of outcomes rather than precluding full comprehension altogether. We might avoid framing effects, for example, by expressing the outcome in terms of both losses and gains; or by giving examples of what an outcome would mean for the respondent, for example, what severe problems with mobility would prevent them from doing or what moderate pain feels like; or by providing alternative interpretations of outcomes in order to give a more rounded or multifaceted impression of the outcome to respondents.

But this misses the force of the problem that framing presents for utility theory. The features which frame, and impact on, the eventual preference ranking—such as the particular words used to describe outcomes, the context within which the choice is set, the order of questions, the amount and type of information given—are *inevitable* features of any choice scenario. There is no neutral way of phrasing questions. Questions must be asked in *some* order. Giving respondents more information, or presenting various different interpretations of an outcome does not neutralise the presentation of the information—the outcome is, regardless, being presented in *some* manner. Kahneman gives an example of a study, where participants played the role of a judge adjudicating in a custody case between divorcing parents, which makes this point explicitly:

Each parent was described by a list of attributes. One of the descriptions was richer than the other: It contained more negative and more positive attributes...Some respondents were asked which custody request should be accepted; others decided which request should be rejected. *The rich description was selected under both instructions* (Kahneman 2003, pp.702–3).<sup>45</sup>

More information does not, then, straightforwardly equal better ability to choose correctly. Indeed, the concept of ‘sufficient’ or ‘full’ information is a somewhat elusive one. It seems as though the way in which the information is framed affects how the options are understood, and ultimately affects decisions about those options—regardless of how ‘full’ the information is, or how much of it there is.

#### 4.5 *Eliminating irrational preferences*

In light of such framing effects, the utility theorist cannot simply argue that people are *in fact* accurately representing the expected value of the outcome when, for example,

---

<sup>45</sup> Emphasis added.

they strongly prefer surgery when the outcome is described in terms of survival, and also when they equally prefer surgery and radiotherapy when the outcome is described in terms of death. Unlike in the examples discussed in the previous section exhibiting risk aversion in relation to losses, and risk seeking in relations gains, the outcomes in each case are the same. They can't be understood as in fact representing different outcomes. The reason for this is that utility theory is committed to *description invariance*, that is, that different representations of the same outcome should lead to the same preference (Tversky and Kahneman 1986, p.S253–4). Tversky and Kahneman take this to be an unstated assumption of utility theory, rather than an explicit and testable axiom; it is implicit in the way that options are characterised.

How, then, can the utility theorist respond to such framing effects? In order to assess and choose between the different descriptions and different preferences, we need some standard against which to judge the representativeness or appropriateness of descriptions. One solution would be to judge a description to be appropriate if it results in quality weightings which accurately represent the utility of a given outcome. But this obviously begs the question—preference elicitation was supposed to be the method by which we determine the utility of outcomes, and so determine how to numerically represent the value of those states. If we need an external standard of utility to judge the outcomes of this process, the whole point of the exercise is undermined. Given this, utility theory can attempt to protect itself non-circularly against framing effects is by appealing to an *internal* standard to assess preferences: the axioms of rational choice. Only *rational* preferences maximise utility, for the utility theorist, so if preferences and resulting utility values fail to conform to the axioms they can be judged to be irrational, and so excluded from relevant measures of utility. I have shown that the requirement of full information is not achievable but, in spite of this, filtering out or rectifying preferences which don't meet the standards of transitivity,

completeness and independence may rule out many of the contradictory preferences generated by framing effects.

Some framing effects produce preferences which violate the axiom of independence of irrelevant alternatives. In one study, participants were asked to predict how much they would enjoy eating potato crisps in a few minutes time (Kahneman and Thaler 2006, p.226). Some participants could see a chocolate bar near to the crisps, whereas other participants could see a tin of sardines; these items were not, however, mentioned. The irrelevant foods influenced the participants' predictions of their enjoyment of crisps: their preference for crisps was increased by the presence of sardines, and decreased by the presence of chocolate. The utility theorist would characterise these preferences as irrational, and so could justifiably exclude them from any social choice or aggregation function.

Other framing effects violate the principle of transitivity. A study asked participants to choose between two public policy initiatives in response to a disease which was expected to kill 600 people (Kahneman 2003, p.702). The participants were split into two groups. The first group was presented with two policy outcomes as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

The second group was presented with the same two policy outcomes, differently worded:

If Program A\* is adopted, 400 people will die.

If Program B\* is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

In the first group, a clear majority favoured program A, the risk avoidance option. In the second group, a clear majority favoured program B\*, the risk seeking option. When participants were asked to directly compare the two options, however, they almost always agreed that the same policy should be adopted in both cases. This outcome appears to violate the principle of transitivity because it suggests that the same outcome under different descriptions can appear at different places on an overall utility ranking. This leaves open the possibility that an outcome be ranked both high and lower than other outcomes. For example someone might prefer Program A, that 200 people be saved, to Program B, that there is a one-third probability that 600 people will be saved, yet prefer Program B to Program A\*, that 400 people die; so if  $A > B > A^*$  and  $A = A^*$ , then  $A > B > A$ .

So, in both of these examples, the resulting preferences ought not to be counted when aggregating individual utilities, insofar as they are irrational. But in neither scenario is there a straightforward way of choosing between the outcomes under different descriptions, or in different contexts: neither the preferences elicited by crisps-near-chocolate and by crisps-near-sardines can be said to represent an outcome unaffected by irrelevant alternatives. And neither the preference for policy A over B, nor the preference for policy B\* over A\* represent preferences over a neutrally described outcome. When faced with conflicts of preferences, it's not clear how to choose between the two, so the utility theorist appears to have to drop *both*. The utility theorist may suggest that some other description or contextualisation of an outcome can be provided which *does* elicit representative preferences: predictions about levels of enjoyment from crisps-on-their-own, for example, or considering A and B and A\* and B\* all together. These new descriptions may be developed by considering the information that is required to prompt participants to recognise the irrationality of the preferences elicited in response to the outcomes under their original descriptions. So, for example, if consideration of both outcomes side by side is enough to prompt



recognition that they in fact represent the same outcome, this should form the basis of the new description.

There are two possible reasons for preferring the new description to the previous descriptions. The first is on the grounds that it constitutes a better, more accurate, or more neutral description of the outcome. However, given the discussion about the impossibility of full information in the previous section, this line of reasoning is not available to the utility theorist. These purportedly neutrally described outcomes may well result in *different* preferences than result from the framed descriptions, but if there is no way of defining a standard of full information, and if more information isn't necessarily better, then the claim of neutrality doesn't provide a good reason for choosing this description over the allegedly less neutral descriptions. Eliciting preferences under new descriptions generates a proliferation of utility values, rather than a benchmark for assessing utility values.

The second reason for preferring the new description is on the grounds that it does not result in irrational preferences. The new description is better in virtue of the preferences that it elicits. This line of reasoning implies that irrational preferences might be avoided by filtering out those descriptions that engender or are deemed likely to engender irrational preferences (Goodin 1986, p.76). This would involve identifying those descriptions which produce, or are likely to produce, irrational preferences—such as putting a question about dating immediately before a question about overall quality of life, as in the example discussed in the previous section—and avoiding asking these kinds of questions. The same approach might be taken in cases where particular 'hot' words prompt wildly different preferences in comparison to apparent synonyms, or certain features of the measurement context prompt divergent responses. Avoiding using hot words, contextual features or question orders avoids the possibility of eliciting inconsistent preferences further down the line.

However, this response fails to fully recognise the problem posed by the existence of framing effects for two reasons. Firstly, the reason that the utility theorist takes the divergent disease policy preferences and potato crisp preferences to be irrational is because they take the outcome over which the preferences range to be the same in all cases, but to be preferred differently under different descriptions. The irrationality is not in the preferences associated with the 'hot' words and orderings *per se*, but in the contradiction between these preferences and the preferences associated with the same outcome under other descriptions or in different contexts. Simply removing some descriptions from consideration does not make the preferences over remaining descriptions rational. It merely screens their irrationality from view. If we only accept preferences over crisps-on-their-own, or only over the two policy outcomes when they are presented in terms of death and in terms of survival, then we ignore the fact that these *very same outcomes* elicit different preferences under different descriptions. The preferences over the 'neutral' descriptions are also intransitive and fail to be independent insofar as the 'non-neutral' descriptions are descriptions of the very same outcomes. Secondly, the existence of framing effects comes to the surface when framing leads to preferences that violate the rational choice axioms. The cases of greatest interest occur when framing effects are so active that they systematically affect the ranking of preferences. However, framing effects are not limited to such occasions. Framing depends on factors such as the language used to describe outcomes, features of the measurement context, and perception of the kind of question being asked (Tversky and Kahneman 1986, p.S273). If these can affect how agents formulate and state their preferences, then they can do so in all contexts, not just those which lead to irrational preferences.

The existence of framing effects calls into question the assumption of description invariance. The commitment to description invariance implies that outcomes over which preferences range are independent of the way in which they are described in

preference elicitation scenarios. The description of the state in the preference elicitation scenario is a placeholder for some actual (or possible) state of affairs in the world, which exists separately from the description. Respondents are therefore taken to be choosing between, and placing a value on, states, rather than descriptions of states. If preferences are dependent on the way in which outcomes are framed and presented, then preferences must instead be understood as ranging over descriptions of outcomes, or perhaps interpretations or conceptualisations of outcomes, rather than over the outcomes or states of affairs *per se*. Moreover, framing is unavoidable. Even a description that highlights the ‘sameness’ of descriptions previously treated as different—such as the description which includes both sets of disease policies—is a description of the outcome. The utility theory approach to determining the quality weighting of outcomes via preference elicitation assumes that the outcome states over which respondents are required to express their preferences are clearly defined, conceptually pre-existing states, which individuals are then asked to put into a ranking. Utility theory thus presents preference elicitation as a matter of *discovering* one’s preferences through structured scenarios. However, the existence of framing effects indicates that preferences are *constructed* in measurement scenarios, via the presentation of outcomes under certain descriptions.

The existence of framing effects also calls into question the axiom of completeness, which is highly dependent on the assumption of description invariance. Utility theory requires agents to be able to make pairwise comparisons about all possible combinations of outcomes in their choice set, that is, they must have preferences about all of the possible choices they can make, and be able to create a complete ordering of their preferences (Anand 1987, p.192). However, completeness of pairwise comparisons is ruled out by the framing effects exhibited in the elicitation of preferences. The problem raised by the potato crisps experiment helps to highlight this. If people’s preferences over outcomes change when their attention is focussed on

other outcomes, or other features of the measurement context, then each preference cannot be slotted into a complete ranking of preferences. In different contexts, under different descriptions, preferences shift and change. If there is no way of expressing all of the options an agent has to choose between in a neutral or 'full' way, then there is no finite list of precisely defined choices or outcomes. Preferences may be imprecise, ranging over non-specific objects, and may be expressed and qualified in innumerable ways. This makes the utility theorist's requirement of completeness of preferences implausible.

At the beginning of this section I referred to the axioms of rational choice as an 'internal' standard against which to assess preferences. You may well have raised an eyebrow at this suggestion. There's good reason to think that the formal consistency constraints posed by the axioms of rational choice constitute an external, normative standard against which preferences are assessed. Sen (1973) points out that the utility theory conception of rational preferences is not in fact neutral, or a case of 'behaviour without reference to anything other than behaviour.' Rather, the psychological assumptions of rational choice theory are 'well chosen' so as to appear fairly neutral. Utility theory makes implicit assumptions about the relationship between preferences and psychological states—its reliance on notions of rationality, sufficient information and individual utility maximisation entails an account of preferences that is dependent on a particular model of human psychology. And, as Jean Hampton (1994, p.211) puts it, 'the axioms play a role in fixing the ends of action by acting as a normative "sieve" through which our preferences must pass in order to be considered "value-defining".' Though the utility theorist sets out with the aim of defining well-being in subjectivist terms—in terms of what individuals choose or prefer—the consistency conditions imposed by the axioms of rational choice in fact constrain subjective preferences. As the array of examples presented by Kahneman and Tversky show, these consistency conditions, if enforced, will end up significantly influencing the final rankings.

People's choices fail—systematically—to conform to the utility theorist's conception of rationality and, consequently, enforcing rational preference will involve systematic adjustment of rankings. If this is right, there's reason to think that utility theory fails as an explanatory and predictive model, and instead makes a normative claim about preferences—people's preferences *ought to* conform to the axioms of rational choice.

Economist Leonard Savage, when confronted with the Allais Paradox, chose options that are incompatible with the axioms of rational choice (A over B, and B' over A'). Rather than admit a shortcoming of the independence axiom, he concluded that his own preferences were in error (Savage 1954, p.103). This is a telling result. It indicates that establishing the formal elements of utility theory had, for Savage at least, become prior to capturing preferences *per se*. He took the consistency conditions to define what his preferences ought to have been, rather than taking his preferences to define any formal conditions of choice. This reasoning takes the axioms to impose normative constraints on choice—they tell us what our choices should be—and in this sense they are taken to be prior to our actual subjective preferences. The normative purpose of utility theory seems to have eclipsed its function as an explanatory and predictive tool. Emerging out of the ashes of utilitarianism, utility theory sought to provide a practicable way of capturing people's perspicacious interests, against the perceived fantasy of the utilitarian's aim of measuring subjective states directly. However, this purpose seems to be largely forgotten if preferences are filtered through a set of consistency conditions. Attempts to launder preferences via the axioms fails to recognise that the systematic decision-making heuristics identified by Tversky and Kahneman open up a gap between preferences and the formal constraints posed by the axioms of rational choice. The axioms should be understood as external, normative constraints, rather than internal, neutral ones.

The intention of the discussion in this chapter is not to entirely undermine the use of preference elicitation as a measure of utility, but rather to draw attention to some

implications of this practice. If utility values attach not only to outcomes under a particular description, but also to outcomes in a particular survey setting, given a particular ordering and phrasing of questions, given a particular method of survey—by telephone, or face-to-face, or on paper—and so on, then utility values don't straightforwardly capture some state of persons. Instead they capture a relationship between the people whose well-being is being measured and the survey tool that is used to elicit preferences. They also reflect features of the environment within which preferences are elicited, and in particular the purposes and context of the people making the measurement. The people responsible for designing surveys and eliciting preferences have a crucial role to play in constructing the preferences of the people whose well-being they seek to measure. In non-experimental contexts, people's preferences won't be constructed in such a direct way by anyone in particular, but will nonetheless be responsive to the context within which they are formed, including features of their own mental state.

One implication of this discussion is that the utilities attached to given outcome states should be taken to range over outcomes under a particular description, or in a particular context. The way in which these outcomes are constructed is largely down to the decisions of those designing the surveys or measuring the preferences of participants. In the absence of a complete ordering of precisely-specified preferences, it's not possible to extrapolate from a utility value for an outcome generated using one survey design, or under one description, to a utility value for that outcome *per se*—that is, a state of being, rather than a description of a state of being. This implies that resulting utilities from one survey cannot necessarily be taken to say anything about the preferences over those outcomes, where these are understood as actual states that people are in, rather they tell us something about preferences over outcomes under a given description, or in a given context. This places a significant limitation on the generalisability of utility values—if utility values are related to the measurement

context in this complex manner, then they cannot be plucked out and transposed to other contexts, or taken to apply to outcomes under slightly different descriptions, without consideration of how this might affect the preferences in question.

## CHAPTER 5

### Subjective experience

My brief discussion of hedonism in chapter 3 indicated that I take hedonist theories to be a subset of objective list theories. While Sen and Nussbaum do explicitly take internal states to be partly constitutive of well-being, the paradigmatic doings and beings in their accounts are not psychological states. And while some objective list accounts *do* explicitly specify goods that are partly constituted by positive psychological attitudes, very little is said about their identification and measurement (Fletcher 2013). The measurement of subjective experience raises questions and concerns that are different to those raised by the measurement of other objective goods, and it is for this reason that I provide a separate treatment of hedonist theories in this chapter. Despite being objective theories of well-being, which take the constituents of well-being to be good for us independently of whether we think they are or not, there is a sense in which hedonist theories are manifestly subjective. Hedonist theories typically take well-being to be identical with some particular mental state or attitude, which means that well-being cannot be identified or measured without examination and appraisal of people's subjective experience. Some of the discussion in this chapter will also be relevant to desire-satisfaction theories of well-being, insofar as it concerns the measurement of subjective psychological states and attitudes in general, rather than feelings of pleasure or attitudes of enjoyment in particular.



Hedonist theories of well-being, in general, take people's well-being depend on how they feel or what they experience. Traditional hedonist views take well-being to consist in a person feeling happy, or having other pleasurable feelings (Crisp 2006). Attitudinal hedonism takes well-being to consist in a person having a positive attitude, such as enjoyment, towards states of affairs (Feldman 2002). More idealised theories take well-being to consist in a person having desirable subjective states—enjoyment, happiness—where those states are not based on ignorance of the facts, and are freely desired and chosen (Sumner 1996). Hedonist psychology has undergone something of a revival in recent years, as technological and statistical developments have introduced the apparently genuine possibility of directly measuring subjective states. One of the central attempts to defend a neo-hedonist conception of well-being—*experienced utility theory*—is proposed by Kahneman and his collaborators as a way of avoiding many of the problems associated with preference elicitation models discussed in the previous chapter (Dolan and Kahneman 2008; Kahneman 2000a, 2000b; Kahneman and Thaler 2006). Unlike the preference elicitation model discussed in the previous chapter, which is explicit about its attempts to be nonspecific about the content of preferences when eliciting utility values, the experienced utility model takes utility to consist in positive subjective psychological states. For Kahneman, well-being is characterised by a greater number and a greater intensity of positive affective states, and a lower number and lesser intensity of negative affective states (Kahneman 1999). Other psychologists take well-being to be best understood in terms of the composite concept of 'subjective well-being,' which comprises high positive affect, low negative affect and high life satisfaction; still others take more specific kinds of subjective experience, such as flow states or fulfilment, to be constitutive of well-being (MacLeod 2015).

In this chapter I chiefly focus on Kahneman's theory of well-being. In line with my strategy in the previous two chapters, I have selected this account because it is a well-

developed attempt at operationalising well-being, which takes seriously the need to identify and measure well-being in people. Kahneman takes a fairly ecumenical approach to hedonism. While he defends the view that well-being should be understood in terms of moment-by-moment subjective experience, where positive subjective experiences improve well-being and negative subjective experiences detract from it, he does not place restrictions on the kind of positive subjective states that can constitute well-being. He also considers a number of different tools for measuring subjective psychological states, some of which take well-being to consist in particular feelings or emotions, and others which seem to characterise it in terms of positive attitudes. Consideration of Kahneman's approach will, then, help to shed light on various possible hedonist approaches. In fact most of my discussion focusses on the direct measurement of subjective experience in general, rather than the measurement of any specific feeling or attitude, so will be relevant to many varieties of hedonism.

The chapter proceeds as follows. In 5.1, I outline Kahneman's experienced utility theory. Kahneman distinguishes between two distinct ways of measuring experienced utility. *Remembered utility* is a memory-based approach, which takes an individual's retrospective assessment of past states to reveal their experienced utility in those states. *Moment or instant utility* is advocated by Kahneman as a more accurate and reliable measure of experienced utility. It attempts to capture a series of real-time measurements of the 'valence'—that is, positive or negative value—and intensity of subjective psychological states throughout an episode, from which a utility profile is subsequently derived. These two metrics result in different, and sometimes contradictory, measurements of utility. In 5.2 I describe the experiments which Kahneman takes to show the primacy of moment utility over remembered utility. In the following three sections, I go on to critique Kahneman's account of moment utility on three accounts. First, in 5.3, I argue that discrepancies between reports of moment utility and remembered utility do not in themselves suggest the primacy of moment

utility as a more accurate or representative measure of utility. Rather, the argument in favour of moment utility depends on the inviolability of *temporal monotonicity* in measures of utility. Temporal monotonicity is the rule that increasing the duration of a painful event should not improve its overall utility (Kahneman and Thaler 2006, p.228). I argue that temporal monotonicity about utility assumes that utility is understood in hedonic terms, and so should not necessarily be taken to apply to remembered utility. Secondly, in 5.4, I show that Kahneman fails to identify a clear distinction between remembered and moment utility. I argue that there is reason to think that Kahneman's own *day reconstruction method* is not in fact a measure of moment utility. Finally, in 5.5, I argue that measures of moment utility fail to avoid the focussing and anchoring effects observed in measures of remembered utility and decision utility. Moment utility is favoured as a measure of utility in part because it is alleged to avoid many of the measurement biases of other utility metrics, so this raises significant problems for the theory. In 5.6, I briefly sum up, arguing that expected utility theory fails to provide a method of measuring subjective psychological states *directly*. There are a number of different ways of framing experience which result in different measures of experienced utility. Instead of treating one time point as most representative of subjective psychological states, we should reject the idea that there is a neutral or correct way of measuring experienced utility.

### 5.1 *Experienced utility*

Kahneman's experienced utility theory finds its roots in Bentham's notion of hedonic utility, and the classical utilitarian claim that people ought rationally to desire, and to act so as to achieve, things that maximise their utility (Read 2007, p.46). Benthamite models of hedonic utility lost traction against rational choice theory models of decision utility in the 20th century in part because of the apparent impossibility of observing and measuring subjective experience of pleasures and pains (Kahneman et al. 1997, p.375; Read 2007, p.46). In the early 1990s, however, Kahneman revitalised a

hedonic account of well-being with his *experienced utility theory*, an attempt to recast the measurement and interpretation of hedonic utility in the language and methods of modern psychology (Nussbaum 2008, S82). Experienced utility theory proposes a way of understanding utility which, its proponents claim, allows for the direct measurement and comparison of subjective psychological states (Kahneman et al. 1997).

Like Bentham, Kahneman takes utility to be best understood in terms of hedonic experience, where experiences are characterised as pleasurable and painful, or experiences that a subject wants to continue and those that they want to stop (Kahneman et al. 1997, p.379). Utility is, on this view, a sensation, which can be measured on a single axis, from painful to pleasurable, with a neutral midpoint (Nussbaum 2008, pp.82–3). The utility associated with different kinds of experiences is taken to be quantitatively variable, enabling comparisons, rankings and aggregation of different activities and states. Maximising utility is thus understood as individuals making decisions that maximise the amount of pleasure they feel (Kahneman and Thaler 2006 p.222). Experienced utility—as in utility theory—is defined in terms of subjective psychological states, but—in contrast to utility theory—those outcomes which maximise experienced utility are not necessarily coextensive with preferred or chosen options (Read 2007, p.51). This means that someone can make a mistake about what will maximise their utility, or systematically be mistaken about what will maximise their utility, while still acting rationally, that is, by continuing to *try* to act so as to maximise their total utility. If cognitively normal decision-makers do not maximise experienced utility through their choices or preferences, then the utility theory model of utility comes apart from experienced utility, where this is understood in hedonic terms (Kahneman et al. 1997, p.375). Like utility theory, experienced utility theory claims that individuals ought to and generally do attempt to act so as to maximise their utility. Experienced utility theorists, however, do not think that

optimal utility for an individual can necessarily be read off their choices; instead, it has to be measured more directly, by measuring subjective experience.

Kahneman rejects Bentham's view that pleasure and pain are 'sovereign masters of human action' and does not endorse a fully Benthamite utilitarian reasoning about utility, that is, he doesn't endorse maximisation of utility as the ultimate goal of morality, or as a benchmark for evaluating all decisions (Kahneman et al. 1997, p.377). Instead he takes experienced utility theory to have rather a smaller scope, claiming that there are some situations where experienced utility is a relevant criterion for evaluating outcomes, but that a separate judgement is required to determine whether this is the case. Though Kahneman does not take experienced utility to be the sole constituent of well-being, it nonetheless forms a significant part of it. He suggests that the concept of well-being that is used in individual and public policy decision making should prioritise immediate subjective psychological states above remembered well-being or other global assessments (Kahneman 2000b, pp.691–2). In his own words: 'There is an obvious sense in which present experience is real and memories are not' (2000a, p.692).

Kahneman argues that experienced utility can be measured by constructing a picture of the *total utility* of an event or state, composed of a series of measurements of utility on a moment by moment basis throughout its duration (Kahneman 2000a, p.677). The basic unit of experienced utility is *moment utility*, which is a temporally contextualised measurement of someone's hedonic state. Of course, each moment or episode is not without extension, but is a short snapshot of the psychological present—lasting, it is suggested, up to three seconds (Kahneman and Riis 2005, p.285). A group of consecutive episodes form a *temporally extended outcome*, and each time-point is assigned a level of moment utility to give a *utility profile* for the outcome (Kahneman et al. 1997, p.389). The total utility for that outcome is the integral of the utility profile, that is, the area under the curve if the utility levels of all episodes were mapped out on

Cartesian coordinates. On the experienced utility account, there is no need to account for adaptive preferences, or issues like hunger, arousal, anger or tiredness biasing judgements, because all of these will be encompassed in the record of total utility. Reported experienced utility at any given moment will reflect all relevant aspects of experience, including the state that the individual is in when making the report, and their anticipation of future states.

Experienced utility theory makes three explicit normative assumptions about the measurement of total utility. The first is *temporal monotonicity*, the rule that increasing the duration of a painful event will not improve its overall utility (Kahneman and Thaler 2006, p.228). Adding an unpleasant experience to the end of an event should make the overall experience worse, while adding a pleasant experience should improve the overall experience. The second is *separability*, the idea that the order in which moment utilities are experienced is not relevant to the total utility of an outcome; the third is *time neutrality*, that all moments are weighted equally regardless of where they come in the utility profile (Kahneman and Riis 2005, p.290). These latter two assumptions might at first seem to be improbable claims—of course the order in which events and psychological states happen matters to utility, and clearly some moments contribute more to overall utility than others.

However, advocates of experienced utility claim that measurements of moment utility already take into account the impact of the order of events and the fact that some moments affect the utility of other moments, and so these don't need to be accounted for separately by weighting moments based on their ordering (Kahneman and Riis 2005, p.293). So, for example, if someone reports a utility profile made up of three consecutive episodes of moment utility, measuring 50, 40, and 30 respectively, experienced utility theory says that the utility of this profile is identical to a profile made up of three consecutive episodes of moment utility measuring 30, 40 and 50 respectively. However, the stimulus required to generate these levels of utility may not

be the same in both cases, so, in order to generate an episode with utility level 40 after an episode with utility level 30 may require a different degree of stimulation than that required to generate an episode with utility level 40 after an episode with utility level 50. Because experienced utility theory aims to capture only the intensity of the sensation of pleasure and pain experienced and not anything about its content or the activities which precipitated it, an individual's declaration that she experienced a moment utility of 40 captures all of the relevant information about the utility experienced in previous moments. In this way, measurements of moment utility are made within their temporal context and so can be taken to measure total utility, despite not measuring the way or extent that individual actions or inputs contribute to that utility (Read 2007, p.53). Similarly with respect to time neutrality: of course some moments are more important than others, insofar as they have greater impact on the utility of other moments, but this is captured by the utility levels of each moment (Kahneman and Riis 2005, p.293). Therefore, moments should not be assigned more or less importance based on when they occur in the series, as the recorded utility levels of each moment should reflect this.

There are a number of different ways of measuring moment-by-moment experienced utility discussed in the literature. One is the *experience sampling method* (ESM), where participants carry a handheld computer, which randomly prompts them to answer questions about their current psychological state throughout the day (Kahneman and Riis 2005, p.291; Larson and Csikszentmihalyi 1983). Another method for measuring experienced utility is the *day reconstruction method* (DRM) where people are asked to divide their day into a series of episodes, based on the different activities they performed, and to rate their 'affect' during these periods (Dolan and Kahneman 2008 pp.228–30). Of course, these 'moments' or episodes have a longer duration than the psychological present, but the conjecture behind this approach is that people have a privileged access to their mental states and affect at points in time close to the time at

which they occurred, which they do not necessarily have at later points in time. Kahneman also refers to a method of measuring experienced utility where subjects report the intensity of the pain of a medical procedure at 60-second intervals, in real time (Kahneman 2003, pp.713–4). All of these methods assume that individuals are able to judge the utility of each moment consistently and coherently (Kahneman and Riis 2005, p.291). If individuals are inconsistent in their judgement of pleasure and pain across time, then it would not be clear that the same thing was being measured each time, or being measured in the same way. There seems to be an assumption, then, that individuals cannot be wrong—or at least are not, in general, wrong—about the type of subjective experience they are having or its intensity.

Kahneman and Riis also gesture towards the physiological measurement of utility, citing an index of ‘prefrontal cortical asymmetry in the electroencephalogram,’ developed by Richard Davidson, which can predict affective responses based on observation of the electrical activity in the brain (Kahneman and Riis 2005, p.292). In one of Davidson’s experiments, electrical activity in the brain and facial behaviour were recorded via non-invasive means during periods of emotional stimulation, elicited by short film clips designed to stimulate happiness and disgust (Kahneman and Riis 2005, pp.132–135). Participants reported the primary emotion elicited by films, and this was compared with the recorded findings of brain activity and facial expression. The study showed that feelings of disgust are associated with more activation in the frontal region of the right hemisphere than is observed with feelings of happiness, and that feelings of happiness are associated with more activation in the frontal region of the left hemisphere than is observed with feelings of disgust (Davidson 1992, pp.133–4). Davidson also notes that despite consistency in the direction of the asymmetry of brain activity between feelings of happiness and of disgust, there was large variation between individuals, in terms of both the magnitude and the direction of asymmetry observed (Davidson 1992, p.139). So, for example,



clinically depressed subjects exhibit less left frontal activation than non-depressed subjects when measured at a resting baseline. These findings seem to support the claim that prefrontal cortical asymmetry is a measure or approximation of affect. Kahneman and Riis argue that the correlations between Davidson's measure and measures of evaluated well-being indicate that EEG measurements can be used as an 'anchor' to calibrate and verify reports of experienced and evaluated utility (Kahneman and Riis 2005, p.298). They claim that such physiological measures are not susceptible to linguistic bias, so don't come up against some of the framing effects which haunt other measures of utility.

Kahneman (2000a) contrasts the moment-by-moment view of experienced utility with a memory-based approach, which takes a subject's retrospective evaluations of events to be valid data about their subjective experience. The *remembered utility* of a given event is assigned by the subject's retrospective assessment of the situation. Remembered utility is, in some sense, a measure of experienced utility: like moment utility, it is a measure of the actual utility attached to an event, as judged by the person who experienced it. However, remembered utility differs from moment utility because, as the evaluation occurs after the event has finished, remembered utility judgements are global evaluations—evaluations of the event as a whole—rather than evaluations of separate parts of it pieced together (Kahneman 2000a, p.673). Moment utility and remembered utility need not coincide, as someone's memory of an experience may be different from their direct reports of experience. These differences will be discussed in some detail in the following section.

Advocates of remembered utility take well-being to consist in a favourable overall assessment of a particular event or period of time, rather than requiring overall positive subjective psychological experiences throughout the episode. Kahneman's description of the memory based approach as privileging a 'retrospective assessment' of an event leaves open an ambiguity. On the one hand, the retrospective assessment

might be a matter of recalling my subjective psychological state during a particular event or time period. On the other hand, a retrospective assessment might involve reflective consideration of whether a particular experience was, in hindsight, good for me, or whether it was a happy one. The former interpretation *does* suggest a global assessment which tries to capture something similar to Kahneman's experienced utility—a measure of how a particular experience feels—just from the perspective of a different time point. The latter suggests something quite different. It might be understood to be trying to capture something quite different, what Nussbaum (2008) calls a 'reflective judgement' and what Alexandrova (2005) calls 'subjective well-being.' Such an understanding of well-being might allow that a trialling event or period of one's life—such as battling a disease, ending a relationship with a long term partner or struggling to write a book—is considered to have been a good thing, all things considered, despite the physical and emotional pain and suffering which characterised it. On such an account of well-being, positive subjective psychological experience does not necessarily lead to well-being, indeed it might be the case that some degree of striving and suffering is necessary for well-being (Nussbaum 2008, p.87). This act of reflection entails an evaluation of psychological experiences in the context of those things that an individual takes to be valuable, or their conception of success, flourishing or self-actualisation. It may involve discounting or re-evaluating the way an experience felt, in light of what it resulted in, or why those feelings came about. Reflective utility, then, can be understood as prioritising the commitments—sometimes explicitly normative commitments—and context in light of which experiences are retrospectively judged over immediate subjective experience.

It is not clear whether Kahneman is thinking about one or the other of these perspectives in particular when he talks about remembered utility. In reality, the difference between the two may be somewhat indistinct. Memories of past experiences will inevitably involve a certain amount of reflection—remembering something

involves framing it as in the past, as completed, and contextualised in light of subsequent events and experiences. This suggests that even if retrospective assessment is merely understood as recall, it necessarily captures something different from moment utility—a global assessment, rather than a moment-by-moment awareness.

## 5.2 *Divergent experienced utility values*

This section will outline the observed divergence between measures of moment utility and remembered utility. Kahneman and his many collaborators argue that measures of remembered utility violate temporal monotonicity, and are thus not accurate measures of utility (Dolan and Kahneman 2008; Kahneman 2000a; Kahneman 2003; Kahneman and Krueger 2006; Kahneman and Riis 2005; Kahneman and Thaler 2006; Kahneman et al. 1997). They claim that moment utility, which conforms to temporal monotonicity, should be prioritised as a more accurate measure of utility in both personal and political decision making.

Kahneman's experienced utility theory takes moment-by-moment subjective evaluation of experience to be more reliable than remembered utility, and also takes subjective states to be measurable to a high degree of accuracy (Hausman 2015 p.106). While Kahneman and Jason Riis (2005, p.289) acknowledge the importance of remembered or evaluated utility in decision making, they argue that it does not always reflect the actual experience in question. When we use our past experience as a guide for future decision making, we do so using our memory of what our previous experiences were like, and if this is biased or incorrect then we may continue to fail to maximise the utility of future experiences. Kahneman and Richard Thaler (2006, p.227), for example, argue that evaluations of past events are biased because they are anchored to the psychological state the individual is in at the time of the evaluation. In one study, reported life satisfaction was raised significantly for respondents who found a dime, which had been planted by the research team, shortly before answering the questionnaire (Kahneman and Krueger 2006). Another study found that reported life

satisfaction was influenced by the weather (Kahneman and Krueger 2006). A third found that high levels of pain on the day of assessment affected reports of pain experienced over the past week (Stone et al. 1997). These results indicate that remembered evaluations reflect the affective state that the respondent is currently in, and so do not reliably reflect what the evaluated experience was actually like.

One of the key examples used to illustrate the primacy of experienced utility over remembered utility is an experiment on patients undergoing either colonoscopy or lithotripsy (Redelmeier and Kahneman 1996). During the procedure, patients were asked to report the pain they were currently experiencing on a scale of zero to ten (no pain to intolerable pain) every 60 seconds. Then, within an hour of the procedure ending, they were asked to provide an assessment of the pain they had suffered during the episode as a whole. Colonoscopy patients made a further retrospective evaluation of the pain felt during the procedure one month later, and lithotripsy patients made a further retrospective evaluation one year later. As both interventions are physically unpleasant, it seems reasonable to expect that those patients who had longer procedures would experience *more* pain than those who has shorter procedures. However, there was no correlation found between global evaluations of pain and the duration of the procedure. Instead, global evaluations were better explained by the *peak-end rule*: the remembered utility of pleasant or unpleasant episodes is predicted by averaging the most extreme moment utility value recorded during an episode (its 'peak') and the moment utility value recorded at the end of the experience (Kahneman 2000b, pp.697–8; Kahneman and Thaler 2006, pp.227–8).

A similar experiment explicitly tested the impact of adding an additional, but less painful, part to the end of a colonoscopy procedure (Redelmeier et al. 2003). Patients were randomised to two groups: one in which the probe was removed at the end of the procedure, and another in which the tip of the colonoscope remained inserted for up to three minutes at the end of the procedure. Again, patients reported the level of pain

they were experiencing at 60 second intervals throughout the procedure, and provided retrospective global assessments of the discomfort experienced from the procedure (Redelmeier et al. 2003, p.189). The presence of the probe was uncomfortable, and patients reported continued—though decreasing— levels of pain in the final part of the episode. Earlier parts of the procedure were given similar pain ratings by patients in both groups. Again, it seems reasonable to expect that those patients who had the probe left in for longer would experience *more* discomfort than those who had the probe promptly removed—they had effectively the same procedure plus another procedure that was less painful, but still somewhat painful. However, when making a retrospective evaluation of how painful the experience was, those patients who had the probe left in for longer reported the experience to be *less* painful than those in whom it was inserted for a shorter period of time, a phenomenon which is again explained by the peak-end rule (Redelmeier et al. 2003, p.190).

Kahneman and Thaler (2006, p.228) argue that the peak-end rule violates *temporal monotonicity*, the rule specifying that increasing the duration of a painful event will not improve its overall utility. The longer interventions had the same total disutility as the shorter interventions, plus an episode of additional discomfort. It would seem, then, that participants are exhibiting preferences for a dominated option, that is, one that is superior in no respects and inferior in at least one respect. Furthermore, participants recognise the validity of temporal monotonicity, and, when they are explicitly made aware of it, conform to the rule in their choices and utility predictions (Kahneman and Thaler 2006, pp.228–9). Participants’ assessment of remembered utility also violates time neutrality and separability by using a peak-end rule to estimate utility. Rather than basing their evaluation of an event in terms of its total utility, participants are in effect prioritising certain moments in the event (the moment with highest utility/disutility, and later time points) over others, and prioritising later time-points over earlier ones.

Kahneman (2000a, p.675) claims that, looking at a construction of the total utility of patient A, who had the probe removed earlier, and patient B, who had it removed later, '[i]t is immediately apparent...that patient B had a worse experience than patient A,' and that '[n]o one would deny that it is generally better for a colonoscopy to be short than to be long.' Kahneman essentially means by this that the area under patient A's utility profile is smaller than the area under patient B's utility profile. He takes it to be obvious that the real-time or close-to-real-time experience of less pain or more pleasure should be prioritised over a later, remembered claim of more pain or less pleasure. Therefore, he takes there to be something incorrect or misguided about the evaluation of the longer episode as having higher utility overall, given a commitment to temporal monotonicity and time neutrality.

A second example, where experienced utility and reflected assessments of well-being seem to come apart is an experiment conducted by Kahneman which showcases the DRM (Kahneman et al. 2004). In this experiment, 909 working Texan women were asked to write a diary of the previous day, dividing it into a number of episodes, noting how long the episode lasted, what they were doing, with whom. They were then asked to answer structured questions about the episode: when it occurred, what they were doing (selecting from a list of 16 activities), where they were, who they were with, and how they felt, using 12 'affect descriptors,' each on a scale of zero to six. These retrospective appraisals were taken to be good approximations of experienced utility, being close in time to the activity, and to provide a detailed account of the utility levels associated with different activities in the respondents' lives. The participants reported the experienced happiness of childcare to be little better than the lowest rated activities of housework, working, and commuting, and somewhat lower than watching television, eating, relaxing, socialising and the comically euphemistic 'intimate relations.' This is in direct contrast to surveys of attitudinal happiness where time spent with children is reported to be greatly satisfying compared to activities like

housework, and is also in tension with commonly held beliefs about valuable activities and life choices (Hausman 2015, p.112).

As in the colonoscopy experiment, the suggestion is that the day reconstruction method represents how different episodes and activities actually make people feel more accurately than remembered utility through attitudinal surveys, or decision utility through preference elicitation, and so better shows their impact on people's well-being (Dolan and Kahneman 2008, p.230). If people misjudge the utility actually experienced in a particular episode when they make global assessments, either due to use of heuristics such as the peak-end rule, or due to other kinds of biases, such as social pressures, then what these retrospective assessments measure is not *actually* utility. Kahneman and Alan Krueger (2006, p.21) suggest that people should reduce the amount of time they spend performing unpleasant activities in order to reduce misery, and that policy interventions should be introduced to support this end. They stop short, however, of suggesting that people refrain from having children or reduce time spent with them, instead advising that 'the potential benefit of these interventions needs to be compared with their costs' (Kahneman and Krueger 2006, p.21).

### 5.3 *Temporal monotonicity and remembered utility*

The previous section outlined how the rejection of remembered utility in favour of experienced utility depends on the claim that the former violates temporal monotonicity (Kahneman and Thaler 2006, p.228). In this section, I will argue that the violation of temporal monotonicity exhibited in measures of remembered utility does not in fact reveal the primacy of experienced utility. Temporal monotonicity assumes that utility is understood in terms of hedonic states, but we need not expect remembered utility to conform to this rule. I will show that Kahneman tries to evaluate measures of remembered utility as if they are measures of moment utility, and it is this that leads to him labelling them less accurate and representative of experienced utility than measures of moment utility.

Kahneman (2000a, p.676) claims that ‘no one would deny that it is generally better for a colonoscopy to be short than to be long.’ This seems to amount to the claim that temporal monotonicity is intuitively or obviously true. He also claims that an ‘idealized objective observer’ would conclude that ‘[i]ncreases of moment-utility do not decrease the global utility of a utility profile’ and that ‘[i]n a concatenation of two utility profiles, replacing one [total utility] profile by another with a higher global utility will increase the global utility of the concatenation’ (Kahneman 2000a, p.680; Kahneman et al. 1997, p.390). However, there are plenty of cases in which (presumably non-ideal) people’s evaluations of their own states are incompatible with temporal monotonicity. Indeed, in Kahneman’s own example, the reports of the very people who judge the longer colonoscopy to have been less painful overall, despite reporting more pain from moment to moment, violate temporal monotonicity. How should we treat this apparent evidence that self-reported utility violates temporal monotonicity? There are three options: first—the route that Kahneman takes—to say that these self-reports are in some sense mistaken; secondly to reject temporal monotonicity altogether; or thirdly to show that in fact we should take retrospective reports of utility to measure something different from direct measures of utility. I will argue for the latter: to conclude that experienced utility is a more reliable or primary measure of utility fails to capture what is shown by the examples, which identify points of contrast between different reports of utility. Remembered utility captures a sense of utility which should not be expected to conform to temporal monotonicity. This should not imply that temporal monotonicity is false, just that it does not necessarily apply outside of the context of moment utility.

Someone participating in an experiment which elicits a violation of temporal monotonicity may have their attention drawn to the fact that the total experienced disutility of, for example, patient B (who had the probe left in for longer) is greater than that of patient A. People generally subsequently agree that they should indeed



prefer the shorter and less painful intervention, or that they should have judged the shorter intervention as bringing about less disutility than the longer intervention (Kahneman and Thaler 2006, pp.228–9). Kahneman and Thaler seem committed to the idea that evaluating the utility of an episode from this third perspective, namely assessing utility by comparing graphical representations of utility profiles, or descriptions of utility profiles, verifies the true utility associated with an episode. They argue that the fact that individuals, on reflection, recognise and regret the violation of temporal monotonicity is an indication of the primacy of moment utility over remembered utility. But far from being conclusive evidence that experienced utility *does* better capture the utility of a given episode, this example seems only to tell us that different methods of judging utility result in different evaluations of overall utility. That some of these methods corroborate each other, or that their evaluations of utility are correlated, does not in itself discredit other, non-corroborating accounts of utility. If evaluations of experienced utility and remembered utility are not in agreement, an assessment of which one provides the better measure of utility may depend on the time at which the assessment is made and the purposes of assessment, as much as any facts about the authentic character of utility.<sup>46</sup>

The experienced utility theorist might argue that the violation of temporal monotonicity in measures of remembered utility shows that experienced utility is more appropriate in those instances where this effect is noticed. However, the claim

---

<sup>46</sup> Kahneman and Riis (2005) in fact make a somewhat similar point, and acknowledge that this idea is in tension with Kahneman's earlier view that measures of evaluated well-being are simply flawed indicators of experienced utility. They counsel that remembered utility and experienced utility should be measured separately. I am broadly in agreement with this new view, which casts a significant shadow over Kahneman's earlier conclusions. However, in a 2006 paper Kahneman and Thaler revert to calling temporal monotonicity an 'an elementary principle of rational evaluation,' which appears to once again relegate remembered utility evaluations to the realm of irrationality (2006, pp.227–8).

that remembered utility violates temporal monotonicity implies that temporal monotonicity (as well as time neutrality and separability) is a rule which applies to remembered utility as well as to moment utility. Kahneman supposes that temporal monotonicity applies to all experienced utility, but there is reason to think it only applies to certain examples thereof. Namely, because of the restrictions that temporal monotonicity places on the relation between the duration of the experience and pleasure, it looks as though the concept of utility that it uses is one which explicitly conceives of utility as additive. It may be the case that increasing the duration of a painful event cannot improve its total utility profile, but this doesn't mean that increasing the duration of a painful event does not improve its overall remembered utility, if we take the two to differ in a relevant sense. When Kahneman claims that the violation of temporal monotonicity points to the inaccuracy of measures of remembered utility, he is judging the accuracy of remembered utility by standards relevant to the measurement of moment utility, such as the shape of the total utility profile. That is, he assumes that 'accurate' remembered utility is identical with total utility, where this is understood in terms of moment utility.

When people make retrospective assessments of events or time periods, one should not assume that they are trying, and failing, to report total utility. A retrospective assessment of the pain of a colonoscopy might be affected by how the experience chimed with my expectations, whether it was longer or shorter, more or less frightening that I had anticipated, how I felt afterwards, and by comparisons with other similar experiences. Some of these assessments may only be able to be made after the procedure has ended. These ways of evaluating an experience are not irrelevant to well-being, and do not seem to be able to be captured through the measurement of moment utility alone. Even if each measurement of moment utility captures the relation between that moment and previous moments, so in some way accounts for

past experiences, moment utility cannot capture re-evaluation of the experience of past moments, or at best can only capture it in terms of present experience.

Indeed, when faced with evidence that a longer colonoscopy will be more painful at the time, but will be remembered as less painful than a shorter intervention, it is not clear which procedure it is more ‘rational’ to choose, or which procedure brings about the greatest utility. Assuming that the colonoscopy episode has a significantly shorter duration than my subsequent memory of the episode, there may be good reason to think that the longer procedure will have a less negative impact on my self-perceived well-being. Given that the peak-end rule is a reliable predictor of reported utility, why not take it into account when deciding between unpleasant or pleasant episodes? If a period of pain added to the end of an episode will improve the remembered utility of the episode, perhaps I *should* choose the longer episode even though this may seem counter intuitive. As Daniel Hausman puts it, ‘[w]hy not regard the additional period of discomfort as an *investment* in achieving a life that seems to have had less discomfort in it?’ (Hausman 2015, p.113).<sup>47</sup> Whether Hausman’s suggestion is the right way to go about decision making might depend on factors such as the duration of the episode itself, the expected duration of life following the intervention, the level of pain or suffering experienced during the episode, and how this can be expected to impact on my mind and body in the future. Examples like the colonoscopy experiment should not straightforwardly lead us to the primacy of experienced utility—indeed there may be equally strong intuitions which lead to the primacy of remembered or reported utility. All it shows is that there are competing accounts of utility, given at different time points.

Kahneman and Alan Krueger (2006, p.21) suggest that individuals should spend less time doing things that they do not enjoy, in order to improve their overall utility. This

---

<sup>47</sup> Hausman’s emphasis.

means that people should maximise their experienced utility, so, according to the DRM experiment, Texan women should spend more time socialising, having sex, relaxing, eating and watching television, and less time cleaning, working and looking after their children (and perhaps not have them at all). However, this seems to prioritise certain kinds of immediate well-being or satisfaction, whilst failing to recognise the significance of what Kahneman and Riis (2005, p.287) call the ‘eudaemonic conception of wellbeing,’ which is measured in terms of longer term psychological properties such as life-purpose, self-actualisation and optimism.

To be fair to the proponents of experienced utility theory, there is some recognition of the limitations of experienced utility. Kahneman and Krueger (2006, p.22) claim that the potential benefits of interventions which reduce the amount of time people spend in an unpleasant emotional state, or otherwise improve total utility, need to be weighed up against their costs. Elsewhere, Kahneman claims that total utility ‘is not proposed as a comprehensive concept of human well-being, only as a significant constituent of it’ (2000a, p.691), and that a separate judgement is required to determine whether experienced utility is a relevant criterion for evaluating outcomes in a particular instance (Kahneman et al. 1997, p.377). There is, however, no direction given as to how to apply this cost benefit analysis, or how to decide whether experienced utility is the appropriate measure of well-being in a given case. In any case, Kahneman and Krueger (2006) clearly think experienced utility should play a much greater role in public policy decision making, and healthcare decision making, including a prioritisation of mental health interventions over physical interventions, and think that the outcomes of such decision making will be significantly changed by such a focus.

#### *5.4 Distinguishing remembered from moment utility*

The argument for experienced utility theory depends on the assumption that experienced utility can be clearly characterised, and can be distinguished from

remembered utility. There are some inconsistencies in the distinction that is drawn between remembered utility and experienced utility in the examples given, and none of the examples cited give a clear indication of how to identify a particular assessment as an instance of remembered utility or experienced utility. The most straightforward way of drawing this distinction seems to be the time at which the assessment is made: if the assessment is made at the time of the event, or very close to the time of the event, it measures experienced utility. If the assessment is made after the event has finished, it measures remembered utility. However, recall that in the colonoscopy studies, retrospective assessments were made within an hour of the end of the procedure (Redelmeier and Kahneman 1996, p.4), whereas in the Texan women's DRM experiment, reports made the following day were taken to be measures of experienced utility. If less than an hour's delay after the end of an event is enough for memory bias to creep in in the case of colonoscopy, why not in the case of the DRM? In both cases, the episode(s) had ended at the point of evaluation, so the possibility of participants using remembered utility heuristics such as the peak-end rule, or other framing effects impacting on their judgements, remains open.

Another way of distinguishing between instances of remembered utility and experienced utility might be on the basis of the type of questions that are asked of respondents, or the method by which reports of affect or utility are recorded. So, for example, experienced utility is characterised by questions which ask how a respondent feels now, exemplified by the ESM (Kahneman and Riis 2005, p.291), or by asking them to represent on a calibrated scale what level of pain they are currently experiencing, as in the Gottman–Levenson method of assessing pain, which was used in the colonoscopy experiments (Redelmeier et al. 2003, p.188). On the other hand, remembered utility is characterised by questions about the experience as a whole—the 'total pain or discomfort from the procedure' (Redelmeier et al. 2003, p.189)—or questions which require respondents to aggregate a set of experiences (Kahneman and

Krueger 2006, p.5). However the DRM does not seem to fare any better on this account either, as participants were required to recall the previous day, break it down into episodes, and evaluate how they felt during each episode. This involves a retrospective global assessment of a particular episode, and aggregation of affect experienced within each activity (Kahneman et al. 2004, p.1777). The episodes in the DRM case were no more than two hours long, whereas the average length of colonoscopy was under 30 minutes (Redelmeier et al. 2003, p.189). If the accuracy of a retrospective assessment of the colonoscopy is in doubt, then we should also question the accuracy of the assessment of the DRM episodes.

There seems reason, then, to question whether the DRM should in fact be taken to be a measure of experienced utility. The DRM study found that life circumstances (income, education, marital status) had little impact on reported enjoyment of activities, whereas other personal characteristics, such as amount of sleep the previous night and depression, did have a large impact (Kahneman et al. 2004, p.1778). Furthermore, local features such as time pressure at work and the identity of the people with whom respondents were interacting (partners, children, co-workers) also influenced reported enjoyment of activities (Kahneman et al. 2004, p.1779). So it seems that DRM does not measure global life satisfaction or general affect associated with life circumstances—which is perhaps not surprising as it asks respondents to consider their feelings about a particular event or activity—but this doesn't show that it measures direct experience. Unfortunately there is no direct comparison between DRM-reported affect and another experience-utility metric such as ESM. This makes it difficult to assess whether it does in fact possess all of the features which allegedly define measures of experienced utility: in particular, 'the goal of approximating the results of continuous real-time experience measurement,' 'accurate retrieval of specific episodes' and 'elicitation of the objective circumstances of episodes' (Kahneman et al. 2004, p.1779).

Some attempts are made to show that DRM does in fact have these features, by comparing the results of DRM with the results of existing experience sampling in similar contexts (Kahneman et al. 2004). But it wouldn't be surprising if the remembered utility associated with a particular activity were correlated with experienced utility for that activity: experienced utility theorists acknowledge that retrospective evaluations are expected to be correlated with measures of momentary psychological states, and the two methods of evaluation are not taken to be independent of one another (Kahneman and Riis 2005, p.287). The DRM does reflect results of experience sampling in its reportage of positive and negative affect (Kahneman et al. 2004, p.1777); however, elsewhere Kahneman and Krueger claim that retrospective evaluations do tend to correctly characterise episodes as good or bad, so this should not alone be taken as evidence that DRM measures experienced utility (2006, p.5). The study write-up also suggests that the fact that respondents' reports of their level of tiredness throughout the day was similar to patterns reported in experience sampling indicates that they accurately recalled their experiences (Kahneman et al. 2004, p.1778). Even granting that respondents did not base their accounts of tiredness on existing beliefs about how tired they are at different points in the day, this similarity isn't enough evidence to show that all episodes were accurately remembered.

There may not, then, be evidence, that the DRM is a good metric of experienced utility, or that it is not a measure of remembered utility. But this does not necessarily affect other, apparently more direct, measures of experienced utility, such as ESM and pain assessments. Though there may not be a precise line between moment and remembered utility, a working distinction may be able to be made on the basis of the time of evaluation and the kind of questions asked. That is, the closer to the event utility is measured, and the more specific the questions that are asked, the more likely a measure of utility is to be a measure of moment utility. On this basis, some of the

tools which capture immediate responses to events may be rightly said to capture moment utility. However, whilst these direct measures may avoid violating temporal monotonicity and time neutrality, they do not seem to be free from framing biases.

### *5.5 Framing and focussing effects in experienced utility measures*

One of the key reasons for which moment utility is preferred over other utility metrics is that it allegedly avoids some of the framing effects which afflict models of remembered utility and preference-elicitation models of utility. In this section, I will argue that measures of experienced utility that are plausible candidates for direct measures of experience avoid the violation of temporal monotonicity, but they remain open to other kinds of framing and focussing effects. As well as violating time neutrality and temporal monotonicity, experienced utility theorists argue that measures of remembered utility are open to a number of framing effects. These include anchoring reported utility to features of the reporting context and to the respondent's affective state at the time of reporting, and focussing on particular salient characteristics of the episode or in evaluating overall experience (Kahneman et al. 1997, pp.381–5; Kahneman and Krueger 2006, p.6; Kahneman and Thaler 2006, p.227–9; Hausman 2015, p.110; Oliver 2017). These are alleged to lead to misreporting of experienced utility.

The colonoscopy experiment is an example where the time of reporting is extremely close to the experience, with reports made every 60 seconds throughout the intervention, starting 30 seconds after the beginning of the procedure (Redelmeier et al. 2003, p.188). Measures such as this, and ESM, seem plausible candidates as measures of moment utility, if we assume that what is being asked of patients is for them to report their pain, affect or subjective psychological state in the psychological present, rather than over the last minute, or over the duration of the current activity. Rather than create a complete picture of the utility experienced by an individual, these methods take frequent snapshots and build up a picture of total utility by assuming



that the instant measurements of utility captured are broadly representative of utility levels over the episode or over a period of time.

Kahneman (2000a, p.692) takes present experience to be ‘real’ and tangible in a way that memories of past experience are not. There is an assumption in experienced utility theory that there is a reality as to an individual’s subjective psychological state, and that different measurement tools capture it in more or less detail. So, for example, an ESM survey which captures a respondent’s subjective psychological state a few times a day will create a very rough picture of well-being, which is accurate to the extent that the instants that it measures do in fact correctly record her utility levels at that point, but which are open to error because they are full of holes. The DRM—if it can be retrieved as a measure of experienced utility—may capture a rough approximation of subjective well-being, which is more complete than an ESM profile but more open to memory related biases. A measure like the pain assessment used in the colonoscopy experiments, which captures pain reports continually throughout a relatively short intervention, creates a much more detailed picture of well-being, and is less likely to give a misrepresentative account. However, it cannot be maintained over long periods of time, due to the intensive measurement requirements, and, presumably, can only be used to record affect with respect to particular activities. Furthermore, it only captures pain, which may give an incomplete picture of overall subjective experience. Nonetheless, these methods, assuming their validity and reliability, are essentially taken to capture the underlying reality of the respondent’s subjective well-being in some sense.

In the previous chapter, I argued that preference-based measures of well-being are subject to a host of framing effects. Respondent’s answers to surveys and questions are influenced by features of the measurement context that are not strictly relevant to their preferences, such as the ordering and context of the questions, the particular words used to describe outcomes, and the amount and type of information given. ‘Focussing’

effects occur where asking someone questions about a particular part of their lives, or a particular aspect of health status, or a transition from one state to another leads them to overstate the importance of that state or feature in the broader context of their lives (Dolan and Kahneman 2008, p.223). So asking someone to report her preferences relating to losing her hearing might lead her to focus on the transition period immediately after becoming deaf and the inevitable difficulties in communication that would involve, and to focus on the ways in which deafness will negatively affect her life, such as the inability to hear her favourite music, listen to the radio, or hear the voices of her partner and children. This may result in an exaggeration of the impact deafness would have on her overall well-being, by failing to recognise the ability of deaf people to have just as full and happy lives as anyone else. This was seen, in the previous chapter, in the study which showed that students' answers to questions about how their life is going in general changed in the context of questions about their love life.

It is not clear that measures of experienced utility are free from focussing effects. Respondents to experienced utility tools are asked questions about how they are feeling, which refer to particular aspects of their psychological state. Some tools require respondents to record their pain levels, others ask how they are feeling, or ask them to indicate the presence or absence of a number of different emotions or feelings (Kahneman and Krueger 2006, p.9). If systematic focussing and framing effects exist in the context of preference elicitation, then why not here? If a respondent is given a list of different emotions and feelings, and asked whether and to what extent she is experiencing them, she may lump together various related feelings under those headings, or may discount related but distinct feelings because she takes them to be relevantly different from the named feeling. So, when asked to report on whether she feels angry, she may conclude that she feels irritable and short-tempered, but wouldn't say that she was exactly *angry*, because she knows she is just in a bad mood, and her

annoyance is not directed at anything in particular. If she reports that she does feel angry, will she have misrepresented her feelings, and thus her well-being? Or will she have more gravely misrepresented her well-being if she reports that she does not feel angry?<sup>48</sup> Experienced utility theory does not seem to give an answer to this: there is an unquestioned assumption that introspection will reveal with clarity the psychological states that a person is experiencing, allowing her to accurately answer questions relating to her subjective well-being. But even if there is some sense in which introspection can reveal her psychological state and subjective well-being to her, it does not follow that she will be able to categorise that state in the way that is demanded by an experienced utility survey. Categories of emotion and feeling may be taken to refer, on the one hand, to specific emotional states which everyone can recognise whether they are or are not feeling or, on the other hand, to more general categories of emotional states which cover a wide range of descriptions of subjective psychological experience.

This issue is well illustrated by the DRM experiment with Texan women. The 16 categories of activity (including exercising, watching television, napping, shopping, childcare, working) were pre-specified by the researchers designing the study, and the participants categorised their daily tasks (to which they had previously given their own descriptions) against this list (Kahneman et al. 2004, p.1777). As the descriptions of the activities were pre-determined, the experiment seems open to obvious framing

---

<sup>48</sup> Leah McClimans (2010) discusses similar examples to this, where patients responding to quality of life surveys query the scope and terms of the questions asked. McClimans suggests that surveys about quality of life or subjective experience should not be taken as determinate, nor used to make determinate assessments. She argues that such tools should be used as a basis for better communication, rather than a way of establishing fixed answers to questions about quality of life.

problems.<sup>49</sup> If the category of ‘childcare’ had instead been called ‘spending time with children,’ women might have focussed on the positive aspects of looking after their children when rating the episode, rather than the tiresome aspects. If the category of ‘relaxing’ had been called ‘lying on sofa’ or ‘time spent doing nothing,’ again it would likely have resulted in different responses. The categories used to describe episodes of the day are not neutral—they already evoke positive or negative affect. Furthermore, it is difficult to know how respondents assigned their own descriptions to given categories when they initially constructed their diary of the previous day. If a respondent identified a period of time spent ‘relaxing with children,’ or ‘playing with children,’ or ‘watching TV with children,’ would they then assign this to the ‘childcare’ category (lower than average affect), or to one of the ‘relaxing,’ ‘socialising after work’ or ‘watching TV’ categories (all higher than average affect)? It is not clear exactly where the boundaries of the categories lie, whether there is any overlap, or even in some cases—‘intimate relations’ being the most obvious example—what exactly they refer to. The way that activities are assigned to categories may reflect pre-existing biases regarding the positive and negative experiences associated with particular activities, and may not be consistent across respondents. Furthermore, participants’ descriptions of their affect while performing particular tasks and activities may be coloured by their prior expectations of how they should feel when looking after their children, or working, or having sex. If such focussing effects exist in preference-based surveys, there is no reason to think they won’t also occur here.

---

<sup>49</sup> Even if the DRM is rejected as an adequate model of experienced utility, the use of finite lists of types of activities and of types of emotion or feeling is used in other measures of experienced utility, such as the ESM (Larson and Csikszentmihalyi 1983), so this critique remains pertinent to experienced utility theory more generally.

Even experienced utility measures which don't use finite lists of emotions and activities seem to fall foul of framing effects. Consider again the colonoscopy experiment, in which patients underwent an uncomfortable procedure, with an average length of almost half an hour (Redelmeier et al. 2003). The patients measured their pain by holding a device that controlled the position of a marker on a calibrated scale from 'no pain' to 'extreme pain.' The scale was displayed on a computer screen, and participants were prompted to rate their pain every 60 seconds. Were they asked to record how worried or anxious they felt, or how much they were enjoying themselves, or how they felt about their health, the resulting profile would very likely be different. This is unsurprising, as in each case, people would be asked to report on different aspects of their subjective psychological state. It may be that in a particular case, one or another of these aspects is most salient to well-being, so in the case of a colonoscopy, pain is the most relevant feeling, whereas in other cases depression, or anxiety, or an overall assessment of well-being (How are you feeling on a scale of 1–10?) might be more relevant. But if utility is understood as a measure of subjective psychological experience, then it should not be possible to know in advance of measurement the most salient feature of a person's psychological state against which to measure her well-being in a particular instance. Knowing how someone's psychological state pertains to their well-being seems to pre-suppose already knowing something about the subjective psychological state which she is in. Determining that pain is the relevant measure of experienced utility in a given case assumes an understanding of the subjective psychological states that contribute to utility in that instance.

Moreover, the focussing issues which affect stated preferences over outcomes seem also to come into play here. If a patient is asked to report every 60 seconds the precise amount of pain she is feeling, this will likely focus her attention on the pain associated with the procedure, which may lead to reporting of higher pain levels than would have

been noticed had she been chatting to the doctor, or looking out of the window, or thinking about what she was going to have for lunch. In surveys such as this, where respondents are asked to record their pain level very frequently over a protracted period of time, and in a way which requires them to concentrate, and to consider their responses, the pain associated with the procedure will unavoidably be one of the chief psychological states characterising their experience. It seems quite possible that this will affect the experience itself.

A further way in which the fact of measurement might impact on the subjective psychological state of respondents is if the survey and act of surveying itself starts to impede on the activities they are engaged in. The reason that measures such as DRM and ESM are proposed and used as measures of experience is because it is practically impossible to measure experience on a frequent basis in anything other than short isolated episodes. Measuring experience directly, as in the colonoscopy experiment, is highly invasive, and can only plausibly be done in instances where the participant does not have to be focussing on any other activities. ESM, which randomly asks participants to report their psychological state once during every two-hour block of time between 8am and 10pm for one week (Larson and Csikszentmihalyi 1983, p.23), is deemed to be very burdensome on participants (Kahneman et al. 2004, p.1776). If a participant's ESM device wakes her up, or interrupts a meeting, or disturbs her while she is trying to perform a difficult and intricate task it is quite plausible that her subjective experience will be changed by the fact of its being measured.

This could mean that there are only certain types of activity that can be measured by ESM or only certain times of day that the instrument can be used, in order to minimise such disruption. It may be thought to be reasonable to assume that instances of disruption will be randomly distributed across all participants and activities, and therefore will not affect the validity of any aggregate measures of utility. However certain kinds of activities— namely, ones that require a level of focus, or absence of

disturbance, such as sleeping, working, cooking, watching a film or a sports game—will be systematically affected by direct measurement of utility. This rules out a random distribution of disturbance, if the reason for disturbance is related to features of the activity in question. ESM requires participants to complete a report, when prompted, ‘if at all possible’ (Larson and Csikszentmihalyi 1983, p.23). If there are likely to be common features of those episodes which a respondent does not deem it possible to respond, the ability of ESM to capture the utility of those types of activity will be very limited, if not entirely precluded.

It looks, then, as if moment utility remains open to critiques that are very similar to those Kahneman levels at preference elicitation models of utility, on the grounds that the design of the measurement tool and act of measurement impacts on the measurements of utility that are generated. This is a problem for Kahneman, who grounds his support for experienced utility theory in part on the failure of decision utility to provide an unbiased account of preferences. If the fact of measuring experienced utility and the specific features of the measurement tools impact on the reported utility values generated by the survey, then measures of experienced utility will fail to reliably capture subjective psychological states in the promised manner.

### *5.6 The varieties of experienced utility*

The arguments in this chapter have endeavoured to show that there is not good justification for thinking that moment utility is a more accurate or representative metric of utility than remembered utility or preference-based models of utility, either on the grounds that it conforms to temporal monotonicity, or because it avoids framing effects, which lead to unreliable or misrepresentative measurements of utility. The fact that moment utility conforms to temporal monotonicity does not itself elevate moment utility above measures of remembered utility, as the latter uses a concept of utility which does not entail temporal monotonicity at all. And assessing the value of remembered utility in terms of temporal monotonicity judges it according to an

inappropriate standard. Accepting that moment utility is the best way of measuring utility in a given case requires some external reason for thinking that taking utility to consist in moment-by-moment subjective psychological states is the most reliable or accurate or useful way of conceptualising utility. Those examples which exhibit the divergence between remembered at moment utility only go to show that experience can be evaluated in different ways, using different measures of utility.

Nussbaum (2008) accuses Kahneman of failing to recognise the limitations of Bentham's conception of happiness, and not recognising that happiness, at least sometimes, requires reflection. Sometimes recognising value and genuine well-being may require a global assessment, contextualising moments of pleasure within a framework of values, and it is not controversial that there is value and well-being in things other than moment-by-moment pleasure. We all do things which don't improve our psychological states on a moment-by-moment basis, but which we take to be good for us, or to bring us happiness. Experiences which are not pleasant at the time may later be judged to have been good for us overall, or to have been necessary contributing factors to events, decisions and other psychological states or dispositions which, all things considered, we don't regret. Parents might gain satisfaction and a sense of achievement in procreating and bringing up children, which is not experienced in the everyday tasks of childcare and supervision, but which does nonetheless form a significant part of their self-identity and underpin some of their most important relationships. Or an adult might come to recognise the well-being that hours of much-resented piano practice have eventually brought her, resulting in a lifelong disposition to be diligent and focused in her endeavours, and a talent that brings pleasure to her and others around her. Factors other than immediate and moment-by-moment reported well-being might come into play in an assessment of the well-being associated with them, or in assessment of their value. Indeed, sometimes struggle and pain is a necessary part of achieving things of value—in some



instances, it may be that only through striving for something does it become a valuable end (Nussbaum 2008, p.87). Well-being may not be best understood as merely a moment-by-moment psychological state, but rather as—or also as—a longer-term assessment, or one which bears out over the course of a life-time.

Kahneman appears, at points, to recognise this, but nonetheless suggests that moment utility should be used as the basis for personal and policy decisions. The argument for this is not entirely clear. If he recognises that other forms of experienced utility, and other forms of non-experienced utility, are also constitutive of well-being, and legitimate forms of utility, then there is no reason to think that the divergence between moment utility and other measures of utility should lead to the prioritisation of moment utility. Experienced utility theorists' prioritisation of moment utility seems unprecedented, in light of evidence (evidence that experienced utility theorists themselves present) which seems to show that people's choices do not tend to maximise experienced utility. Kahneman and Tversky's prospect theory was disruptive and brilliant in its recognition of the systematic failure of utility theory to function as an explanatory and predictive tool. It exposed the axioms of rational choice as functioning as normative constraints, rather than descriptive constraints, on reasoning. The formal constraints imposed on preferences by utility theory conceal people's actual preferences, and thus, one of primary aims of utility theory is lost. In his defence of experienced utility theory, however, Kahneman seems to make much the same mistake. That is, he imposes a number of formal constraints on experienced utility—temporal monotonicity, time neutrality, and separability—even though the experiments he cites indicate that people systematically violate these constraints in assessments of their own subjective experience. Once again, a theory which ostensibly seeks to prioritise personal experience as a determinant of well-being ends up dismissing it in light of failure to meet external, formal constraints. The formal constraints end up playing a normative role, rather than a descriptive one.

If measures of moment utility are also subject to framing effects, then generalising on the basis of measures of moment utility, as with preferences, should proceed with caution. Moment utility, despite its claims to be more direct and unmediated than other measures of utility, nonetheless fails to bypass the unavoidable influence of the context of measurement. That moment utility is measured under some purposive description, in some non-neutral context, by a particular set of individuals, means that what is being measured is not best understood as a pre-existing state of persons. Rather, as in the case of preference-based measures of utility, measures of moment utility are highly dependent upon the context of measurement.

## Part III



## CHAPTER 6

### The construction of well-being

Though divergent in their subject matter, each of the last three chapters came to a similar conclusion about the conceptualisation and measurement of well-being. Whether well-being is understood in terms of preferences, subjective experience or objective goods, a close examination of attempts to identify and pick out well-being in particular people or groups indicates that what is being identified and measured is shaped, mediated, and to some extent determined by factors external to them. Well-being is not best understood as a property or attribute of persons which is discovered, but rather as something which is mediated and constructed. The small selection of examples that I have looked at by no means exhausts all of the possible ways of describing and evaluating well-being. However, my discussion of objective list, preference-based and experience-based accounts intentionally mirrors the conventional three-way characterisation of theories of well-being, in the hope of convincing readers of the occurrence of a similar phenomenon across the board. That is, whether we conceive of well-being in subjectivist or objectivist terms, whether we understand it in terms of desires, experiential states, or goods, well-being is in some sense constructed, rather than determined. In this chapter I will build on the specific findings and arguments of the previous three chapters to develop a relational conception of well-being, and to draw out some general conclusions about the nature of well-being.

The chapter proceeds as follows: in 6.1, I defend my positive thesis, that well-being is best understood as a representation of a relation between an object of measurement, a measuring instrument, and an environment. In 6.2 I discuss the centrality of the role of measuring subject in the well-being relation, and their role in theorising about the relation in order to produce information about well-being. In 6.3 and 6.4 I consider the possibility of a stance independent account of well-being. In 6.3 I discuss an argument for moral objectivity from John McDowell, which indicates one way of justifying stance independence about well-being. In 6.4 I discuss Richard Boyd's naturalist moral realism, which suggests another approach. I offer reasons to reject both of these views. In 6.5, I consider the objection that my account is unduly permissive, and doesn't sufficiently enable discrimination between better and worse theories of well-being and claims about well-being. I explore Alexandrova's suggestion that a speaker-relative conception of well-being should define the context of well-being claims in terms of the objective practical environment of the speaker, rather than the speaker's psychological state. I argue that conceptual promiscuity is not necessarily a bad thing, and suggest that my account nonetheless allows for discrimination between conceptions of well-being on the grounds of their appropriateness.

### *6.1 Well-being represents a relation*

The structural similarity of my analysis in the previous three chapters is, I hope, clear. In chapter 3, I argued that assessments of well-being via the measurement of objective goods are highly dependent on features of the measurement context, and particularly on the purposes and intentions of the people attempting to measure it. However the objective goods are defined, the identification of their presence or absence in a person involves the construction of an information base—the data which tell us whether someone has the good in their life and to what extent. All objective goods are multiply realisable across a variety—an indeterminate number, perhaps—of information bases, so defining the information base involves a series of normative and pragmatic

decisions which are not given in the specification of the good at a higher level. The measurement of objective goods is thus highly dependent on decisions made in the context of their description, identification and measurement, such that there isn't a fact of the matter about the presence and degree of the good outside of such a contextual description.

In chapter 4, I argued that consideration of the measurement of well-being via preference elicitation indicates that preferences don't range over outcomes, where these are understood to be description-invariant states of the world, but over descriptions of outcomes. Assessments of preference-based utility reflect features of the context of measurement, including not only the wording or phrasing of the statement of an outcome, but other features of the environment within which the choice is made. When we measure preferences we don't capture something which exists prior to the act of measurement. Preferences are thus also tied to the context of their elicitation, and to choices made by the people who attempt to elicit them.

Finally, in chapter 5, I argued that consideration of the measurement of well-being via the measurement of subjective experience should lead to a similar conclusion. Though there appears to be something more 'direct' about the measurement of experience, reports of subjective experience are also mediated by features of the context and act of measurement. When someone assesses their own subjective experience, they do so under a given description, within a given referential context, for a particular purpose—and all of this shapes the resulting reports. Again, there's reason to think that measuring subjective experience isn't a process of ascertaining something which exists prior to its measurement, and that reports of subjective experience are highly context dependent.

In each case, the outcome of the assessment of well-being—the well-being of the person under consideration—depends not just on some facts about the person whose

well-being is being assessed, but on features of the environment in which their well-being is being assessed, the tool with which their well-being is assessed, and, moreover, the attitudes, decisions and purposes of the person assessing their well-being. The resulting measures of well-being are constructed in and through the measurement process, rather than being discovered or determined through measurement. It was clear in all three cases that the conceptualisation of well-being by the people who design the surveys, or who determine the set of well-being-relevant goods and their information bases is central to the assessment of well-being that comes out at the end. This is obviously true in the sense that depending on whether well-being is measured via preference elicitation, measures of subjective experience, or objective measures, different values or assessments of well-being will result. The discrepancies between remembered utility, moment utility and reflective utility discussed in chapter 5 show that different conceptions of subjective experience also produce different assessments of well-being. The same is true for different objective list accounts. But, moreover, the arguments presented in all three cases suggest that the specific conceptualisation of well-being—for example, the way in which surveys are designed so as to elicit well-being values; the choice to focus on pain, life satisfaction, or happiness; the choice of functioning vector used to order capabilities—lead to different assessments of well-being. In each case, the aims, purposes and values of the people measuring well-being appear to be highly influential in shaping the measures, and so shaping the resulting assessments of well-being.

Conceptualising the measurement of well-being in terms of the model theory of measurement, which I outlined in chapter 2, helps to make sense of my analysis. The model theory of measurement takes measurement to involve an interaction between an object of measurement (the system being measured), a measuring instrument, and the measuring context, which includes the measuring subject (the person making the measurement). This relation produces measurement indications. But these indications



must be abstracted and modelled in order to result in any information about well-being. Central to this account of measurement is the idea that measuring something involves placing it in a *parameter space*. This involves characterising the object of measurement in terms of whatever quantity or attribute you seek to measure, and specifying a range of possible states in which the object could be or characteristics that it could have.

Applying this account of measurement to the measurement of well-being helps to make sense of the ways that well-being values are highly dependent on aspects of their measurement context. Measuring well-being—at minimum identifying whether someone is doing well or not, and perhaps also identifying the extent to which they are doing well, or how well they are doing compared to others—first involves an interaction between an environment, which includes a measuring subject (the person measuring well-being), a measuring instrument (for example a preference elicitation tool, a pain assessment instrument, or a means of identifying prudential goods), and an object of measurement (the person or persons whose well-being is being measured). This produces a set of indications. Measurement indications will typically take the form of answers to questions, behavioural observations, the results of medical tests, and data about access to resources or property. Sometimes indications will be quantitative, and sometimes qualitative.

But well-being indications do not alone generate information about well-being. Only through the construction of a model, which invokes background knowledge and assumptions to explain how these indications relate to the well-being of the person in question, will measurement outcomes be generated. Some of the theorising about well-being will be at a very general level—determining what to measure and how to measure it is done in relation to a theoretical framework, which characterises well-being in subjectivist, objectivist or mixed terms, and which identifies particular means for gathering data which will be relevant to well-being claims. Other aspects of modelling

well-being will be more specific. Modelling might involve, for example, consideration of how someone's access to healthcare resources impacts on their well-being; the extent to which their answers to questions about depression contribute to an assessment of their psychological well-being; whether reports about mobility should be taken to provide evidence of well-being, and to what extent they contribute to an overall assessment. Measurement outcomes are under-determined by indications—there are many different well-being measurement outcomes which could be derived from a single set of indications. The generation of measurement outcomes involves theorising, on the part of the people making the measurement, using appropriate theoretical frameworks and evidence, about the nature of well-being, and the relation of the particular indications to it.

Again, sometimes measurement outcomes will be quantitative, and many formal evaluations of well-being, as are used in healthcare, development and policy contexts will endeavour to convert both quantitative and qualitative indications into quantitative measurement outcomes. Some of the tools considered in the previous three chapters were explicitly quantitative—attempting to measure and enumerate preferences, subjective psychological states, or to quantify functionings. But appropriately systematic theories of well-being could also take much simpler and more contextually bounded forms. Just as length can be measured using strides or hand spans, and doing so can make sense in particular contexts without the need for calibration or explicit scientific theorising, well-being could be measured using personal scales—I can ask myself whether I'm doing better than I was last summer, or whether more of my employment related preferences would be satisfied in this job prospect or that one. There may, of course, be questions about the stability of such measures, due to the unreliability of memory for example, and they would be inappropriate for use in policy-making contexts or contexts which require formal accountability. But, insofar as it is possible to make *some* such comparisons reliably

and meaningfully, these more informal measures of well-being are plausibly valid in certain contexts.

This relational account of well-being is well supported by the discussion and examples in the previous three chapters. When well-being is measured, the resulting assessment is shaped by the environment in which the measurement takes place—where this includes the cultural and linguistic context as well as the physical space in which measurement takes place—and by features of the measuring instrument—including the kinds of questions it comprises, the language and terms used, and the specification and definition of the goods that it picks out. Of course facts about the person whose well-being is being measured are crucial—without their responses, or without an application of the tool to them and their lives, there would be no assessment of their well-being. However, the responses that are generated are shaped by the other *relata*. This is true in the obvious sense that no responses would have been given had the measurement exercise not taken place, but also in the sense that the particular responses are dependent on the environment and the tool for their existence.

The idea that measuring well-being involves placing the object of well-being in a parameter space has clear parallels to Sen's argument (1985b) that moral principles impose informational constraints, and that determining how persons and states to world relate to a moral principle requires us to characterise them in informational terms that are relevant to the principle in question (see my discussion in 3.3.ii). The model-based account of the measurement of well-being also makes good sense of the measurement of health state utility via preference elicitation. Measures of health state utility depend on the construction of a complete set of health states over which utility values will range, in order to enable patients to elicit utility values about their health (see 4.2). The set of health states are a set of possible indications—possible states of well-being that the patient could be in—and a good measure of health state utility will provide a coherent explanation of how the set of health states reflect possible states of

persons that are relevant to their health-related quality of life. Indeed, every measure of utility via preference-elicitation depends on a complete and specifiable set of outcomes over which preferences can range, which represents all of the possible utility values a person can have. This set of possible outcomes is the parameter space into which objects of well-being measurement are placed, in order to derive information about their preferences. Moreover, when utility values are assigned to health states, preferences are typically only elicited for a selection of health states, with the rest being imputed based on a derived utility function. Most of the utility values assigned to health states are therefore modelled by the measuring subjects, rather than being directly surveyed. The resulting array of health state utility values represent a theoretical model of the relationships between preferences and well-being, as much as they represent people's actual preferences.

Finally, my account helps to explain the apparently inconsistent assessments of well-being observed in the colonoscopy procedures discussed by Kahneman (see 5.2). Patients who had the probe left in for longer experienced low moment utility (measured by in-procedure pain assessments) and high remembered utility (measured via a retrospective evaluation of how painful the experience was), compared to those for whom the probe was promptly removed. These two utility values only appear to be inconsistent if the same theoretical framework is used to derive measurement outcomes and to make informational claims about well-being. However, if the retrospective values are interpreted in terms of two different theoretical frameworks, the two resulting well-being measurement outcomes will no longer be inconsistent, merely different. The work to be done is in determining which framework is appropriate in the given context, then, and not in determining the true assessment of well-being.

## 6.2 *The role of the measuring subject*

The measuring subject, that is, the person or people who are making the measurement, plays a particularly important role in the measurement of well-being on this account. There are two principal ways in which they construct assessments of well-being: first, they play a role in the selection and often the design of the measuring instrument, and the design of the physical context of measurement. Secondly, their theorising about well-being is what generates measurement outcomes. The first consideration has been covered at some length in the previous three chapters, so I will discuss it only briefly; the second merits a little more extensive discussion. First, then, the measuring subject's selection and design of the measuring instrument is significant for the measurement of well-being insofar as the instrument affects the measurement outcomes. In the measurement of physical attributes, such as distance, time, and mass, the selection of the tool does undoubtedly have an effect on the measurement outcome—using a more precise tool will lead to the differentiation between outcomes that would have been judged equal using a less precise tool, for example. In the case of well-being, the effect is likely to be more pronounced. The difference between instruments used to measure well-being appears to result in not just minor re-orderings of well-being rankings, but significantly different assessments of well-being. Insofar as it makes sense to say that different tools measure the same attribute—well-being—they do so via the measurement of quite distinct constituent parts, and do not obviously represent differently-refined assessments of the same underlying property.

The subject of measurement plays this role in both formal contexts—for example, in those contexts where well-being is measured using validated tools for the purpose of public policy or clinical decision making—and informal contexts. In many instances of formal well-being measurement the measuring subject may not be the designer or author of the measurement tool, but will use an 'off-the-peg' tool developed by someone else. Here, they are nonetheless responsible for selecting an appropriate tool,

and implementing it in an appropriate way. They are also at least partly responsible for determining the physical environment within which measurement takes place—whether it is done over the phone, online, by post, in the presence of a professional, by interview, and so on. In formal contexts where the measuring subject also develops the instrument they have a far greater degree of control over the resulting measurement indications, via their control of the instrument.

In informal contexts, well-being might be assessed as the result of a conversation, or on the basis of casual observation of someone's behaviour or consideration of facts about their situation. Here too the measuring subject has control over the questions they ask, the prompts they provide, the factors that they take into account, and the information they take to be pertinent. They will be constrained, of course, by the limits of linguistic comprehension, as well as cultural limitations—there may be some questions that it is inappropriate to ask, or that the object of well-being will refuse to answer—and practical limitations—they may not have access to the diary, bank account details or medical records of the object of well-being. In formal contexts, creators of instruments will be contained by the limits of linguistic intelligibility, but they will also be constrained by disciplinary conventions and expectations, such as formal validity procedures, as well (Alexandrova and Haybron 2016).

Secondly, the measuring subject plays a fundamental role in constructing measurement outcomes via their role in modelling the interactions between the *relata*, and theorising about how the indications produced by their relation bear on well-being. In relation to more formal measures of well-being, modelling involves statistical analysis, to determine how individual questions relate to the broader construct under consideration, but also judgements about the nature of the underlying phenomenon. Certain theoretical assumptions have to be made about what well-being is, and how various constructs—psychological function, physical function, social-function, global life satisfaction, and so on—relate to well-being in order to get any attempt to measure

well-being off the ground. Making such assumptions is a necessary part of choosing an appropriate range of questions and determining how they should be weighted in relation to one another, for example. Some theorising about well-being takes place prior to any modelling of data—it is needed to identify what is relevant to well-being in the first place.

Crucial to modelling well-being is the placement of the object of measurement in a parameter space, which represents a range of possible well-being states in which it could be. This will involve characterising the object of measurement as the kind of thing that can have different states of well-being, and as the kind of thing whose possible states of well-being are best captured by the possible indications on the measurement apparatus. In the most basic and informal sense, this might mean that when someone says they are doing well, and seems not to be disguising underlying unhappiness, this counts towards a positive assessment of their well-being. And when someone says they are not doing well, and doesn't seem to be joking or exaggerating, this counts towards a negative assessment of their well-being. This characterisation is more complex than it first seems, involving the representation of the object of measurement as an agent, as someone with introspective ability to determine their own states, as a language user who understands the question in more-or-less the way that it was intended, and so on. It also fits in with a broadly subjectivist general theory of well-being—that well-being can be determined, and perhaps is best determined, in a particular context, by asking someone to report on how well they are doing. Modelling well-being by placing an object of well-being in a parameter space thus requires prior theorising about the appropriate conception of well-being for the context under consideration.

The modelling of well-being may differ from the modelling of other measurable attributes—length, time, mass, temperature, and so on—with respect to the way that evidence and theoretical assumptions constrain the ways that the attribute in question

can be represented. I don't view these as differences in kind, but they do result in quite different concepts and measures as outcomes. The measurement of temperature, for example, proceeds on the basis that certain substances expand uniformly with temperature. This must form part of the theoretical grounding of the measurement of temperature using a thermometric fluid, because the very possibility of such an activity depends on this assumption. Similarly, the measurement of length via the concatenation of metal rods presumes that the rod does not expand or contract over time, or when moved through space. Again, the very possibility of this measurement operation depends on such a theoretical assumption. Furthermore, to maintain the assumption that the length of the object in question would be the same if measured by some other means—by measuring the time it takes for light to travel the distance in question, or by measuring the distance using a particular piece of string—further assumptions have to be made about the possibility of calibration across these operations. In order to claim that two different operations measure the same property it is necessary to show that they are consistent under the same theoretical framework. Once such a theory has been established, the kinds of operational rules that yield genuine measurement can be established relative to that theory (van Fraassen 2008, pp.165–6). This is necessary because the concept of length is a fairly tightly constrained one, which will not tolerate the prospect of much inconsistency between measurement operations.

In some ways, the theoretical constraints on the measurement of well-being look similar. For example, the measurement of subjective well-being depends on the assumption that self-reports of affect have a consistent referent, and are commensurable. And the measurement of objective well-being similar depends on assumptions about the stability of measures and the consistency of the good under consideration over time and between persons. In another sense, well-being looks quite different. If we are variantists about well-being, we take the concept of well-being to



be inconsistent across contexts, and this implies that different theoretical assumptions about well-being will apply. Variantism about well-being is motivated by widespread contextual variation in the use of the concept of well-being. This suggests that any overarching theoretical constraint on well-being operations will tolerate inconsistency across contexts. As discussed in chapter 1, the background picture of well-being seems to contain a patchwork of inconsistent intuitions about well-being and diverse applications, which pull in different directions. This background picture essentially reflects a fairly broad concept, which is used in multiple ordinary language- and specialist language contexts. This doesn't mean that there are no limits on the possible meanings of well-being, but it does mean that it is not such a tightly constrained concept as, for example, length or mass. That is, the theoretical constraints on the possible models of well-being, the possible parameter spaces into which the objects of well-being measurement can be placed, are looser, or at very least much more numerous, than the possible parameter spaces into which objects of length or mass measurement can be placed. If this is right, then the role of the measuring subject in the measurement of well-being has the potential to be *more* significant with respect to the measurement outcomes than the role of the measuring subject in the measurement of physical attributes, simply because the constraints on the possible conception of well-being, and thus on the possible placement of objects of well-being measurement in a parameter space, are less restrictive.

This isn't to say, though, that anything goes vis-à-vis the measurement of well-being—that the measuring subject is free from any theoretical constraint in their modelling of well-being. I will discuss this further in 6.5, but a few brief remarks will help to set the stage. A scientist measuring interplanetary distances is constrained by a whole host of theoretical assumptions, disciplinary conventions, gold standard measuring instruments, orthodox mathematical expressions and languages, and evidence from previous studies. These conventions constitute the disciplinary context within which

she is embedded, and they limit the kinds of choices that she can make when calculating the distances in question. In part, these conventions exist because the purpose of certain scientific activities, and the acquisition of scientific knowledge, demands consistency across its constituent projects. Such consistency is necessary for the activity in question to be understood as part of scientific practice, either of a general or specific kind. Peter Winch (1990, pp.85–6) argues that the scientist is influenced both by the phenomena that she observes, and by the scientific community: it is on the basis of the former than she develops her concepts and conclusions as she does, but it is on the basis of the latter that she is able to do this at all. Her observation and measurement occur within an established form of activity with her fellow scientists, and counts as participation in that community insofar as it is *intelligible* to them. This, of course, doesn't mean that her findings cannot contradict existing evidence, nor even that she cannot argue in favour of fairly substantial theoretical alterations. Instead it means that in order for her experiments and arguments to form a part of her scientific practice, they must be understood as contributions to scientific practice, knowledge and understanding by her fellow practitioners.

Winch's account of intelligibility provides a helpful framework for understanding what goes on when someone endeavours to measure well-being. The measuring subject is constrained by his socio-linguistic context. If he is endeavouring to construct a psychometrically valid measure of subjective well-being, which can be used as a formal measure of well-being in a healthcare context, for example, the contextual constraints on his activity will be explicit, fairly strict and numerous. There are psychometric conventions about the processes by which surveys are constructed and tested which limit the possible choices he can make. He must abide by these if his findings are to be intelligible within the psychometric community, and if the instrument is to serve its intended function as a measure of well-being. However, even when some theoretical conventions are explicit, as in conventions of scientific,

technical and academic practice, implicit theoretical constraints will also be imposed by broader social and linguistic practice. In informal contexts, too, the person assessing well-being is constrained by his socio-linguistic context. The conventions about the measurement of well-being in ordinary language contexts are in some sense less demanding than the conventions in formal psychometric contexts. There may be no formally established rules for evidence gathering, statistical modelling and validation in everyday assessments of well-being but, nonetheless, the possible choices of measuring subjects are constrained. If someone consistently judges people to be doing badly when they report that they are doing well, and vice versa, we might think there to be reason to question the viability of the theory of well-being that he is working with. On the other hand, he might justify his assessments of well-being, arguing, for example, that the objects of well-being in question are afflicted by a systematic failure to recognise their own subjective psychological states, because they are anorexic or psychotic. This may make his well-being assessments intelligible to us, despite initially appearing counter-intuitive.

My relational account of well-being differentiates between the measuring subject and the object of well-being. Sometimes, however, the two are one and the same, that is, in those cases where someone introspectively assesses their own well-being. Does this make a difference to my account, or present a problem for it? Like Scanlon (1998, pp.109–10), I think that well-being plays a surprisingly small role in the thinking of individuals who are deciding how to act or what to choose for themselves, and that well-being plays greater role in judgements by a third person. However, clearly it is possible for individuals to make assessments of their own well-being. There is reason to think that introspective assessments of well-being should not be treated differently from third person assessments, that is, that introspective assessments of well-being involve the same kind of theoretical modelling as third person assessments. My discussion in chapter 5 suggested that there is not good reason to think that direct

reports of subjective experience should be understood to be unmediated and unframed in a way that other ways of eliciting well-being reports are not. Reports of subjective experience which depend on introspection are also framed by the questions which prompt them and the context in which they are made. Introspection need not, then, be thought to be particularly special or distinct from other means of eliciting well-being indications.

In order to avoid thinking of introspection as involving two distinct agents—an object of measurement and a measuring subject—the relation involved in introspective assessment of well-being can be conceived as a relation between an object of well-being (oneself); a measuring tool (introspective appraisal); and a measuring environment which, in this case, does not need to be taken to include a separate measuring subject. When someone introspectively evaluates their well-being, their assessment will involve a number of different components: they will think about what it would be for them to do well or badly, identify characteristics of their life or psychological state which are relevant to such an assessment, determine whether these are reliable evidence, and, if so, determine whether they are contributors to or detractors from well-being, and to what extent. Depending on the context, the content of these considerations will be different. This is pretty much identical to what goes on in third person assessments of well-being, and this is because a pluralist concept of well-being requires consideration of such factors in order to give it content, regardless of whether well-being is assessed from the first-person perspective or the third. It is not the case that, from a first-person perspective we have a different, direct perception of well-being that is not mediated by context. Even when no particular question or external prompt is present to frame an introspective assessment of well-being, socio-linguistic conventions will constrain a person's introspective assessment of their well-being. It is these conventions which shape the way in which they conceive of their well-being, and which give rise to a pluralist concept of well-being.

### 6.3 *Stance independent pluralism I: McDowell*

The next three sections consider the prospects of two alternative accounts of well-being. In this and the following section, I consider two ways of justifying a stance independent pluralist view. In this section, I consider an argument for moral objectivity from John McDowell which suggests grounds for maintaining stance independence about well-being. In the following section, I discuss Richard Boyd's naturalist moral realism, which offers a different explanation, and which seems to avoid some of the problems of McDowell's approach. Finally, in 6.5, I consider Alexandrova's view that objective features of the measuring subject should be taken to determine the content and standard of assessment of well-being claims, which takes my account to be too promiscuous about the possible theories and concepts of well-being, and to lack discriminatory and critical capacity.

My account characterises well-being as highly dependent on the attitudes of the persons measuring it, but not as independent of the objects of measurement. Capturing well-being in the language of measurement makes sense of the apparent attitude dependence of well-being ascriptions, whilst not losing sight of the centrality of the object of well-being claims and their features as central to any theorising about well-being. Despite the role of the measuring subject in constructing well-being ascriptions, and because of its focus on the identification and measurement of well-being in individuals, this account places greater emphasis on the object of well-being than accounts which start with general definitions of well-being, and only think about its identification and measurement in people as an afterthought, if at all. Nonetheless one of the likely criticisms of my account is that it shifts the assessment of well-being away from the object of well-being itself and his features, and onto the subject of well-being and her features, and that this compromises the concept. When we talk about well-being, we are talking about the object of well-being, not the subject, so to make

assessments of well-being reflect the latter rather than the former gets something seriously wrong.

In chapter 1, I argued that a commitment to well-being variantism leaves two substantive options on the table: stance independent pluralism and stance dependent pluralism. I characterised stance independent pluralism as the view that what constitutes well-being is different in different contexts, and where the determinants of well-being in any context are given by the objective, attitude-independent features of that context. These will typically be facts about the person whose well-being is in question, including facts about their attitudes and psychological states, and facts about their context, but will exclude the attitudes, values and purposes of the person who is assessing their well-being. Stance independent pluralism thus takes there to be a fact of the matter about the appropriate theory of well-being for any given context that is independent of attempts to define or measure it. The stance independent pluralist looks likely to be a realist about the measurement of well-being—there is a fact of the matter about someone’s well-being before anyone sets about measuring it. Stance dependent pluralism is the view that what constitutes well-being is different in different contexts, where the determinants of well-being are shaped not only by facts about the person whose well-being is in question, and their context, but also by the attitudes, values and purposes of the person who is defining or measuring well-being. The stance dependent pluralist will think that assessments of well-being are in some non-trivial way shaped by and dependent upon their theoretical characterisation.

My conception of well-being as constructed in the context of measurement is, clearly, a stance dependent account. It takes any attribute of well-being to be critically dependent on the attitudes of some agent or group of agents. I think that my arguments about the measurement of well-being via the measurement of measurement of objective goods, preferences, and experienced utility speak, at first glance, in favour of a stance dependent pluralism about well-being. But someone with

an inclination towards the stance independence of well-being is likely to resist my interpretation. The stance independent pluralist about well-being will likely argue that different characterisations of well-being, even at the very specific level, are determined by objective features of the case. That is, the appropriate theory and characterisation of well-being for a given context is an objective matter. There is, for example, a correct way to design a survey in order to elicit those preferences that characterise well-being in a group of patients with chronic kidney disease. There's correct way to design a life-satisfaction survey to determine the well-being of participants in a basic income randomised trial. And there's a correct way to characterise the capabilities of Eritrean citizens in order to determine their well-being. The causal chain, according to this argument, is reversed: it is not that the conceptualisation of well-being by the people measuring it determines, or partly determines, the resulting attributions of well-being. Instead, objective features of the object of well-being assessments and their context determine both the objectively correct characterisation of well-being, and—if they come to have knowledge of this characterisation—the attitudes of those people measuring well-being. People's shared attitudes towards the correct characterisation of well-being are caused by the objective features of the case; they are not the cause of particular characterisations of well-being. This kind of view might, I think, be held by certain perfectionist and objective list theorists, who are willing to accept the need for a pluralist concept of well-being that is led by the variation exhibited in practice, but not willing to accept the constructive role of the measuring subject proposed by my account.

As it stands, there is an epistemic gap in the stance independent pluralist's explanation. If objective context-specific statements about well-being are determined by features of the case, it remains unexplained how the people measuring well-being come to acquire the correct attitudes about well-being, and its appropriate characterisation. It is all very well to insist that there is a correct theory, construct and measure of well-being for

every context, but in absence of some method for coming to know or approximate the right theory, the value of such a claim is fairly limited. The stance independent pluralist seems committed to the idea that ascertaining the correct theory and measure of well-being for a particular case requires a careful consideration of the features of the person or group whose well-being is being assessed. Some stance independent pluralists might think that the same feature always points towards the same conception of well-being, across different cases, others might allow that the same feature counts in favour of different conceptions of well-being across cases. Either way, though, he needs to offer an explanation of how it is known, on consideration of the features of a case, that they lead to a particular conception of well-being. Without some explanation, there doesn't seem to be reason to believe that the specific characterisation of well-being is the correct one. In chapter 3 I discussed claims of self evidence in relation to objectivist accounts of well-being. I argued that it is difficult to defend claims of normative self-evidence without falling back on an intersubjective justification. I will not rehash the argument, but similar considerations apply here, and it would be imprudent for stance independent pluralists to insist that adequate consideration of the facts of the case are sufficient for self-evident appraisal of the correct characterisation of well-being. Without a general explanation of the relationship between features of cases and the derivation of correct theories and assessments of well-being, there may not even be reason to think that there is ever a 'correct' derivation at all. And clearly appeal to intersubjective agreement is not open to the stance independent pluralist as a means of closing the epistemic gap—if the correct characterisation of well-being depends on the attitudes of a set of people, it will not be stance independent in the relevant sense.

In his critical notice of Bernard Williams' *Ethics and the Limits of Philosophy*, McDowell argues that the objectivity of normative beliefs may be the best explanation for convergence on them (McDowell 1986). So intersubjective agreement may be an indication of the objectivity of claims about what is good for people in a



given context. McDowell suggests that the explanatory power of claims of normative truth depends on the fact that the reasons for holding them are good ones. This, he says, warrants the idea that 'the belief is held because the fact that things are as it represents them impresses itself on the converging believers' (McDowell 1986, p.379). McDowell suggests that although ethical explanations don't amount to proofs, we can nonetheless make a parallel between ethical explanations and mathematical explanations, such that in both cases the explanatory power of an explanation of belief is grounded in the endorsement of the reasons for which the belief is held. Of course—and McDowell recognises this—if the reasons for which the belief is held are only endorsed at a local level, then explanations thereof will only seem to be good ones from within a given ethical outlook. However, insofar as there is no better explanation of ethical convergence, a kind of objectivity in ethics can nonetheless be grounded, insofar as it constitutes the 'best explanation.'

The stance independent pluralist about well-being might well use a similar argument to explain convergence on conceptions of well-being. While well-being usually plays a role, often a central role, in theories of value, claims of well-being aren't typically taken to have independent normative force. An additional judgement about the normative value of well-being in a particular case is needed to move from a claim about someone's well-being to a claim about how one ought to act, or a claim about right and wrong. It might, then, be thought that well-being claims don't need the same kind of explanation that normative claims need, to ground their objectivity. If well-being claims are merely descriptive claims, then their objectivity is grounded in the objectivity of the facts about persons and their well-being that they describe. But well-being claims are somewhere between strictly factive and obviously normative claims: they say something about what is the case, rather than what ought to be the case, but they also make claims in relation to the good—what is good for someone, what it is for them to do well or badly. Any attribution of well-being to a person is grounded in a claim about

how we *ought* to understand well-being. Whether or not this should be understood as a moral ought, the facts in question seem to overdetermine well-being claims. Given this, some explanation is needed for why we are justified in defining well-being as we do—this is the explanatory gap which the stance independent pluralist needs to bridge. This indicates that something like McDowell’s argument should be taken to be necessary for explaining the objectivity of well-being claims, as well as quintessential normative claims. A commitment to stance independent pluralism entails an explanation such that the appropriate characterisation of well-being is objective.

McDowell’s argument hangs on the idea that the objectivity of ethical claims is the best explanation for their convergence. He doesn’t present an argument for this, nor is it clear why this should be thought to be the case, nor how it could be known to be the case. Intersubjectivists about ethical claims can also explain convergence (insofar as it does in fact exist), likely on the grounds that having shared norms and values is central to successful collective living. Why think this is a worse explanation than the purported objectivity of ethical claims? One reason is that ethical principles appear to us to be objective. I take this to be the idea, expressed by McDowell, that ‘the best explanation of the convergence would involve the ideas that things are as the belief represents them’ (McDowell 1986, p.378). However, while it is plausible to say that some ethical beliefs represent moral obligations as really having normative force, that ethical principles really do express rights and wrongs, and really ought to guide our practice, it’s less clear this amounts to an appearance of objectivity. For our normative seemings to imply objectivity, it would have to be the case that, were our normative claims grounded in intersubjectivity rather than objectivity, we would not perceive them to have such normative force, or to express rights and wrongs and oughts in the same way. However it’s not at all clear that this is the case, and it seems difficult to show it without a question-begging assumption of normative objectivity. I am reminded of Elizabeth Anscombe’s anecdote about Wittgenstein:

[Wittgenstein] once greeted me with the question: ‘Why do people say that it was natural to think that the sun went round the earth rather than that the earth turned on its axis?’ I replied: ‘I suppose, because it looked as if the sun went round the earth.’ ‘Well,’ he asked, ‘what would it have looked like if it had *looked* as if the earth turned on its axis?’ This question brought it out that I had hitherto given no relevant meaning to ‘it looks as if’ in ‘it looks as if the sun goes round the earth.’ (Anscombe 1959, p.151)

Similarly, ethical claims appear to us to be objective if we view our beliefs about them through an objectivist framework. We experience real moral obligations and prohibitions, but whether we perceive them to be objective or not depends on our theoretical interpretative framework. If sensible verification will not verify or falsify a claim, we cannot appeal to how things appear to us to determine its truth or meaning.

Without a plausible ‘best explanation’ argument, concerns about epistemic access to objective normative claims again rear their head. McDowell is right to suggest that the dichotomy between science, as a discipline in which we can make our inquiry from outside our historical and social standpoint, and ethics, as a discipline in which we cannot, is false. But ethical objectivity is not the only, or the best, response to the breakdown of a science/ethics distinction. The possibility remains of getting on with our various endeavours, scientific and ethical, without objectivity, of developing and holding coherent, purposive, justified and justifiable beliefs which help us to make sense of the world and to live in it, without them needing to represent or correspond with something which has an existence separate from our inquiry. It looks as though the stance independent pluralist about well-being will have considerable difficulty offering an adequate explanation of how we have knowledge of the objective truth about well-being without reliance on an intersubjective grounding.

#### 6.4 *Stance independent pluralism II: Cornell realism*

But perhaps the stance independent pluralist need not resort to arguments from best explanation. A group of philosophers known as the *Cornell realists* argue that moral terms correspond to natural properties, and we can thus come to have knowledge of moral facts via observation, just as we can come to have knowledge of the physical world via observation of physical properties (Sturgeon 1985; Brink 1986; Railton 1986; Boyd 1988). This approach avoids relying on intuition and self-evidence to explain our moral knowledge, and so seems to identify a way of bridging the epistemic gap. It seeks to provide a firm basis for the claim that our normative beliefs are caused by and responsive to a normative reality, which does not invoke any special faculties.

Richard Boyd (1988) provides the most detailed explanation of the naturalism of moral properties and moral terms. His account is premised on a nuanced account of scientific realism, which asserts that many scientific concepts are *homeostatic cluster terms*. This means that they refer to natural kinds ‘whose natural definitions involve a kind of property cluster *together with* an associated indeterminacy in extension’ (1988, p.196).<sup>50</sup> So the definition of a homeostatic cluster term is provided by a cluster of different properties, where possessing some indefinite number of these properties is sufficient for falling within the extension of the term. He takes the concepts ‘healthy’ and ‘biological species’ to be examples of homeostatic cluster terms, characterised by a number of imperfectly shared and homeostatically related features. The features are homeostatically related insofar as the presence of some of the properties tends to favour the presence of the other properties, or there are underlying mechanisms which tend to preserve the co-occurrence of several properties. To try to resolve the vagueness in these natural kinds, by offering a strict set of necessary and sufficient conditions for healthiness or for biological species, would be to specify an *unnatural*

---

<sup>50</sup> Boyd’s emphasis.

kind, because the terms capture kinds that are extensionally vague in nature. Determining which properties are part of the cluster is an a posteriori task, and will be a contingent matter, depending on the actual operation of homeostatic mechanisms, which may be different in different contexts.

Boyd argues that moral terms are also homeostatic cluster terms. He prioritises developing a naturalist reading of ‘good,’ but his account can easily be extended to a naturalist reading of ‘well-being.’<sup>51</sup> Moral goodness, then, is a homeostatic property cluster. It is characterised in terms of a cluster of goods, and for something to be correctly characterised as ‘good,’ it must possess some indeterminate number of these goods. For Boyd, the goods are things that satisfy human needs—some physical, others social and psychological (1988, p.203). The goods are homeostatically clustered, such that the satisfaction of the goods for one individual tends to be conducive to their satisfaction for others. The homeostasis of moral goodness is thus social, rather than individual. On this account, we come to know the truth and falsity of moral claims via observation, in exactly the same way as we come to know the truth and falsity of complex scientific claims via observation of a series of indicative properties. Just as, in order to determine whether someone is healthy or not, we identify the properties that they have that are indicative of good and ill-health, to determine whether something is morally good or not, we identify the properties that it has that are indicative of goodness.

Central to Boyd’s scientific realism is the claim that scientific methodology is theory dependent. Changes to measurement operations for theoretical magnitudes—of unobservables, for example—are often made on the basis of theoretical developments

---

<sup>51</sup> Indeed, there is little distance between Boyd’s definition of moral goodness and many perfectionist definitions of well-being. At points he appears to take well-being and moral goodness to be synonymous (1988, p.204, p.211).

(1988, p.188). This methodology is justified by the realist on the assumption the measurement procedures are reliable, and that the theoretical developments reflect increasingly accurate knowledge of theoretical entities. This process of improvement is understood as a feedback loop: the reliability of measurement operations explains the success of our observations and experiments in discovering new theoretical knowledge, which may in turn produce further improvements in experimental techniques (1988, p.189). Realist scientific methodology is, then, theory dependent, and it is objective because it is theory dependent. That is, it is only reliable because and insofar as the background theories upon which it rests are approximately true. Our observations of the properties which constitute moral goodness and other normative entities are theory dependent, in the same way. That is, moral observations are theory dependent in the sense that our moral observations are reliable because our moral theories are approximately true.

Boyd's account is very credible. It provides a subtle and plausible grounding for a stance independent pluralist conception of well-being, which avoids the epistemic problems encountered by other approaches by committing to a robust naturalism about moral properties. The most controversial aspect of his account, with respect to both scientific realism and moral realism, is the claim that the theory dependence of observation does not undermine its reliability because and to the extent that our theories are approximately true. I focus on this claim with respect to moral realism only. Boyd's account obviously demands some demonstration that our moral theories are indeed approximately true, but his argument for this is very quick. Assuming a homeostatic conception of morals, such a demonstration would involve showing that 'our background beliefs about human goods and the psychological and social mechanisms which unite them have been good enough to guide the gradual process of expansion of moral knowledge envisioned in that conception' (1988, p.207). That is, it would involve showing that we now have greater knowledge of our collective needs

and how to meet them than we did earlier in our history. Boyd claims that it is pretty clear that we have improved our understanding of our needs and our individual and social capacities. Our current moral theories better comprehend our needs than our previous ones, and this improvement shows that they must have been approximately true to begin with.

I think, in part, the credibility of Boyd's account comes from his very broad conception of human needs—his 'fundamental' human goods, and the corresponding needs that they satisfy, are not just physical, but complex social and political goods. Boyd's account also depends on the attitude independence of the goods that characterise moral properties—this is central to its realist status. This leads him to say some peculiar things about 'fundamental' human goods. I quote him at length:

Knowledge of fundamental human goods and their homeostasis represents basic knowledge about human psychological and social potential. Much of this knowledge is genuinely *experimental* knowledge and the relevant experiments are ('naturally' occurring) political and social experiments whose occurrence and whose interpretation depends both on 'external' factors and upon the current state of our moral understanding. Thus, for example, we would not have been able to explore the dimensions of our needs for artistic expression and appreciation had not social and technological developments made possible cultures in which, for some classes at least, there was the leisure to produce and consume art. We would not have understood the role of political democracy in the homeostasis of the good had the conditions not arisen in which the first limited democracies developed. (Boyd 1988, p.205)

The claim here is, I think, that experiments in living reveal to us human goods, which satisfy important human needs. Some of these, such as the need for friendship or the need to engage in cooperative efforts, will likely emerge early on, and relatively little

experimentation is required to determine their role in the satisfaction of human needs. More complex needs, such as the need for a fungible and non-perishable medium of exchange, the need for a free press in a political democracy, or the need for a 4G wireless network in a digital society, develop later, following more experimentation, and only in the context of contingent technological, social and political developments. We know that these truly are human goods, which satisfy human needs and constitute the good, because when we treat them as properties of moral goodness in our moral theories we are able to acquire deeper insight into our human needs and potential.

This move troubles me. Boyd concedes that his view is premised on ‘optimistic empirical claims’ (1988, p.217). One of these is, presumably, the claim that our moral observations are generally reliable, that is, that they reflect an attitude independent moral truth. What Boyd views as optimistic, I regard as unsubstantiated. In the passage quoted above, Boyd stipulates that the experiments which enable knowledge of fundamental human goods should be ‘naturally’ occurring, and that their occurrence and interpretation will depend ‘both on ‘external’ factors and upon the current state of our moral understanding.’ He accepts, then, that moral development will be grounded in current moral attitudes and theories. This attitude dependence is only consistent with moral realism if our moral theories are approximately true. But the argument that he provides in support of this claim does not necessarily support moral realism. There seems to be no way of knowing whether our beliefs that certain goods are fundamental human goods truly do allow us insight into our nature, or whether the goods that we contingently acquire and our beliefs about them change our nature and needs, and partially fix, or at least direct, their future development. On Boyd’s story, while the contingencies that enable democracy and art to develop may be partly accidental, it is no accident that they are fundamental human goods. It has always been the case that political democracy will be part of the good, given certain conditions. It is our understanding and knowledge of our needs and the goods that fulfil them over



time that changes. We know that our theories are truth tracking because our understanding of our needs improves over time. On my alternative reading, it is not just our understanding of our needs and the goods that fulfil them that changes over time, but those needs and goods themselves. The nature of humans and the nature of their needs changes in the context of contingencies, so the human goods that satisfy our needs, and our needs themselves, could have been otherwise. Our understanding of our needs doesn't necessarily improve over time, and it would be difficult to tell whether it does improve: if our needs are not static, the possibility for comparison may be limited. If our moral theories are responsive to our changing needs, rather than responsive to our understanding of our needs, then Boyd's conception of moral truth no longer seems so applicable. Insofar as there is moral truth which is tracked by our moral theories, it is highly constrained—temporally, spatially, culturally, and by specific social and political developments.

My reading might be not be entirely unacceptable to Boyd, so long as the needs and goods in question remain attitude independent—that is, he may allow that our needs could have been otherwise, and could be otherwise in the future, and even that different people can have different needs in different contexts, as long as it is not the case that their variability is dependent on our attitudes. However, it is not clear to me that the distinction between attitude independent contingencies and attitude dependent contingencies can be so clearly drawn. Boyd takes moral development to require experimentation in collective living. If deliberate structuring of human societies through social and political development can change our needs, affect which goods will fulfil them, and so lead to the revision of moral theories about our fundamental human goods, then there is a clear sense in which our goods, needs and moral theories are dependent on the intentions and actions of people. But, moreover, if experimentation in collective living is done with an aim of improving understanding

and fulfilment of our needs, and so driven by some conception of what our needs are, then clearly attitudes about the good will be determinative of good.

Boyd later recognises that the extent to which he allows people's needs to be dependent on their social circumstances opens the door to a kind of relativism (2003). He argues that if the nature of flourishing is different for people who grow up in different social conditions, we may think of the term 'good' as partly denoting each of the different versions of human flourishing (Boyd 2003, p.511). While he still regards each of the different kinds of human flourishing to be homeostatic cluster properties, this marks a very serious concession on Boyd's part. So serious, indeed, that it is difficult to see how this remains a realist view. If human needs are path dependent, and so in part determined by earlier moral theorising, as well as other contingencies, then the realist claim that moral progress involves discovering facts about human need seems unwarranted. James Lenman (2013) suggests that the theory-ladenness of moral progress prevents it from playing the role that Boyd requires it to in confirming our moral theories. Our notion of what constitutes moral progress is driven by our actual moral theories, and so the confirmation of our theories as true via the observation of progress is almost inevitable. Lenman (2013, p.272) argues that '[m]ore or less any evolution in our moral outlook is apt to look, *ex post*, like progress.' This strikes me as slightly exaggerated, but it certainly looks as though our current moral theories are liable to characterise our current moral practice as progressive in relation to previous moral practice, insofar as moral practice keeps step with moral theories.

One alternative to Boyd's account, which is consistent with some of more compelling aspects of his naturalism, would be to think of ethics, as do John Dewey and William James, as responding to the most urgent problems of the age, rather than trying to determine the moral truth once and for all. For Dewey, moral and political philosophy involves analysing what is going on in the world, and how it is going on, and trying to determine the consequences of human action (1927, pp.20–1). He argues (1929,

p.102) that philosophy should adopt an experimental method, which ‘is interested in the mechanism of occurrences instead of in final causes.’ Philosophy takes place within a social and historical context: ‘Experimental knowledge is a mode of doing, and like all doing takes place at a time, in a place, and under specifiable conditions in connection with a definite problem’ (1929, p.102). For James too the moral philosopher must respond to the world as he finds it: ‘The subject-matter of his study is the ideals he finds existing in the world; the purpose which guides him is this ideal of his own, of getting them into a certain form’ (1891, p.331).

One concern with an account of moral philosophy as a means of problem solving, rather than as a means of seeking and attaining knowledge of the immutable, is that it precludes the possibility of moral progress. On Boyd’s account, observation of moral progress is the key to vindicating our moral theories—it is because we can observe moral progress that we know them to be approximately true. It is, for Boyd, the fixity of the fundamental human needs which enables us to deepen our knowledge of the goods which fulfil them, and to improve our understanding of the good. Kitcher, drawing on the American pragmatist tradition, seeks to show how a pragmatist moral philosophy can hold on to a notion of moral progress (2012; 2015). Kitcher (2015, p.478) conceives of a pragmatic progress as a form of *originary progress*—progress *from*, rather than progress *to*. For the pragmatist, then, progress consists in overcoming problems in the current state. This, of course, is no straightforward matter. Solutions to current problems may themselves breed new problems, or may increase the likelihood of future problems. Identifying and characterising problems and solutions can only proceed from a particular standpoint, which means that solutions are only ever ‘ends-in-view,’ which may need to be reconsidered further down the line (Kitcher 2015, p.480).<sup>52</sup> Pragmatic progress is not teleological, and its

---

<sup>52</sup> ‘Ends-in-view’ is Dewey’s expression (1925, pp.102–3).

goals cannot be defined in abstraction from a particular point in social history, nor perhaps in abstraction from a particular perspective.

A similar account of moral progress has recently been defended by Allen Buchanan and Russell Powell (2018). Buchanan and Powell argue that defining moral progress in terms of determinate fixed content is epistemically irresponsible. Instead, we should recognise that humans are fallible with respect to the nature of morality, and there is no reason to think that the sources of our past errors about moral demands have been eliminated. This leads them to defend a 'pluralist, provisional' account of moral progress. Moral progress consists in better compliance with valid moral norms, of which there are a plurality. However, we should regard our current beliefs about the validity of moral norms and our current conceptions of moral progress and moral reasoning as 'only provisional, subject to revision over time' (Buchanan and Powell 2018, p.93). For Buchanan and Powell we must treat not only the content of our moral theories as provisional, but also our concept of moral progress. Since we might come to understand morality in radically different ways, our understanding of moral progress might also shift. If the pragmatist can provide an account of moral progress, then it looks like Boyd's argument that observation of moral progress indicates the approximate truth of our moral theories misses the mark. Observation of moral progress might instead indicate that our moral theories have enabled us to solve a series of problems, with respect to particular ends-in-view.

So where does this leave the prospects of stance independence about well-being? Despite its attractiveness, the Cornell realist approach can't secure the attitude independence of well-being. Boyd does, to an extent, recognise the complexity of the relationship between our social and political structures, our attitudes and purposes, and our needs and theories about moral concepts. But it is difficult to both acknowledge the complexity and reciprocity of this relationship and also maintain that the truth of moral claims is independent of our attitudes towards them. Boyd's

argument for moral realism—as opposed to his argument for moral naturalism—relies on a claim about the truth of our moral attitudes and theories which is not well justified by the brief argument he provides. An alternative pragmatist reading points towards a stance dependent account of moral claims and concepts. And given the possibility of an alternative explanation, something like McDowell’s ‘best explanation’ argument is required to explain the primacy of the realist picture. This undermines much of the advantage of the Cornell realist approach, which sought to avoid the epistemic gap faced by other moral realist views. So the outlook for stance independence about well-being is little changed—stance dependence remains a live option, and, moreover, an option with a plausible explanation for the apparent role of the subject of measurement in the generation of assessments of well-being.

I don’t think the arguments in this and the previous section are, strictly speaking, decisive. Plenty of objectivists about well-being, and about normativity more generally, will be willing to bite the bullet. Even if it is difficult to show how we come to know objective truths about well-being without supposing their existence, this doesn’t entail that there are no such objective truths. However, the presence of the epistemic gap, and the recourse to concepts like self-evidence and claims about the objectivity of normative claims or the truth of moral theories in order to explain it, puts the stance independent pluralist on the back foot.

### *6.5 The extent and limits of well-being*

Another concern, likely the major concern, with stance dependent pluralism about well-being is that it is too permissive about what counts as a measure of well-being. While well-being measurement outcomes *are* constrained by facts about the object of measurement and their environment—so a measuring subject can’t just pick an evaluation of well-being theory out of their head—the only constraint on the possible conception of well-being is that imposed by the boundaries of the meaning of the concept of well-being, in natural language or in technical language, depending on the

context of measurement. In this section I will argue that although my view is undoubtedly permissive about the possible conception of well-being, this is not in itself a bad thing. Its relatively laissez-faire approach does not entail that no ways of measuring well-being are better or worse, more or less appropriate, more or less intelligible, than others. It is not, therefore, indiscriminately permissive. Furthermore, that the boundaries of the concept of well-being are not always clear, and not immutable, does not mean that there are no limits to its possible invocation.

The concern can be understood thus: in the relational view set out in 6.2, the measuring subject plays a significant role in shaping both well-being indications and well-being measurement outcomes. Given the theoretical pluralism permitted by well-being variantism, there isn't a clear restriction on the content of well-being claims. Well-being variantism opens up the possible context of well-being claims significantly by accepting that the scope of the definition of well-being is determined by practice, but my stance dependent pluralism seems go much further. This is chiefly because it weakens the link between facts about the object of well-being and the measurement thereof, by taking those facts to be shaped by the measuring subject, through their theorising about well-being. It thus seems to preclude scepticism about the well-being status of idiosyncratic and objectionable well-being claims.

Alexandrova, in defence of her contextualist account of well-being, suggests that accounts of well-being that allow the semantic content of well-being claims to depend entirely on the psychological state of the speaker of well-being claims—that is, the person who makes assessments about well-being—have ugly consequences (2013).<sup>53</sup>

---

<sup>53</sup> Alexandrova calls the person whose well-being is being measured the *subject of well-being* claims, and the person who assesses their well-being the *speaker*. I, using the language of contemporary measurement theory, call the person whose well-being is being measured the *object of measurement*,

She argues that the relevant context for determining the semantic content of well-being claims is the speaker's context, rather than the context of the person whose well-being is being measured. A claim that someone is doing well means different things in the mouths of different people, because different things count towards their well-being in different contexts, different contexts demand different standards of justification, and the threshold for doing well is different in different contexts. The person whose well-being is being measured is the same across different contexts, in which their well-being may be differently assessed—it is not changes in them across contexts, changes to their properties and attributes, that account for changes in their well-being when they move from context to context. Instead it is the context of the person who is assessing well-being, and in particular the evidential and justificatory standards and well-being-threshold relevant to it, that determine the semantic content of well-being assertions.

Alexandrova argues that semantic content of well-being claims is fixed by objective features of a speaker's practical environment, rather than on the speaker's psychological state. She argues that the meaning of well-being claims should not depend on the salience to the speaker of the possibility of certain errors in their ascription of well-being, but on 'features which make it the case that a certain notion or standard of well-being applies, whether or not they fall within anyone's field of awareness' (2013, pp.320–1). While the speaker's attitudes, purposes and values comprise some of these 'objective features,' there are other contributory factors as well. The implication of this is that the speaker's judgement about well-being claims is not the only, or the definitive, authority on the matter. Alexandrova takes this position in

---

and the person who assesses their well-being the *measuring subject*. This is potentially confusing so in this section I avoid talking about subjects at all. I do refer to the *speaker* of well-being claims in relation to Alexandrova's view.

order to avoid the ‘unwelcome consequence’ of allowing speakers’ idiosyncratic conceptions of well-being to determine the semantic content of well-being assertions (2013, pp.321–2). This might be the consumerist who insists that wealth is the only contributor to well-being; the pessimist who thinks the threshold for adequate well-being to be very low; or the eccentric who thinks that well-being should be measured in terms of hirsuteness. She suggests that if well-being claims depend on this radically subjectivist standard, and the possibility of a certain error about a given assessment of well-being is not salient to a person, she will be entitled to treat the assessment as knowledge.

Alexandrova’s focus is slightly different from mine—she is chiefly concerned here with the truth conditions of well-being assertions—but her criticism is nonetheless relevant to my position. While the view that I have presented does not take the standard of assessment of well-being claims to be dependent solely on the speaker’s psychological state, I do award more authority to the speaker in determining the conception of well-being in any given case than Alexandrova seems willing to permit. Alexandrova does take the values, attitudes and purposes of the measuring subject to be part of their objective practical environment, but argues that the agent’s actual judgment about what standard of well-being is locally appropriate should not be the final authority (2013, p.321). She gives an example of a person who has ‘developed a morbid fascination with the Gulag camps and as a result adopted a very dark view of human life according to which well-being consists in minimal survival’ (2013, p.321) This person judges her friend Julia’s impending depression to be ‘a symptom of Western effeteness,’ and so not relevant to her (Julia’s) well-being. Alexandrova is clear that this is not an acceptable theory of well-being, and our account of well-being should draw the line so as to exclude such theories from grounding knowledge claims about well-being. I, on the other hand, think that such theories are intelligible as theories of well-being and shouldn’t be ruled out as candidates for grounding knowledge claims about



well-being. This by no means implies that they are the best theory for understanding well-being in a particular context, nor even that they are particularly useful at all.

I don't think it to be particularly problematic that idiosyncratic and objectionable theories of well-being can ground well-being assertions, for two main reasons. First, the fact that something can be understood as an assertion about well-being does not necessarily award it any particular significance. Secondly, allowing assertions about well-being that are justified in terms of idiosyncratic and objectionable theories of well-being to count as theories of well-being to stand does not undermine the possibility of assessing the appropriateness or inappropriateness of a well-being claim in a particular context.

First, then, allowing theory-conditional assertions of well-being to stand does not award them any particular significance. It does grant that the speaker has made an intelligible statement about well-being, that is, that their claim that can be understood by someone else who is part of the same linguistic community. This accommodates the possibility that the claim is not thought to be relevant or appropriate by others. The claim that justice involves meting out harsh punishments for minor indiscretions is an intelligible claim. It also might be a commitment that is entailed by a particular theory of desert and human morality. That it is intelligible, though, does not entail that it is appropriate in many—indeed, any—contexts. Its intelligibility does entail that it is the kind of claim that is apt for assessment as appropriate or inappropriate in any context in which it is given as a justification, as explanation, or as evidence. The speaker in Alexandrova's example does appear to present an intelligible claim about well-being—she broadly takes well-being to be dependent on features of a person, to say something about how they are doing, or about what is good for them. The unacceptability of her claims for Alexandrova lies in the fact that she sets the well-being threshold in the wrong place, taking things that are often understood to be necessary for achieving a threshold level of well-being to be well above the threshold.

That a theory of well-being sets a surprisingly low threshold for the attribution of well-being, or that it classifies things usually taken to be as relevant to well-being as in fact irrelevant, does not make it unintelligible, though it may make it unconvincing. If someone ascribed well-being to people on the basis of whether it was raining in Kentucky, or took well-being to be a record of someone's height in centimetres, or understood well-being to be an attribute of wooden furniture rather than people, her well-being claims may be thought unintelligible. That is, she appears not to understand the meaning of well-being, nor the linguistic rules and conventions governing its use. In such a case it is not clear that her claims about well-being should be taken to be false or inappropriate, so much as nonsensical. They are simply not well-being claims, or at least, not well-being claims within our linguistic context, and so do not admit of the kind of assessment that is relevant to well-being claims.

The line between intelligibility and unintelligibility is not, I think, entirely clear. Some extremely eccentric conceptions of well-being—which define well-being in terms of hirsuteness, or blood pressure, or the number of siblings a person has—may toe the line between intelligible and unintelligible. Whether they are intelligible or not may not be determinable by appreciation of the claim in isolation, and might depend on a more general assessment of their speaker's theory and conception of well-being. But regardless, whether or not a claim is intelligible does not alone tell us anything about its relevance. Given this, the consequences of sanctioning idiosyncratic well-being claims are somewhat less problematic. It merely opens the door to an assessment of the appropriateness of the theory and the claim in question, rather than endorsing its content.

The second reason follows on from this. Even if assertions about well-being that are grounded in idiosyncratic theories of well-being are taken to count as theories of well-being, it is still possible to criticise or discount the relevance and appropriateness of such assertions and theories in a given context. Classifying something as an intelligible

claim about well-being or an intelligible theory of well-being doesn't mean it's a useful claim or a good theory. Given the constructive role that the measuring subject plays in theorising about well-being, the appropriateness and usefulness of a theory of well-being and any resulting claims about well-being will depend on the purpose they are intended to serve. For a psychometrician seeking to develop a new measure of well-being for use in children living with foster carers, the appropriateness of her theory will be judged relative to the conventions of her discipline, on the basis of its psychometric validity, for example, or on its not merely repeating the work of other academics, or on its logistical and practical applicability. The appropriate standards for assessing theories of well-being and claims about well-being for the psychometrician will be quite different compared to the parent who seeks to assess whether his children will be better off if he moves to the countryside, or continues to live in the city. In both cases, the measuring subject's theory and claims will be assessed on the basis of whether they adequately take the needs and interests of the objects of well-being into account; include relevant evidence to justify their claims and exclude irrelevant evidence; and justify the inference made on the basis of their findings. However, the standards that they will be expected to adhere to in assessing needs, collecting and evaluating evidence, and offering justification will be quite different. What might be interpreted as sloppy reasoning, poor selection of evidence or inadequately thorough consideration of needs for the psychometrician might be entirely appropriate, and even supererogatory, in the case of the parent. Furthermore, depending on the context in which well-being is measured, there may be a fairly fixed received view on the appropriate theoretical framework to use, or, typically in informal, personal settings, there might be almost no restriction at all on the theory of well-being that a given measuring subject can use. There might even be a best possible answer about how to understand well-being in a given context—the most appropriate way of understanding well-being—but this will always be determined relative to the purposes of the speaker. Indeed, on a variantist view, without considering the purposes

of the speaker, it's difficult to make sense of a particular ascription of well-being, insofar as they form part of the context in which well-being claims acquire their semantic content.

The purposes of a speaker are, of course, also open to criticism—we might think, for example, someone's purpose is misguided or even harmful. But the value of such criticism may depend on the scope and character of the purpose in question. A particular theory of well-being may, for example, serve the purposes of a given speaker in developing their own view of the world, in making sense of events in their life, or in making a personal decision, without necessarily serving any wider public purpose. In such a case criticism may be justified if it seeks to help the person in question—if we deem our friend to be in danger of harming herself through her well-being assessments, either of herself or of other people. However, it's less clear that such criticism has value when it merely seeks to assert the superiority of some other purpose—our own, perhaps. Of course, typically people's purposes are not entirely private, when assessing the well-being of others, and their assessments may have far reaching and significant implications. In such cases, and likely in relation to the extent and scope of impact that their assessments will have, appraisal of the purposes, and appraisal of people's theorising about well-being, will become increasingly important. The parent appraising their children's future well-being may be accountable to their partner, their children, their own parents, and in some cases to legal and social care professionals. The psychometrician developing a measure of well-being will be accountable to their disciplinary community, to the people who will use their instrument, and perhaps to the people whose lives will be affected by the operationalisation of their instrument.

So, the ascription of theory-conditional knowledge claims to people need not imply that we find their claims are relevant, or useful, or that we endorse their purposes. The relevance and appropriateness of a theory of well-being, and of particular assessments

of well-being, for a given context will largely depend on the purposes of the measuring subject—the reasons for which they are measuring well-being. But relative to such purposes, there is plenty of space for criticism of theoretical frameworks and particular assessments of well-being. This ought to assuage some fears about the ugly consequences of permitting such knowledge claims to stand. If the consequences of allowing just the features salient to the speaker herself to be either constitutive of well-being or to determine the right threshold for ascribing well-being aren't as ugly as Alexandrova supposes, then there's less reason for thinking that this option should be ruled out.

Cartwright and Runhardt argue that affording value laden choices to those designing and using measures has drawbacks: the resulting proliferation of measures will be confusing, will allow people to cherry pick measures for their purposes, and will make it difficult to accumulate knowledge when different people use different measures and thus turn out to be studying different things (Cartwright and Runhardt 2014, pp.275–6). They argue that this should drive us to produce common metrics for central social science concepts, though they acknowledge that this may entail distortions of meaning, failures of fitness for purpose, and lack of nuance and detail. The same may be said to apply to well-being concepts. I broadly agree with Cartwright and Runhardt's enjoyment that a balance should be found among different aims. Common metrics can play an important role in enabling comparisons across people and groups. They allow, for example, measures of well-being to be used in decision making about resource distribution. That a particular conception and measure of well-being has been used by someone else for a purpose related to mine may well provide some reason for using that same measure. But other factors must be considered—does the measure and conception of well-being suit my purposes in other ways; is the person or group whose well-being I am measuring relevantly similar to theirs; will comparison with other people or groups produce useful, meaningful information?

Answers to such questions may speak against the use of common metrics. Cartwright and her collaborators seem to think that balancing fitness for purpose against a drive towards the use of common metrics will tend towards the use of fairly standardised measures in social sciences (Cartwright and Runhardt 2014; Cartwright and Bradburn 2011; Chang and Cartwright 2008). My intuitions tend to pull in the opposite direction, at least in the case of well-being. Comparison of well-being across different groups is of course necessary for certain purposes, but it is not valuable for its own sake, and ensuring that conceptions of well-being are fit for purpose and meet the specific requirements of the context may end up being more important in most cases. Common measures will not lead to the accumulation of knowledge if the measures are inappropriate for the context in which they are employed.

Alexandrova (2013) is chiefly concerned with the way in which a speaker's context determines the truth values of well-being claims. Part of the reason that she wants to avoid allowing the salience of the possibility of error to the speaker to determine the truth value of well-being claims is that it would permit the Gulag-fixated speaker to claim knowledge about Julia's well-being on the basis of her theory. This seems to make knowledge ascriptions depend on the speaker's psychological state, which sets the bar for knowledge far too low. If we allow a speaker's psychological state to determine the semantic content of well-being claims, we end up allowing implausible knowledge claims. My chief concern with Alexandrova's argument that the objective practical environment of the speaker should fix the semantic content of their well-being claims, rather than their psychological state, is that their practical environment is not plausibly understood as 'objective' for the purpose of well-being measurement. That is, if my relational conception of the measurement of well-being is right, then the features of the environment that are determinative of well-being are always shaped by some person or group of people, insofar as they can only be as relevant to well-being under some theoretical framework. This means that there is no such thing as the

‘objective practical environment’ of the speaker—any attempt to object to a speaker’s particular assessment of well-being requires further theorising of the object of measurement, a measuring instrument and its environment. If this is right then, insofar as it is possible to make knowledge claims about well-being, their content will be inescapably dependent on the psychological state of the speaker.

Alexandrova clearly wants to allow for the possibility of criticism of particular individual’s theories and ascriptions of well-being, but criticism too requires the modelling of a practical environment, including the object of well-being and the well-being instrument, in light of some theoretical framework. Evaluation of the practical environment in order to contest a particular conception of well-being inevitably involves a theoretical construction, which models well-being indications in order to generate information about well-being. The criticism of a particular measuring subject may garner significant agreement among a relevant community, but this does not make their interpretation of the practical environment objective. In this sense, the semantic content of well-being claims is never fixed by the objective practical environment, but a theoretical model of the practical environment *from some perspective*. On my account, the final authority about what standard of well-being is locally appropriate *is* some agent’s actual judgment about well-being, insofar as it makes sense to call the standard of well-being that is agreed by the relevant community to be best suited to a particular context a ‘final authority’ (Alexandrova 2013, p.321). One reason to avoid such language is that my account takes any theory and assessment of well-being to be defeasible—a well-being claim currently deemed appropriate may be supplanted by a more convincing theory, a better characterisation of the object of measurement, new evidence, new methods of collecting and evaluation evidence, and so on. The authority is never final, and the possibility of alternative, and better, characterisations remains ever-salient.

All this suggests that, if well-being claims are to be understood as knowledge claims, the bar for knowledge claims must be set 'low' enough for the salience of the possibility of error to the speaker to determine their truth value. Otherwise, if this really does constitute too low a bar for knowledge, well-being claims cannot be taken as knowledge claims. I am not too concerned as to whether well-being claims should be understood to be knowledge claims or not, and whether they should ever be thought to be true or false. Without ascription of knowledge, truth, objectivity, and so on we can go quite a long way in criticising and defending particular well-being claims, and can be justified in using or rejecting them in particular contexts.<sup>54</sup> Whether they meet the standard of knowledge or not might not matter that much, so long as they continue to play a useful role in both private and public decision making.

In this final chapter, I have endeavoured to develop a coherent stance dependent pluralist account of well-being, and to defend it against anticipated criticism. I have not sought to provide a knock-down argument against stance independence—indeed, I regard such a project as likely fruitless. The disagreement between stance dependent and stance independent pluralism about well-being is a localised version of a realism/constructivist debate in moral philosophy more broadly, and theorists on both sides of the divide are largely driven by competing and mutually exclusive intuitions, and are prepared to bite the bullet and attempt to explain away concerning implications of their view. I have, instead, tried to make space for a stance dependent alternative by exploring the form that it might take, how it could make sense of the apparent attitude- and context- dependence of measures of well-being, and how it

---

<sup>54</sup> Richard Rorty (1998, particularly introduction and chapter 1) takes the pragmatist to be unable to drop the notion of truth, given the role it plays in people's interactions with the world. But she can imagine and eulogise about a world where this concept no longer exists, and can attempt to show the futility of a metaphysical belief in truth. This seems to me to be the right way to think about truth- and knowledge- claims about well-being.



could remedy the inadequacies of stance independent pluralism. Initial concerns about stance dependence, turn out, I argued, to be less problematic than they might at first have seemed. Stance dependent pluralism therefore stands as a plausible and relevant alternative to existing accounts of well-being.

## Conclusion

I have sought to contribute to the philosophical literature on well-being, not by developing an alternative theory of well-being, but by proposing an alternative way of thinking about theories of well-being. The thrust of my thesis can be distilled down to three main claims. The first is a claim about context-dependence. Well-being means different things in different contexts, and there is no defining characteristic that is shared by all and only instances of well-being. The second is a claim about measurement. Well-being is a measurable attribute, where measurement is not just incidental to its definition but central to understanding the concept of well-being. Adequately specified philosophical theories of well-being must say something about the measurement of well-being. The third is a claim about stance dependence. Assessments of well-being in particular contexts are highly dependent not only on features of the context and environment, but also on the attitudes and purposes of the measuring subjects. There is no objective way of characterising someone's well-being, as measuring well-being involves modelling the relationship between the object of measurement and their well-being.

Adopting a stance dependent pluralism about well-being has practical, as well as philosophical, implications. Stance dependent pluralism looks able to justify psychometric validation and reliability tests, the widely-accepted gold standard in social and clinical sciences for assessing the appropriateness of a measure of well-being, quality of life, or life satisfaction (Alexandrova 2017a). Such methods rely on correlation between constructs, measures, and people's attitudes about well-being to

justify claims that measures of well-being do in fact capture well-being, and not something else. My account also provides a reason to resist a move, in policy contexts, towards the standardisation of quality-of-life measurement. Using the same measure across contexts might be appropriate for certain purposes but it may be entirely inappropriate for other purposes. In the healthcare context, for instance, whereas standardisation of well-being measures may be well justified in the context of decision making about resource distribution, which demands a particular kind of public accountability, it might be inappropriate for determining the benefit of treatments and interventions for clinical decision making.

For the person seeking to measure well-being, a stance dependent conception of well-being is both a blessing and a curse: a blessing because it frees them from striving for correctness in their attempts to develop an appropriate conception of well-being and allows them to question and deviate from existing conventions and approaches, but a curse because it foregrounds the amount of power that they, as measuring subjects, have in determining both the form and the content of the information they produce about well-being. Awareness of this power might seem like it would lead to paralysis: if there is an indeterminable diversity of ways in which you could approach the measurement of well-being in any given context, and a variety of ways in which you could describe and interpret your purposes, how can you go about choosing how to proceed? But such concern would be unwarranted. Those attempting to define and measure well-being do so within existing practices, whether those are broad socio-linguistic practices, or more strictly defined disciplinary practices, and these confine, and largely or partially determine, what their purposes are and how they should be conceived. Our practices of well-being measurement are already stance dependent, and while drawing attention to this fact might encourage scrutiny of conventions about well-being measurement previously assumed to be inviolable, it should not precipitate extensive changes to those practices.

One concern with my approach is that it passes the buck by replacing the work involved in determining the appropriate account of well-being for a given context with the similarly burdensome undertaking of determining the appropriate purposes and attitudes for a given context. Justifying a particular conception of well-being involves specifying and justifying one's purposes, and showing how those purposes occasion this conception of well-being rather than another one. This is potentially concerning for public policy purposes, as determining what one's purposes are, let alone determining what one's purposes should be, is a complex task. It will likely involve decision-making frameworks, deliberation, and will undoubtedly be a distinctively normative activity. My arguments about the stance dependence of about well-being suggest, moreover, that specifying one's attitudes and purposes will also be a constructive task. If attempts to introspectively identify one's subjective psychological states are mediated and framed, so too will be attempts to determine one's purposes. I regret that my thesis cannot cover this aspect of the definition and measurement of well-being, as it is undoubtedly a large part of the picture, given the account of well-being I have defended. However, insofar as I have passed the buck, I have gained some ground. If defining and measuring well-being in an accountable and justified manner is taken to involve defining and specifying one's purposes and one's theoretical framework, the terms of the debate have shifted onto firmly pragmatist ground. The question is no longer whether a concept and theory of well-being accurately reflects objective facts about people and the world—whether it captures the way things really are—but whether a concept and theory of well-being adequately captures our attitudes and interests, and whether it enables us to do what we want to do.

In one sense, parts of my argument are circular. Perhaps this is not something I should admit to. The project depends on an instinct to develop an inclusive, rather than exclusive, account of well-being. And it is therefore no great surprise that I end up with an ecumenical account of well-being. I also place the practice of defining and

identifying well-being for particular purposes in the foreground of philosophical considerations about well-being, so the suggestion that measurement is central to understanding well-being is likely equally unsurprising. At the end of chapter 1, I argued that there is a deep philosophical chasm between the *a priori* approach to the philosophy of well-being and my own approach. I view the prospects of bridging the divide via a straightforward deductive argument to be remote, if not entirely mythical. Our respective positions are characterised by substantive metaethical disagreements and reflect, I think, different conceptions of not only the relationship between moral claims and the world, but also the relationship between language and the world more generally. The divergence is also characterised by a realism-antirealism debate, which bubbled to the surface at several points in the thesis, not only specifically with respect to prudential value, but also in the contexts of measurement theory and metaethical theory. Evidence, counter-examples, and arguments are not guaranteed to change anyone's mind, as defendants of either account will be inclined to interpret them through their own framework, and thus evaluate their meaning and implications quite differently. In light of this, a certain degree of circularity looks to be difficult to avoid in certain kinds of philosophical disagreement and argument. Instead of trying to construct a knock-down argument, I have endeavoured to build a promising picture of what a pragmatist conception of well-being might look like—to flesh out some of its details, and to explore how it overcomes some of the deficiencies of alternative approaches. If a critical argument looks doomed to failure, building a positive picture to show how an alternative to the status quo might look like seems to me to be the best way to proceed.

In spite of this, I do intend for my argument to be accessible—perhaps even convincing—to some people who are not already disposed to agree with it, even if it is unlikely to win over anyone who is fully committed to the view that logic reflects, rather than regulates, language. There is something difficult to pin down about well-

being, I think partly because of the space that it occupies between fact and value. Many theories of well-being recognise this—perfectionist and objectivist theories in particular—and this leads them towards a pluralism about well-being, which better reflects the ways in which the concept of well-being is used. I have endeavoured to show that the path that leads towards pluralism also leads away from attitude independence. For those who accept the need for a pluralist account of well-being and the need for philosophy about well-being to consider the manifestation of well-being in the people and in the world, I have shown that the identification of well-being involves much more than providing high-level definitions, and that the prospect of a fully specified account of well-being remaining objective in any meaningful sense is very limited.



## Bibliography

- Alexandrova, Anna (2005), 'Subjective Well-Being and Kahneman's 'Objective Happiness'', *Journal of Happiness Studies* 6(3): 301–324.
- (2008), 'First-Person Reports and the Measurement of Happiness,' *Philosophical Psychology* 21(5): 571–583.
- (2012a), 'Values and the Science of Well-Being: A Recipe for Mixing,' in Harold Kincaid (ed.) *The Oxford Handbook of Philosophy of Social Science*, Oxford: Oxford University Press.
- (2012b), 'Well-Being as an Object of Science,' *Philosophy of Science* 79(5): 678–689.
- (2013), 'Doing Well in the Circumstances,' *Journal of Moral Philosophy* 10(3): 307–328.
- (2017a), 'Is Well-being Measurable After All,' *Public Health Ethics* 10(2), Symposium on Daniel Hausman's Valuing Health: Well-Being, Freedom and Suffering: 129–137.
- (2017b), *A Philosophy for the Science of Well-Being*, Oxford: Oxford University Press.
- Alexandrova, Anna and Daniel M. Haybron (2016), 'Is Construct Validation Valid?' *Philosophy of Science* 83(5): 1098–1109.
- Anand, Paul (1987), 'Are the Preference Axioms Really Rational,' *Theory and Decision* 23(2), 189–214.
- Anand, Sudhir and Amartya Sen (1994), 'Human Development Index: Methodology and Measurement,' *Human Development Report Office Occasional Paper 12*, New York: UNDP.



- Anscombe, G.E.M. (1959), *An Introduction to Wittgenstein's Tractatus*, London: Hutchinson & Co.
- Aristotle (2009), *The Nicomachean Ethics*, W.D. Ross (trans.) and Lesley Brown (ed.), Oxford: Oxford University Press.
- Arrow, Kenneth J. (1982), 'Risk Perception in Psychology and Economics,' *Economic Inquiry* 20(1), 1–9.
- Audi, Robert (2015), 'Intuition and Its Place in Ethics,' *Journal of the American Philosophical Association* 1(1): 57–77.
- Ben-Menahem, Yemima (2006), *Conventionalism*, New York, NY: Cambridge University Press.
- Bentham, Jeremy (1843), 'An Introduction to the Principles of Morals and Legislation,' in John Bowring (ed.) *The Works of Jeremy Bentham, published under the Superintendence of his Executor Vol.1 (1788–1843)*, Edinburgh: William Tait.
- Berlin, Isaiah (2013), 'The Pursuit of the Ideal' in his *The Crooked Timber of Humanity*, Princeton: Princeton University Press.
- Bermúdez, José Luis (2009), *Decision Theory and Rationality*, New York: Oxford University Press.
- Borges, Luis Jorge (1964), *Other Inquisitions 1937-1952*, Ruth L.C. Simms (trans.), Austin: University of Texas Press.
- van der Bos (1994), 'Parametric statistical model-based measurement,' *Measurement* 14(1): 55–61.
- Bowling, Ann (2005), *Measuring Health: A review of quality of life measurement scales (third edition)*, Berkshire: Open University Press.
- Boyd, Richard N. (1988), 'How to Be a Moral Realist,' in Geoffrey Sayre-McCord (ed.) *Essays on Moral Realism*, Ithaca and London: Cornell University Press.
- (2003), 'Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part I,' *Philosophy and Phenomenological Research* 66(3): 505–553.
- Bradburn, Norman M., Nancy L. Cartwright and Jonathan Fuller (2017), 'A Theory of Measurement,' in L.M. McClimans (ed.) *Measurement in Medicine: Philosophical*

- Essays on Assessment and Evaluation*, London: Rowman & Littlefield International.
- Bradford, Gwen (2015), 'Perfectionism,' in Guy Fletcher (ed.) *The Routledge Handbook of Philosophy of Well-Being*, Abingdon: Routledge.
- Bradley, Ben (2009), *Well-Being and Death*, Oxford: Oxford University Press.
- Brandt, Richard B. (1979), *A Theory of the Good and the Right*, Oxford: Clarendon Press.
- Bramble, Ben (2016), 'The Experience Machine,' *Philosophy Compass* 11(3): 136–145.
- Brazier J., J. Roberts and M. Deverill (2002), 'The estimation of a preference-based measure of health from the SF-36,' *Journal of Health Economics* 21(2): 271–292.
- Bridgman, Percy (1927) *The Logic of Modern Physics*, London: Macmillan.
- Brink, David O. (1986), 'Externalist Moral Realism,' *The Southern Journal of Philosophy* 24 (supp. vol.): 23–41.
- Broome, John (1991), 'Utility,' *Economics and Philosophy* 7(1): 1–12.
- (2004), *Weighing Lives*, Oxford: Clarendon Press.
- Bruckner, Donald W. (2016), 'Quirky Desires and Well-Being,' *Journal of Ethics and Social Philosophy* 10(2): 1–34.
- Buchanan, Allen and Russell Powell (2018), *The Evolution of Moral Progress: A Biocultural Theory*, Oxford: Oxford University Press.
- Campbell, Norman Robert (1928), *An Account of the Principles of Measurement and Calculation*, London: Longmans, Green and Co.
- Campbell, Stephen (2015), 'The Concept of Well-Being,' in Guy Fletcher (ed.) *The Routledge Handbook of the Philosophy of Well-Being*, Abingdon: Routledge.
- Cartwright, Nancy (1999), *The Dappled World: A Study of the Boundaries of Science*, Cambridge, UK: Cambridge University Press.
- Cartwright, Nancy, and Norman M. Bradburn (2011), 'A Theory of Measurement,' in Rose Maria Li (ed.) *The Importance of Common Metrics for Advancing Social Science Theory and Research: A Workshop Summary*, Washington: National Academies Press.

- Cartwright, Nancy and Rosa Runhardt (2014), 'Measurement,' in Nancy Cartwright and Eleanora Montuschi (eds.) *Philosophy of Social Science: A New Introduction*, Oxford: Oxford University Press.
- Chang, Hasok and Nancy Cartwright (2008), 'Measurement,' in Stathis Psillos and Martin Curd (eds.) *The Routledge Companion to Philosophy of Science*, London: Routledge.
- Cohen, G.A. (2003), 'Facts and Principles,' *Philosophy & Public Affairs* 31(3): 211–245.
- Crisp, Roger (2006), *Reasons and the Good*, Oxford: Oxford University Press.
- Dancy, Jonathan (1986), 'Two Conceptions of Moral Realism,' *Proceedings of the Aristotelian Society* 60 (supp. vol.): 167–187.
- Davidson, Richard J. (1992), 'Anterior Cerebral Asymmetry and the Nature of Emotion,' *Brain and Cognition* 20(1): 125–151.
- Dasgupta, Partha (2001), *Human Well-Being and the Natural Environment*, Oxford: Oxford University Press.
- Dewey, John (1925), *Experience and Nature*, New York: W. W. Norton and Co.
- (1927), *The Public and Its Problems*, Chicago: Sage Books.
- (1929), *The Quest for Certainty: A Study of the Relation of Knowledge and Action*, New York: Minton, Balch & Company.
- Dolan, Paul and Kahneman, Daniel (2008), 'Interpretations of Utility and their Implications for the Valuation of Health,' *The Economic Journal* 118: 215–234.
- Dupré, John (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Cambridge, MA: Harvard University Press.
- Dworkin, Gerald (1988), *The Theory and Practice of Autonomy*, Cambridge, UK: Cambridge University Press.
- Ellis, Brian (1966), *Basic Concepts of Measurement*, Cambridge, UK: Cambridge University Press.
- (1992), 'Conventionalism in Measurement Theory,' in C. Wade Savage and Philip Ehrlich (eds.) *Philosophical and Foundational Issues in Measurement Theory*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- The EuroQol Group (1990), 'EuroQol - a new facility for the measurement of health-related quality of life,' *Health Policy* 16(3): 199–208.

- Feldman, Fred (2002), 'The Good Life: A Defense of Attitudinal Hedonism,' *Philosophy and Phenomenological Research* 65(3): 604–628.
- (2004), *Pleasure and the Good Life*, Oxford: Clarendon Press.
- Ferrans, C.E. and M.J. Powers (1985), 'Quality of life index: development and psychometric properties,' *Advances in Nursing Science* 8(1): 15–24.
- Finnis, John (1980), *Natural Law and Natural Rights*, Oxford: Clarendon Press.
- Fletcher, Guy (2009), 'Rejecting Well-Being Invariabilism,' *Philosophical Papers* 38(1): 21–34.
- (2013), 'A Fresh Start for the Objective-List Theory of Well-Being,' *Utilitas* 25(2): 206–220.
- (2015), 'Objective List Theories,' in Guy Fletcher (ed.) *The Routledge Handbook of the Philosophy of Well-Being*, Abingdon: Routledge.
- (2016), *The Philosophy of Well-Being*, Abingdon: Routledge.
- Foot, Philippa (2001), *Natural Goodness*, Oxford: Clarendon Press.
- Forster, Michael (2010), 'Wittgenstein on family resemblance concepts' in Arif Ahmed (ed.) *Wittgenstein's 'Philosophical Investigations': A Critical Guide*, Cambridge, UK: Cambridge University Press.
- van Fraassen, Bas (2001), 'Constructive Empiricism Now,' *Philosophical Studies* 106: 151–170.
- (2008), *Scientific Representation: Paradoxes of Perspective*, Oxford: Oxford University Press.
- Frigerio, A., A. Giordani, and L. Mari (2010), 'Outline of a general model of measurement,' *Synthese*, 175(2): 123–149.
- Gauthier, David (1986), *Morals by Agreement*, Oxford: Clarendon Press.
- Goldbart, Juliet and Sue Caton (2010), *Communication and people with the most complex needs: What works and why this is essential*, Mencap, online resource, [https://e-space.mmu.ac.uk/198309/1/Mencap%20Comms\\_guide\\_dec\\_10.pdf](https://e-space.mmu.ac.uk/198309/1/Mencap%20Comms_guide_dec_10.pdf), accessed 02/08/2018.
- Goodin, Robert (1986) 'Laundering preferences,' in Jon Elster and Aanund Hylland (eds.) *Foundations of Social Choice Theory*, Cambridge, UK: Cambridge University Press.

- Green, C.P., C.B. Porter, D.R. Bresnahan and J.A. Spertus (2000), 'Development and Evaluation of the Kansas City Cardiomyopathy Questionnaire: A New Health Status Measure for Heart Failure,' *Journal of the American College of Cardiology* 35(5): 1245–1255.
- Griffin, James (1988), *Well-Being: Its Meaning, Measurement and Moral Importance*, Oxford: Clarendon Press.
- Grove, Nicola, Karen Bunning, Jill Porter, and Cecilia Olsson (1999), 'See What I Mean: Interpreting the Meaning of Communication by People with Severe and Profound Intellectual Disabilities,' *Journal of Applied Research in Intellectual Disabilities* 12(3): 190–203.
- Gudex, Claire (ed.) (1994a), 'Time Trade-Off User Manual: Props and Self-Completion Methods,' *Centre for Health Economics Occasional Paper 20*, York: The University of York.
- (1994b), 'Standard Gamble User Manual: Props and Self-Completion Methods,' *Centre for Health Economics Occasional Paper 21*, York: The University of York.
- Hampton, Jean (1994), 'The Failure of Expected-Utility Theory as a Theory of Reason,' *Economics and Philosophy* 10(2), 195–242.
- Hare, D.L. and C.R. Davis (1996), 'Cardiac Depression Scale: Validation of a New Depression Scale for Cardiac Patients,' *Journal of Psychosomatic Research* 40(4): 379–386.
- Harsanyi, John C. (1977), 'Morality and the Theory of Rational Behavior,' *Social Research* 44: 623–656.
- (1998), 'A Preference-Based Theory of Well-Being and a Rule-Utilitarian Theory of Morality,' in W. Leinfellner and E. Köhler (eds.) *Game Theory, Experience, Rationality: Foundations of Social Sciences, Economics and Ethics. In Honor of John C. Harsanyi*, Vienna Circle Institute Yearbook vol. 5, Dordrecht: Springer.
- Hausman, Daniel M. (2015), *Valuing Health: Well-Being, Freedom, and Suffering*, Oxford: Oxford University Press.
- (2017), 'Responses to my Critics,' *Public Health Ethics* 10(2), Symposium on Daniel Hausman's Valuing Health: Well-Being, Freedom and Suffering: 164–175.

- Haybron, Daniel M. (2016), 'Mental State Approaches to Well-Being,' in Matthew D. Adler and Marc Fleurbaey (eds.) *The Oxford Handbook of Well-Being and Public Policy*, Oxford: Oxford University Press.
- Hazlett, Allan (2013), *A Luxury of the Understanding: On the Value of True Belief*, New York, NY: Oxford University Press.
- Heathwood, Chris (2005), 'The Problem of Defective Desires,' *Australasian Journal of Philosophy* 83(4): 487–504.
- (2006), 'Desire Satisfaction and Hedonism,' *Philosophical Studies* 128(3): 539–563.
- (2015), 'Desire-Fulfillment Theory,' in Guy Fletcher (ed.) *The Routledge Handbook of the Philosophy of Well-Being*, Abingdon: Routledge.
- Herdman, M., C. Gudex, A. Lloyd, M.F. Janssen, P. Kind, D. Parkin, G. Bonsel, X. Badia (2011), 'Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L),' *Quality of Life Research* 20(10): 1727–1736.
- Hunt, S.M., S.P. McKenna, J. McEwen, E.M. Backett, J. Williams, and E. Papp (1980), 'A quantitative approach to perceived health status: a validation study,' *Journal of Epidemiology and Community Health* 34(4): 281–286.
- Hurka, Thomas (1993), *Perfectionism*, New York: Oxford University Press.
- (2016), 'Objective Goods,' in Matthew D. Adler and Marc Fleurbaey (eds.) *The Oxford Handbook of Well-Being and Public Policy*, Oxford: Oxford University Press.
- James, William (1891), 'The Moral Philosopher and the Moral Life,' *International Journal of Ethics* 1(3): 330–354.
- Kahneman, Daniel (1999), 'Objective happiness,' in Daniel Kahneman, Ed Diener and Norbert Schwarz (eds.) *Well-Being: The Foundations of Hedonic Psychology*, New York: Russell Sage Foundation.
- (2000a) 'Experienced Utility and Objective Happiness: A Moment-Based Approach,' in Daniel Kahneman and Amos Tversky (eds.) *Choices, Values and Frames*, New York: Cambridge University Press and the Russell Sage Foundation.

- (2000b), 'Evaluation by Moments: Past and Future,' in Daniel Kahneman and Amos Tversky (eds.) *Choices, Values and Frames*, New York: Cambridge University Press and the Russell Sage Foundation.
- (2003), 'A Perspective on Judgement and Choice: Mapping Bounded Rationality,' *American Psychologist* 58(9): 697–720.
- Kahneman, Daniel and Alan B. Krueger (2006), 'Developments in the Measurement of Subjective Well-Being,' *Journal of Economic Perspectives* 20(1): 3–24.
- Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone (2004), 'A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method,' *Science* 306(5702), 1776–1780.
- Kahneman, Daniel and Jason Riis (2005), 'Living, and thinking about it: two perspectives on life,' in Felicia A. Huppert, Nick Baylis and Barry Keverne (eds.) *The Science of Well-Being*, Oxford: Oxford University Press.
- Kahneman, Daniel and Richard H. Thaler (2006), 'Anomalies: Utility Maximization and Experienced Utility,' *The Journal of Economic Perspectives* 20(1): 221–234.
- Kahneman, Daniel and Amos Tversky (1979), 'Prospect Theory: An Analysis of Decision under Risk,' *Econometrica* 47(2): 263–291.
- Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin (1997), 'Back to Bentham? Explorations of Experienced Utility' in *The Quarterly Journal of Economics* 112(2): 375–406.
- Kagan, Shelly (1988), 'The Additive Fallacy,' *Ethics* 99(1): 5–31.
- (1992), 'The Limits of Well-Being,' *Social Philosophy and Policy* 9(2): 180–198.
- (1994), 'Me and My Life,' *Proceedings of the Aristotelian Society* 94(1): 309–324.
- (2009), 'Well-Being as Enjoying the Good,' *Philosophical Perspectives* 23(1): 253–272.
- Kitcher, Philip (1999), 'Essence and Perfection,' *Ethics* 110(1): 59–83.
- (2012), 'The Lure of the Peak,' *The New Republic*, February 2.
- (2015), 'Pragmatism and Progress,' *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy* 51(4): 475–494.
- Korsgaard, Christine M. (1983), 'Two Distinctions in Goodness,' *The Philosophical Review* 92(2): 169–195.

- Kraut, Richard (2007), *What is Good and Why: The Ethics of Well-Being*, Cambridge, MA: Harvard University Press.
- Larson, Reed and Mihaly Csikszentmihalyi (1983), 'The Experience Sampling Method,' in Mihaly Csikszentmihalyi (ed.) *Flow and the Foundations of Positive Psychology*, Rotterdam, NL: Springer.
- Lenman, James (2013), 'Science, Ethics and Observation,' *Royal Institute of Philosophy Supplement* 72: 261–274.
- Lin, Eden (2018), 'Welfare Invariabilism,' *Ethics* 128(2): 320–345.
- London School of Hygiene and Tropical Medicine (2009), *The use of epidemiological tools in conflict-affected populations: open-access educational resources for policy-makers*, online resource, <http://conflict.lshtm.ac.uk/>, accessed 02/08/2018.
- Lynch, Michael P. (1998), *Truth in Context: An Essay on Pluralism and Objectivity*, Cambridge, MA: The MIT Press.
- Mach, Ernst (1986), *Principles of the Theory of Heat*, Thomas J. McCormack (trans.), Dordrecht: D. Reidel.
- MacLeod, Andrew K. (2015), 'Well-Being: Objectivism, Subjectivism or Subjectivism?' *Journal of Happiness Studies* 16(4): 1073–1089.
- Mari, Luca (1996), 'The meaning of "quantity" in measurement,' *Measurement* 17(2): 127–138.
- (2000), 'Beyond the representational viewpoint: a new formalization of measurement,' *Measurement* 27(2): 71–84.
- Marschak, Jacob (1950), 'Rational Behavior, Uncertain Prospects, and Measurable Utility,' *Econometrica* 18(2): 111–141.
- Marshall, Alfred (1920), *Principles of economics*, New York: Macmillan & Co., Ltd.
- McClimans, Leah (2010), 'A theoretical framework for patient-reported outcome measures,' *Theoretical Medicine and Bioethics* 31(3): 225–240.
- McDowell, John (1986), 'Critical Notice' of Bernard Williams' *Ethics and the Limits of Philosophy*, *Mind* 95(379): 377–386.
- Mill, J.S. (1998), 'Utilitarianism,' in John Gray (ed.) *On Liberty and Other Essays*, New York: Oxford University Press.



- Miller, K.H. and C.G. Grindel (2004), 'Comparison of Symptoms of Younger and Older Patients Undergoing Coronary Artery Bypass Surgery,' *Clinical Nursing Research* 13(3): 179–193.
- Mulhern, Brendan, Yan Feng, Koonal Shah, Mathieu F. Janssen, Michael Herdman, Ben van Hout and Nancy Devlin (2018), 'Comparing the UK EQ-5D-3L and English EQ-5D-5L Value Sets,' *Pharmacoeconomics* 36(6): 699–713.
- Murphy, Marc C. (1999), 'The Simple Desire-Fulfillment Theory,' *Noûs* 33(2): 247–272.
- Nagel, Thomas (1979), 'Death,' in his *Mortal Questions*, New York: Cambridge University Press.
- von Neumann, John and Morgenstern, Oskar (1944), *Theory of Games and Economic Behaviour*, Princeton: Princeton University Press.
- Nussbaum, Martha C. (1988), 'Nature, Function, and Capability: Aristotle on Political Distribution,' *Oxford Studies in Ancient Philosophy* 6 (supp. vol.): 145–184.
- (1992), 'Human Functioning and Social Justice: In Defense of Aristotelian Essentialism,' *Political Theory* 20(2): 202–246.
- (2001), *Women and Human Development: The Capabilities Approach*, Cambridge, UK: Cambridge University Press.
- (2008), 'Who Is the Happy Warrior? Philosophy Poses Questions to Psychology,' *Journal of Legal Studies* 37 (supp. vol. 2): 81–113.
- (2011), *Creating Capabilities*, Cambridge, MA: Harvard University Press.
- Oliver, Adam (2017), Distinguishing between Experienced Utility and Remembered Utility, *Public Health Ethics* 10(2), Symposium on Daniel Hausman's Valuing Health: Well-Being, Freedom and Suffering: 122–128.
- Parfit, Derek (1984), *Reasons and Persons*, Oxford: Oxford University Press.
- Pigou, A.C. (1932), *The Economics of Welfare*, fourth edition, London: Macmillan.
- Plach S.K. and S.M. Heidrich (2001), 'Women's perceptions of their social roles after heart surgery and coronary angioplasty,' *Heart and Lung* 30(2): 117–127.
- Qizilbash, Mozaffar (1997), 'Pluralism and Well-Being Indices,' *World Development* 25(12): 2009–2026.

- Raghavan, Ramesh and Anna Alexandrova (2015), 'Toward a Theory of Child Well-Being,' *Social Indicators Research* 121(3): 1–16.
- Railton, Peter (1986) 'Facts and Values,' *Philosophical Topics* 14(2): 5–31.
- (2003), 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action,' in his *Facts, Values, and Norms: Essays toward a Morality of Consequence*, Cambridge, UK: Cambridge University Press.
- (2012), 'That Obscure Object, Desire,' *Proceedings and Addresses of the American Philosophical Association* 86(2): 22–46.
- Rawls, John (1972), *A Theory of Justice*, Oxford: Clarendon Press.
- (2005), *Political Liberalism: Expanded Edition*, New York: Columbia University Press.
- Read, Daniel (2007), 'Experienced utility: Utility theory from Jeremy Bentham to Daniel Kahneman,' *Thinking and Reasoning* 13(1): 45–61.
- Rector, T.S., S.H. Kubo and J.N. Cohn (1987), 'Patients' self-assessment of their congestive heart failure. Part 2: Content, reliability and validity of a new measure, the Minnesota Living with Heart Failure questionnaire,' *Heart Failure* 5(2):198–209.
- Redelmeier, Donald A. and Daniel Kahneman (1996), 'Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures,' *Pain* 66(1): 3–8.
- Redelmeier, Donald A., Joel Katz and Daniel Kahneman (2003), 'Memories of colonoscopy: a randomized trial,' *Pain* 104(1): 187–194.
- Rice, Christopher M. (2013), 'Defending the Objective List Theory of Well-Being,' *Ratio* 26(2): 196–211.
- Roberts, Debbie (2013), 'Thick Concepts,' *Philosophy Compass* 8(8): 677–688.
- Rorty, Richard (1998), *Truth and Progress: Philosophical Papers, Vol. 3*, Cambridge, UK: Cambridge University Press.
- Ross, W.D. (1930), *The Right and the Good*, Oxford: Oxford University Press.
- Savage, Leonard (1954), *The Foundations of Statistics*, New York, NY: Wiley.
- Savage, C. Wade and Philip Ehrlich (eds.) (1992), *Philosophical and Foundational Issues in Measurement Theory*, Hillsdale, NJ: Lawrence Erlbaum Associates.

- Scanlon, T.M. (1998), *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
- Schroeder, Mark (2016), 'Value Theory,' in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), online resource, <https://plato.stanford.edu/archives/fall2016/entries/value-theory/>, accessed 02/08/2018.
- Searle, J. R. (1969), *Speech acts: An essay in the philosophy of language*, London: Cambridge University Press.
- Sen, Amartya K. (1973), 'Behaviour and the Concept of Preference, *Economica* 40(159): 241–259.
- (1977), 'Rational Fools: A Critique of the Behavioural Foundations of Economic Theory,' *Philosophy & Public Affairs* 6 (4): 317–344.
- (1985a), *Commodities and Capabilities*, Professor Dr. P. Hennipman Lectures in Economics, vol. 7: Theories, Institutions, Policy, Amsterdam: North-Holland.
- (1985b), 'Well-Being, Agency and Freedom: The Dewey Lectures 1984,' *The Journal of Philosophy* 82(4): 169–221.
- (1987), *The Standard of Living*, Cambridge, UK: Cambridge University Press.
- Shafer-Landau, Russ (2005), *Moral Realism: A Defence*, Oxford: Oxford University Press.
- (ed.) (2007), *Ethical Theory: An Anthology*, Chichester: John Wiley & Sons.
- Stevens, Stanley Smith (1946), 'On the Theory of Scales of Measurement,' *Science* 103(2684): 677–680.
- Stone, Arthur A., Broderick, Joan E., Porter, Laura S., and Kaell, Alan T. (1997), 'The Experience of Rheumatoid Arthritis Pain and Fatigue: Examining Momentary Reports and Correlates Over One Week,' *Arthritis Care and Research* 10(3): 185–193.
- Sturgeon, Nicholas L. (1985), 'Moral Explanations,' in David Copp and David Zimmerman (eds.) *Morality, Reason and Truth*, Totowa: Rowman & Allanheld.
- Sugden, Robert (2006), 'Taking unconsidered preferences seriously,' in Serena Olsaretti (ed.) *Preferences and Well-Being*, Cambridge, UK: Cambridge University Press.

- Sumner, L. W. (1996), *Welfare, Happiness, and Ethics*, Oxford: Clarendon Press.
- Swoyer, Chris (1987), 'The Metaphysics of Measurement,' in John Forge (ed.) *Measurement, Realism and Objectivity*, Australasian Studies in History and Philosophy of Science vol. 5, Dordrecht: Springer.
- Tal, Eran (2013), 'Old and New Problems in Philosophy of Measurement,' *Philosophy Compass* 8(12): 1159–1173.
- (2017a), 'A Model-Based Epistemology of Measurement,' in Nicola Mößner and Alfred Nordmann (eds.) *Reasoning in Measurement*, London: Routledge.
- (2017b), 'Calibration: Modelling the measurement process,' *Studies in History and Philosophy of Science* 65-66: 33–45.
- Thomson, Judith Jarvis (1997), 'The Right and the Good,' *Journal of Philosophy* 94(6): 273–298.
- Tiberius, Valerie (2011), 'How Theories of Well-Being Can Help Us Help,' *Journal of Practical Ethics* 2(2): 1–19.
- Tomlin Patrick (2018), 'Saplings or Caterpillars? Trying to Understand Children's Wellbeing,' *Journal of Applied Philosophy* 35 (supp. vol. 1): 29–46.
- Trout, J.D. (1998), *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences*, Oxford: Oxford University Press.
- Tubert, Ariela (2010), 'Constitutive Arguments,' *Philosophy Compass* 5(8): 656–666.
- Tversky, Amos and Daniel Kahneman (1981), 'The Framing of Decisions and the Psychology of Choice,' *Science* 211(4481): 453–458.
- (1986), 'Rational Choice and the Framing of Decisions,' *The Journal of Business* 59(4, part 2): 251–278.
- Varelius, Jukka (2004) 'Objective Explanations of Individual Well-Being,' *Journal of Happiness Studies* 5(1): 73–91.
- Velleman, J. David (1991), 'Well-being and Time,' *Pacific Philosophical Quarterly* 72(1): 48–77.
- Ware J.E. and C.D. Sherbourne (1992), 'The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection,' *Medical Care* 30(6): 473–83.

- Wiggins, David (1998), *Needs, Values, Truth: Essays in the Philosophy of Value*, Oxford: Oxford University Press.
- Wilson, Mark (2006), *Wandering Significance: An Essay on Conceptual Behaviour*, Oxford: Clarendon Press.
- Williams, Bernard (1981), 'Conflicts of Values' in his *Moral Luck: Philosophical Papers 1973–1980*, Cambridge, UK: Cambridge University Press.
- Winch, Peter (1990), *The idea of a social science and its relation to philosophy* (Second Edition), London: Routledge.
- Wittgenstein, Ludwig (1989), *Philosophical Investigations: 2nd edition*, G.E.M. Anscombe (trans.), Oxford: Blackwell.
- Woodard, Christopher (2013), 'Classifying theories of welfare,' *Philosophical Studies* 165(3): 787–803.