Measurement and Meaning in Health-Related Quality of Life Research

Leah Marian McClimans

London School of Economics and Political Science

Submitted for the degree of Doctor of Philosophy

Declaration

I hereby declare that the work presented in this thesis is my own.

Leah McClimans

Abstract

In this thesis I take up the topic of our understanding of questions in a detailed case study of non-utility measures of health-related quality of life. I argue that efforts to standardize these measures lead to limitations in our ability to understand and measure quality of life.

In the first half of this thesis I describe two types of bias that affect quality of life measures despite efforts to validate them. On the one hand, quality of life measures can perpetuate ethnocentric understandings of quality of life. On the other hand, respondents often understand the questions in these measures very differently than researchers imagined.

I argue that the residual bias found in quality of life measures is the result of two assumptions built into the use of construct validity: 1) when a measure's outcomes confirm our hypotheses, we are warranted in having greater confidence in the accuracy of our theory 2) respondents understand the questions and answers in our measures in the same way as researchers imagined they would.

In the second half of this thesis I argue that the limitations of construct validity stem from the logic of asking questions, a logic which precludes standardization. I propose that quality of life measures ought to be understood differently—they are not independent instruments capable of unambiguous claims, but rather one element in a dialogic framework whose questions and outcomes serve as the starting point for further inquiry.

Finally, I examine what might have motivated the misguided use of construct validity. I suggest that the motivation lies in an erroneous picture of the human subject. I argue for an alternative picture that allows me to introduce an ethical dimension to our questions about quality of life.

Contents

I. *Bas van Fraassen on Understanding Questions*

II. *Treating Respondents as Simple Weighers*

Acknowledgments

No one comes to the end of writing a PhD without the help and support of many people. My deepest gratitude goes to everyone who listened, comforted, and advised me over these years. I would like to personally thank the following people.

I have been exceptionally lucky to be supervised by two strong and brilliant women, Georgia Warnke and Nancy Cartwright. Without your comments, criticism and encouragement this thesis would surely have been impossible. I would also like to thank Larry Wright for his inspiration, his love of philosophy and for reminding me what it is to have a good quality life.

I would like to thank everyone who befriended me in the Clinical Effectiveness Unit at the Royal College of Surgeons, England. Especially I would like to thank: Barney Reeves for giving me my first job, taking an interest in my academic pursuits and writing me a wonderful recommendation; Julia Langham for giving me another much needed job; Lynn Copley for her infinite patience and understanding; Bumbi Singh for being such a good friend during difficult times, for listening and for all those squash games—I owe you a fiver.

I'd like to thank Jeremy Wisnewski and Josh Rust who came to LSE and gave me encouragement, advice and commented on Chapters 5 & 4 respectively. Your help, both personal and professional, was invaluable. Thank you to my students at the London School of Economics, especially Riccardo Finozzi, for trusting me to lead them into the world of European philosophy. I would like to thank Gary Jones for our long conversations, for teaching me so much about philosophy and for his inspiration. Thanks to Karen McKinnon and David Brewster for moving to Twickenham, for being reliable and kind and cooking me so many dinners. You are true friends.

The following people have given me an abundance of love and support. I am forever indebted to: Sally Williams who gave me a home and welcomed me into her life and family with such generosity and warmth—thank you for all the wine; Jay Conway a fellow traveler who reminds me who I am and where I'm going—thank you for that drive to UCSD; Georgia Warnke my role model in philosophy and life—thank you for lending me your car; Braulio Ramirez who is always there for me, always honest with me—thank you for your insight; Ashley Duffalo my *best* friend who knows all and perfectly understands everything—thank you for those years in Riverside.

Finally I would like to thank my family. My partner John Browne who first invited me to think critically about quality of life measurement and has since lived to regret it. Thank you for answering so many questions and taking so much abuse; thank you for endeavoring to make London like home. I would like to thank my father who nurtured my first philosophical instincts by challenging me to think things through carefully; I would like to thank my mother for giving me the ability to articulate what I think. I thank both my parents for giving me the idea that I can do whatever I put my mind to doing. Most of all, however, I need to thank my grandparents—for supporting me in everything I have ever wanted, for giving me money and sending me cards, for keeping me close even when I was far away, for listening and disagreeing and sometimes just not saying anything. You have made so many things possible.

Introduction

> What, is the highest of all the goods achievable in action?...both the many and the cultivated call it happiness, and they suppose that living well and doing well are the same as being happy. But they disagree about what happiness is, and the many do not give the same answer as the wise. For many think it is something obvious and evident—for instance, pleasure, wealth, or honor. Some take it to be one thing, others another. Indeed, the same person often changes his mind; for when he has fallen ill, he thinks happiness is health, and when he has fallen into poverty, he thinks it is wealth.[1]

It is not uncommon for essays on health-related quality of life or "quality of life" as I will refer to it in this thesis to begin with this quote by Aristotle. As I will explore in Chapter 1 for some quality of life researchers this passage signifies the idiosyncratic nature of quality of life; it reinforces the idea that what makes for a happy or good quality life is different for everyone. For others this passage is used to emphasize the difficulty as well as the importance of coming to grips with a concept of quality of life that is applicable to whole populations of people. In both cases, however, this reference serves as a point of departure for the development of different measures designed to determine accurately an individual or a population's quality of life.

In this thesis I take a critical look at these instruments—both those that consider quality of life to be idiosyncratic and those that take it to have a global meaning—from a philosophical point of view. Specifically I question the validity of our quality of life measures because, as I argue, they involve endemic problems with regard to how we are to understand the meaning of the questions and answers

---

[1] Aristotle (1999) (2nd ed) *Nicomachean Ethics*, trans by Terence Irwin, Indianapolis: Hackett Publishing Company, 15-20.

that make up the content of these measures. My aim here is not to solve this problem by implementing a new and better methodology. Rather I want to dissolve the problem by changing how we think about quality of life measures and how we understand the kind of information they provide.

Traditionally quality of life measures are understood along the lines of a sphigmometer or stethoscope: the measures' outcomes—the reading of my blood pressure or heartbeat—are understood to be independent from the theoretical orientation which one approaches a particular person's illness or medicine more generally. Moreover, the meanings of the measure's outcomes are meant to be clear to anyone educated in how to read them—that is they are not inherently vague or ambiguous. In fact as one researcher puts it,

> The measurement science behind quality of life assessment can ensure the collection of reproducible, substantive data that can be analyzed with as much confidence as a blood chemical value.[2]

In this thesis, however, I argue that the analysis of "quality of life" is perhaps not like the analysis of a blood chemical value. On the one hand, quality of life outcomes are significantly influenced by the theoretical orientation with which we bring to bear on them; on the other hand, quality of life outcomes are inherently vague and ambiguous—even to those educated to interpret them. I began this introduction with the excerpt from the *Nicomachean Ethics* not only to introduce two different ways of thinking about quality of life and to emphasize their philosophical roots, but also to remind us that for Aristotle happiness or quality of

---

[2] Cella, D.F. (1992) 'Quality of Life: The Concept', in *Journal of Palliative Care*, 8: 10.

life is an ethical question. In this thesis I will argue that when we measure quality of life we are best understood as making a contribution to an ethical dialogue—not measuring a scientific construct.

Health-related Quality of Life Measures: Its Influences and Uses

Typically when we think of ethical questions we think of questions regarding what it is to live a good life and how we might go about achieving a good life for ourselves and others. These kinds of questions first began to impinge significantly on the practice of medicine in the years following World War II. As I will discuss in Chapter 2 these questions were precipitated in part by the growing awareness of the experimental research that was conducted on human subjects during the war. In part, however, the questions were the results of advances in technology which increased the kinds and numbers of diseases that medics were expected to treat.

Before the Second World War most of our medical endeavors dealt with acute infectious diseases; diseases like tuberculosis or polio from which individuals either recovered or died. But after the war, at least in the northern hemisphere, the majority of medical practice began to shift to the management and care of chronic disease.[3] This trend continues today: diseases like AIDS which 20 years ago killed most of the individuals that contracted it are now manageable with the appropriate cocktail of drugs; the same is true for tuberculosis, many forms of cancer, diseases that cause renal failure and others. As a result of this shift in medicine individuals

---

[3] Ibid, p. 8.

can live longer with a variety of illnesses. But many of the interventions responsible for this change are either temporarily or permanently toxic and debilitating. Thus questions regarding the relative *quality* of one's life have become a significant factor not only in decisions regarding whether or not to accept certain treatments, but also in the characterization of successful treatments.

The notion of quality of life research was born in 1947 when the World Health Organization (WHO) redefined "health" as, 'a state of complete physical, mental and social well-being and not merely the absence of disease and infirmity'.[4] Two years later Karnofsky and Burchenal included mood, well-being and functional status for the first time as criteria by which new drugs ought to be evaluated.[5] Quality of life research has since progressed into a major research program and quality of life measures are used in a variety of contexts. In what follows I look at some of the areas in which quality of life measures are frequently used.

The most common use of quality of life measures is in the evaluation of successful or effective treatments.[6] In these cases quality of life outcomes from different treatments are compared—alongside biomedical outcomes—to determine which intervention(s) best serves patients' interests. For instance, the United States Food and Drug Administration (FDA) routinely uses quality of life measures or what they refer to as patient-reported outcomes as effectiveness endpoints in clinical

---

[4] World Health Organization Constitution (1947) *WHO Chronicle*, 1: 16.

[5] Browne, J. (1997) *Individual Quality of Life in Older People, Conceptual and Methodological Challenges*, PhD Thesis in Psychology, Dublin University p. 3.

[6] Ibid, p. 12.

trials.[7] Anticancer agents can obtain FDA approval even without an advantage in survival compared to the standard therapy as long as the new drug demonstrates an improvement in quality of life.[8]

In addition to its use in evaluating the success of different drugs, the comparison of quality of life outcomes is also used to determine the relative success of different surgical procedures. For example, there are at least two different procedures which aim to reduce the chances of a sub-arachnoid hemorrhage in a weak artery. One procedure requires a surgeon to enter the body through the skull and clamp the weakened artery; the other procedure uses a coil threaded through the nose to brace the walls of the artery. For the purposes of funding or guideline development we may wish to determine which intervention constitutes best practice. In this case both surgeries are effective, but because one surgery is more invasive and the other more experimental quality of life is a relevant concern. A third use for the comparison of quality of life outcomes concerns the effectiveness of elective surgeries like knee replacements and cochlear implants. Since elective surgeries are essentially a quality improvement issue these measures are allegedly suited to evaluate their success.

In addition to the evaluation of drugs and surgical interventions quality of life measures are also used in routine clinical practice to help individuals make

---

[7] Office of New Drugs and the Office of Medical Policy in the Center for Drug Evaluation and Research in cooperation with the Center for Biologics Evaluation and Research and the Center for devices and Radiological Health at the Food and Drug Administration (2005) *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*, p. 1.

[8] Cella, D.F. (1992) 'Quality of Life: The Concept', in *Journal of Palliative Care*, 8: 8.

decisions about their treatment and to enhance doctor patient communication.[9] Here quality of life measures provide information to doctors about how their patients are doing and offer patients tools to help them articulate their concerns. This use of quality of life measures is in part a result of concerns individuals have raised about the quality of their lives given the side-effects their treatments can have. In part it is a result of changes in medical practice and decision-making which aim to give patients a voice in their own health care.

These two reasons for the growth of quality of life measures are not unrelated. As medicine increasingly focuses on the management of chronic illnesses, on one hand, and offers more and more elective interventions, on the other, ethical questions like those I discussed earlier become relevant to treatment decisions. In fact one may argue that issues of self-determination and patient autonomy are *the* motivating forces of quality of life measurement. I discuss this topic in more detail in Chapter 2.

A final use of quality of life measures is in the allocation of healthcare resources. The reasons for the need to ration healthcare resources are multiple: the organs required for transplantation are in short supply; only a few people in a local area may be trained in a relevant procedure; tertiary or intensive care beds in a nearby hospital may be limited; insurance coverage may be prohibited; experimental drug programs may allow only a fixed number of participants; hospital

---

[9] Browne, J. (1997) *Individual Quality of Life in Older People, Conceptual and Methodological Challenges,* PhD Thesis in Psychology, Dublin University p. 6.

budgets may require the elimination of certain services.[10] In many of these cases the need to ration healthcare resources is related to limited financial resources and the rising costs of healthcare. For instance, in May of 2001 a workshop on healthcare costs held by the United States Federal Agency for Healthcare Research and Quality (AHRQ) found that Medicare costs in California would rise from $252 billion in 2002 to $456 billion in 2010. Factors accounting for these rising costs included growth in pharmaceutical expenses, expensive new technologies, aging of the population and increased consumer demand.[11]

In light of limited financial and healthcare resources questions of resource allocation have mainly focused on deriving Quality Adjusted Life Years (QALYs) for different medical interventions. Here medical outcomes are assessed according to the number of years and the relative quality a given intervention adds to a particular life. One year lived in perfect health is worth one QALY unit and various methods are used to give percentage units to years in various states of less than perfect health. Once these percentages are calculated the cost of any medical intervention can be divided by its expected QALY increase to yield a cost per QALY.[12] In this way, different medical procedures are compared according to their cost-effectiveness.

---

[10] University of Washington School of Medicine (1998) 'Ethics in Medicine: Resource Allocation' at http://depts.washington.edu/bioethx/topics/resall.html

[11] Agency for Healthcare Research and Quality (2001) 'Healthcare Costs: Why Do They Increase? What can We Do?', United States Department for Health and Human Services, at http://www.ahrq.gov/news/ulp/costs/ulpcosts1.htm

[12] Edgar, A. *et al.*, (1998) *The Ethical QALY Ethical Issues in Healthcare Resource Allocations*, Haslemere: Euromed Communications, pp. 35-39.

But not all questions of resource allocation focus on QALYs. For instance, the assessment of rehabilitation needs is often achieved via quality of life inventories that help patients itemize their likely problems in coping with everyday tasks. This itemization can help clinicians with limited resources to promote the best quality of care for their patients while giving them the skills to participate in their daily activities.[13]

Quality of Life and the Importance of Philosophical Reflection

Although quality of life measures are used in a variety of contexts and for a variety of reasons each measure faces the same criteria for determining their relative ability to measure quality of life. As I will discuss further in Chapter 1 quality of life measures are usually evaluated in terms of what are called their psychometric properties, namely their reliability and validity. These criteria are taken from classical test theory and are a measure's two most important evaluative properties.

The quality of life literature currently overflows with examinations of the degree to which quality of life measures achieve these standards. Researchers investigate different reliability coefficients and different strategies for achieving validity. They also offer advice on constructing a measure in terms of the choice of questions as well as in terms of their format and their weighting. Nevertheless, there is a little of what we might refer to as philosophical reflection on the way standards of reliability and validity are applied. By a lack of "philosophical reflection" I mean that the methodological portion of the quality of life literature is

---

[13] Cella, D.F. (1992) 'Quality of Life: The Concept', in *Journal of Palliative Care*, 8: 12.

almost entirely concerned with *how* to construct measures whose psychometric properties conform to the standards of classical test theory. Nevertheless, there is very little discussion regarding the application of these standards with regard to the purposes for which we measure quality of life. Nor is there sufficient discussion of the different meanings of reliability and validity in different areas of investigation.

It is understandable that a discipline whose function is to gather and disseminate information about quality of life would focus, methodologically speaking, on questions of how to engineer valid and reliable measures rather than on questions regarding the standards of reliability and validity themselves. As a result, however, engineering difficulties with the validity or reliability of a measure tend to be met with technical solutions which in turn simply assume the suitability of applying classical test theory standards.

I explain in Chapter 1 that the accuracy of a given quality of life measure is generally defined in terms of construct validity and that this definition is rarely questioned. Instead it is often assumed that if a measure achieves positive construct validity and it is reliable, then it is possible to analyze a measure's outcomes as clearly as we do 'a blood chemical value'. Just as a valid and reliable sphigmometer can answer the question, 'What is her blood pressure?' a valid and reliable quality of life measure is supposed to be able to answer the question, 'What is an individual or population's quality of life?' To be sure, a lack of what I am calling philosophical reflection and its consequences are only problematic if our process for determining the psychometric worth of our measures is problematic; if, for instance, some of the assumptions embodied in construct validity are inconsistent with the reality or purpose of quality of life measurement. In this thesis I make precisely this claim. I challenge some of the assumptions of construct validity with respect to quality of life measurement and thus I suggest that there is not only a role for philosophical reflection in quality of life research, but also a need for it.

The Limits of Construct Validity

As I said at the beginning of this introduction I question whether we can analyze quality of life outcomes as we may analyze blood chemistry. I focus on construct validity because of its importance in the literature. Yet, if the results of blood chemistry are clear to anyone trained in reading them, then the analysis of quality of life outcomes is not. In making this argument I articulate and challenge some of the fundamental assumptions that motivate the use of construct validity in quality of life research. I use my criticisms of construct validity to identify endemic problems with appealing to classical test theory to determine validity, problems which occur whenever we use the format of questions and answers to gather information.

In this thesis I identify and critique three assumptions built into the use of construct validity with respect to quality of life measures: 1) when a measure's outcomes confirm our hypotheses, we are warranted in having greater confidence in the accuracy of our theory and the validity of our measure; 2) respondents understand the questions and answers in our measures in the same way as researchers imagined they would; 3) respondents are conceptualized on the model of weak evaluators.

I begin my discussion of these assumptions in Chapter 2 where I argue that quality of life measures may perpetuate ethnocentric or biased notions of quality of life. I suggest that the application of construct validity sometimes creates a context in which certain assumptions about quality of life may go unchallenged.

Construct validity assesses the extent to which a new measure is related to hypotheses derived from the theory that underpins the measure.[14] For example, say we design a new measure to access the quality of life of cochlear implant recipients. We might theorize that individuals without certain opportunities—for instance, the ability to communicate orally—have a poorer quality of life than individuals with these opportunities. Based on this theory we may hypothesize that cochlear implant recipients will experience an increase in quality of life within a year after their surgery due to a growth in their valuable opportunities. To determine the construct validity of this new measure we would compare the outcomes from our measure with our hypothesis. If there is an inconsistency between our hypothesis and the outcomes, then the logic of construct validity instructs us to investigate our theory or our measure or perhaps both to locate the problem—perhaps some part of our theory is wrong or our measure needs some fine tuning. If, however, the outcomes confirm our hypothesis, then we are supposed to be justified in having greater confidence in both our theory and our measure.

---

[14] Albrecht, G.L. (1994) 'Subjective Health Assessment', Crispin Jenkinson ed, in *Measuring Health and Medical Outcomes*, London: London University Press, p. 21.

But in Chapter 2 I argue that this assumption—confirmation warrants confidence—may lead to ethnocentric understandings of quality of life. Turning again to the quality of life of cochlear implant recipients, if our measure's outcomes confirm our hypothesis, then the logic of construct validity warrants an increase in confidence that certain valuable opportunities such as the ability to communicate orally make for a better quality of life. Because oral communication presupposes the ability to hear our outcomes also support the view that hearing is a prerequisite for a good quality of life. Yet deaf advocates argue that the deaf life can be an equally good quality life and, moreover, that our measure's outcomes reflect both a century of systematic discrimination and the lack of public resources available to teach the deaf.

The logic of construct validity may unintentionally lead us endorse ethnocentric understandings of quality of life, for instance the notion that hearing is a prerequisite to a good quality life. To be sure, deaf advocates could be wrong, but without further investigation—even when our outcomes confirm our hypothesis—quality of life measures and their outcomes may conceal more about quality of life than they reveal.

In Chapter 3 I turn from issues of ethnocentricity and how we understand respondent answers to respondent bias and how individuals understand questions. Respondent bias occurs when individuals answer questions inaccurately. These inaccuracies sometimes arise because respondents understand the questions in a measure differently than researchers imagined they would understand them. When individuals understand questions differently than researchers their answers are in danger of biasing the measure. For instance, in the Nottingham Health Profile (NHP) one question asks respondents, 'I find it hard to reach for things'; 'yes/no'. As I will discuss in Chapter 3 this question is meant to be understood in terms of one's health, but respondents sometimes understand the question in terms of their height. Thus, short people may answer 'yes' indicating morbidity which may not exist.

Valid measures of quality of life require that researchers and respondents understand a measure's questions and answers in the same way. Indeed, the accuracy of our construct validity claims depends on it since differences in how respondents' understand questions cannot always be detected by construct validity: respondent bias can happen when respondents answer *in line* with established hypotheses, but in response to alternatively interpreted questions. To ensure the validity of our measures researchers strive to create questions that constrain how respondents will understand them.

Although valid measurement does require that researchers and respondents understand questions and answers in the same way I will look at research that suggests that even our most trusted measures do not achieve this mutual understanding. Moreover, I argue that attempting to constrain how respondents understand the questions in a measure is not the appropriate solution to this problem. It is not the appropriate solution because as I argue we cannot know in advance of applying a quality of life measure to a particular context just what the questions in the measure ought to mean; how researchers imagine questions is not necessarily the best way to understand them.

But if we cannot know in advance of administering a measure how to best understand the questions therein, then the validity of our measures cannot depend on conveying the meaning researchers imagined to respondents. Nonetheless, construct validity does depend on this conveyance of meaning. For if we rely on construct validity in spite of our inability to determine the meaning of questions in advance, then answers which appear to bias a measure may in fact be the result of alternatively interpreted questions which shed light on our understanding of the construct; answers which appear to confirm our hypotheses may, however, be the result of questions understood in such a way as to undermine our them.

Drawing on the work of Bas van Fraassen and Hans-Georg Gadamer in Chapter 4 I situate some of my concerns regarding quality of life research within well-known debates in the philosophy of science and social science. Here I characterize the futility and damage that follows from attempts to limit how respondents understand questions as a limitation that affects all investigations which rely on the structure of questions and answers. To do this I defend van Fraassen's position that we cannot limit the kinds of questions we ask from criticisms made by Wesley Salmon & Phillip Kitcher. Salmon & Kitcher worry that if we do not put a priori limits on the kinds of questions we might pose then we run the risk of adulterating scientific knowledge. I argue, however, that if we attempt to limit in advance the questions we ask—whether in quality of life research or science—then we unduly limit what we might come to know about our subject matter.

Turning to Gadamer I suggest that we can be critical of the questions we ask without limiting certain questions out of hand. To avail ourselves of Gadamer's insight I suggest that we think of that which we are questioning—a measure; an MRI—in terms of a text that we are trying to understand. For Gadamer we approach a text or text-analogue within a dialogic framework in which other texts on the subject matter—other measures or prior experiences with MRIs—gives us a starting point for understanding, thus we approach a text with certain assumptions about what it might mean. Our assumptions about what it might mean, however, are challenged when discrepancies arise between our assumptions and the text. These challenges take the form of questions, for instance, questions about an unusual marking on an MRI or uncommon answers to the questions in our quality of life measures. For Gadamer valid interpretations of a text are those that construe the text as coherent and possibly true, therefore good questions are those whose answers provide a coherent and truth-sensitive understanding of the text and poor questions are those that do not.

In Chapter 5 I return to my discussion of construct validity and here I develop one of the expectations which might buttress its use despite the difficulties I have raised with regard to it. To make my argument I discuss Charles Taylor's distinction between weak and strong evaluations. The distinction between these two types of evaluation turns on whether or not desires or outcomes are distinguished as to their qualitative worth. Weak evaluations deem that something is good just insofar as it is desired; strong evaluations determine that something is good insofar as the desire itself is worthy.

For Taylor these two different kinds of evaluation mark two different understandings of the human subject. Taylor argues that the human subject who is characterized by weak evaluations is unrealistic and ultimately untenable. I will suggest, however, that it is this conception of the human subject which lends support to the use of construct validity in quality of life research. Nonetheless, following Taylor, I too will argue for the inadequacy of this picture of persons and offer an alternative. My alternative is based on the human subject as characterized by strong evaluations. Strong evaluations characterize the self as dialogic; a self whose self-understanding and understanding of what counts as a good quality of life requires deliberation with others.

A Recommendation for Quality of Life Measurement

If, as I suggest in this thesis, the meaning of the questions and answers in our measures are always to a certain extent indeterminate and if construct validity cannot appreciate or arbitrate among the different ways we might understand the questions and answers in quality of life measures, then we cannot analyze quality of life outcomes as we might analyze 'a blood chemical value'. As a result I argue that we stop thinking of quality of life measures as independent and self-contained instruments whose outcomes ought to provide unambiguous answers to questions about quality of life. This is not to say that we should stop measuring quality of life or using quantitative measures, but rather when we do measure it we ought to understand what we are doing differently. We might conceptualize our measures not as a series of questions and answers which have but one correct meaning, but as contributions to a dialogue through which we can become clearer about what the questions and answers themselves mean; we might understand quality of life measurement not as determinate assessments of quality of life, but rather as tools for enhancing communication about it.

As tools for enhancing communication quality of life measures are conceptualized as one element in a framework of quality of life research in which other elements in this framework, for instance, large qualitative studies; individual interviews with respondents; histories of medicine; clinical assessments of good health and so on contribute to our understanding of how the outcomes of quality of life measures might be understood. For example, on this account no longer are qualitative studies which indicate the different ways in which respondents understand certain questions taken as evidence of a certain measure's invalidity. Rather this information is taken to supplement the information from the quantitative measure; to give us a better understanding of quality of life. To be sure, our quantitative measures of quality of life may change as a result of information from other studies—we may eliminate a question, add another—but these changes are not in the service of creating a measure whose outcomes are independently valid; whose outcomes can be clearly and accurately understood by anyone educated in how to read them.

My recommendation for quality of life research is that we understand measures of quality of life as in concert with other studies of quality of life; quantitative measurement is important, but these measures are not self-sufficient; we cannot understand the meaning of the outcomes from quantitative measures independent of a wider, dialogic context. This proposal is based on an argument that claims that our questions and answers about quality of life are inherently vague and ambiguous, that a definitive understanding of these questions and answers is not forthcoming. But my solution—that we understand quality of life measures in terms of a continuously evolving dialogue among different areas of research—is not an exact science. As I will discuss in Chapter 4 it is possible to give better or worse interpretations of our measures, but they are nevertheless *interpretations* and as such we can always learn to understand their meaning and thus the meaning of quality of life differently and more clearly.

But if my solution to the methodological problems of quality of life measurement is not an exact science, quality of life research is still a science, even it is not exact. How to implement my recommendations efficiently and practically is an importance question, but it is not the focus of this thesis. In this thesis I concentrate instead on explaining and developing some of the problems with using the standards of classical test theory to validate quality of life measures and thus provide good reasons for why we ought to understand these measures differently. Moreover, while other social science concepts may face similar difficulties to the ones that I raise for quality of life my arguments apply specifically to quality of life measurement and should not be taken more generally.

Chapter 1

Measuring Quality of Life: The Standard Needs and Individualized Approaches

Introduction

Health-related quality of life measures or as they are sometimes called "subjective health assessments" or "patient-reported outcomes" became important to health evaluation in the mid-1980's as a way to integrate patient perspectives into the decision-making process in a way that is both quantitative and efficient. The motivation for this was the recognition that clinical measures, based on physician or surgical assessment, or laboratory results, do not exhaust the ways in which one can or cannot be ill or unhealthy.[15] In fact it is well-known that clinical or biometric measures of health or illness often do not correlate well with patient-based health assessments. Due to technological advances and the democratization of medicine these differences in assessment matter, for no longer are health and illness considered merely biological concepts, but they also have an experiential dimension.

Although the importance of 'quality of life' in health care and the use of the term 'quality of life measure' are relatively new phenomena, measures which ask patients about their experience have been used in health care since at least 1949.[16] As the concept of 'quality of life' has gained currency, however, older measures, for example, measures previously known as 'health status measures' such as the Short-Form 36 (SF-36) and the Sickness Impact Profile (SIP) have been reconceptualized under the umbrella 'quality of

---

[15] Albrecht, G.L. (1994) 'Subjective Health Assessment' in *Measuring Health and Medical Outcomes*, Crispin Jenkinson ed, London; University College London Press, p. 15.

[16] Browne, J., *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 737.

life measure'.[17] Moreover, as the concept of quality of life has matured the variety of measures which fall under this category has grown. While on some accounts the sheer variety of quality of life measures speaks against the coherence of the concept[18], what quality of life measures do have in common is their attempt to measure subjective experience in the context of impairment, functional status, perceptions and social opportunities that are influenced by disease, injury, treatment or policy.[19]

But while quality of life measures aim to measure subjective experience, how they measure that experience differs. For this reason it is important to make a few preliminary distinctions. Firstly, quality of life measures are usually designed as utility or non-utility measures. Utility measures are always index measures; index measures reduce a patient's quality of life score to a single number and in the context of a utility measure this score represents a health state's utility. A score of 1 usually indicates perfect health and a score of 0 indicates a state equivalent to death. The utility values for each health state are derived by prior preference scoring from the general population using techniques such as the time trade-off and the standard gamble. The EuroQol (EQ-5D) is an example of a utility measure of quality of life which uses the time trade-off technique.[20]

---

[17] Hunt, S.M. 'The Problem of Quality of Life' in *Quality of Life Research*, 6: 206.

[18] Ibid.

[19] Albrecht, G.L. (1994) 'Subjective Health Assessment', in *Measuring Health and Medical Outcomes*, Crispin Jenkinson ed, London: University College London Press, p. 13. Quality of life measures are occasionally taken to have a more narrow scope referring only to those measures whose construct explicitly aims to capture 'quality of life'. Throughout this thesis, however, I will follow the convention of referring to any measure that aims to assess the impact of health and disease on subjective experience as a 'quality of life measure'. This is the convention of the International Society for Quality of Life Research (ISOQoL) and others.

[20] Brooks, R. *et al.* (1996) EuroQol: the Current State of Play, in *Health Policy*, 37: 53-72. The values derived from the EuroQol can be used to create Quality Adjusted Life Years (QALYs), but this requires an estimate of how long the status of one's health state as obtained from the EuroQol will last. If the EuroQol indicates an improvement of .10 after a surgical intervention and this improvement is thought to last for two years then the intervention provides .2 QALYs over two years.

Secondly, non-utility measures can be either profile or index measures. Unlike index measures profile measures provide multiple scores with which to represent a measure's construct. The construct of a measure represents whatever the instrument is designed to measure and in our context the construct broadly represents the impact of health and disease on subjective experience. In non-utility measures this construct is sometimes determined, through techniques like factor analysis, to consist of multiple independent dimensions or components. For example, if we are trying to measure anxiety we might take it to consist of a cognitive dimension, a behavioral dimension and a physiological dimension.[21] When a construct is understood to consist of multiple dimensions, then its measure is a profile measure, which provides a separate score for each dimension. Index measures, as we saw above, give us just one score and in the context of non-utility measures this score represents the value of the measure's construct under a single dimension or component.

Thirdly, non-utility measures have questions which are weighted, generally arbitrarily, according to researcher expertise and sometimes layperson input; they are not weighted according to techniques such as the time trade-off or the standard gamble listed above. For instance, researchers may determine weights based on their experience with a construct; individuals may be interviewed and their preferences for certain quality of life states averaged; or as is happening more and more questions may be given de facto

---

[21] Streiner, D.L. & Norman, G.R. (2003) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 267. The dimensions of a construct are determined using procedures like factor analysis. See Ibid., pp. 265-9.

weights.[22] The scores from non-utility measures represent the severity of whatever dimension they are meant to measure, but unlike utility measures which reduce all constructs to a single metric of utility, non-utility measures have their own unique metrics.

Finally, utility measures of quality of life have typically preoccupied moral philosophers because of the role they often play in cost-effectiveness and the allocation of scare health resources. Utility measures tend to presuppose the theory of expected utility and begin their inquiry with the question, 'How much utility is a particular health state worth?' Non-utility measures, however, are less directly involved in resource allocation and more directly involved in giving us a picture of patient experience. Although non-utility measures do aim to measure 'quality of life' as I will discuss throughout this thesis, what 'quality of life' amounts to—how we should understand it—varies among these measures. Unlike utility measures we might say that non-utility measures lack a mature theory.

In this thesis I take up the issue of non-utility measures. I do this because they open up inquiry into fundamental questions such as 'What is quality of life?' and thus allow us to evaluate different answers to this question. In the first half of this chapter I begin to distinguish two different answers with respect to two different approaches to quality of life: the standard needs approach and the individualized approach. Although both aim to measure the impact of health and illness on subjective experience and both do this in virtue of a non-utility metric they differ in how they characterize quality of life. For proponents

---

[22] Ibid, pp. 102-3; Browne, J., *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 739; Jenkinson, C. (1991) 'Why are we Weighting? A Critical Examination of the Use of Item Weights in a Health Status Measure', in *Social Science Medicine*, 32: 1413-16.

of the individualized approach quality of life is whatever an individual says it is; for proponents of the standard needs approach quality of life is more uniform.

In the second half of this chapter I turn from these different approaches to quality of life to examine the processes by which we typically assess the extent to which non-utility measures of quality of life *do* give us a good answer to the question, 'What is quality of life?' To do this I look at the processes by which we determine the validity of our quality of life measures—do they measure what they purport to measure?

The validation process for quality of life measures, however, is not straightforward. Because quality of life measures aim to assess subjective experience the phenomena that they all attempt to capture is, broadly speaking, unobservable. This means that quality of life measures lack evidence independent of their measures with which to compare their results. Put differently, quality of life measures lack a gold standard. In lieu of a gold standard quality of life measures must be validated using other criteria. Typically this is done using criteria derived from classical test theory, namely content, criterion and construct validity.[23] In this chapter I will argue that the validity of quality of life measures stands or falls with positive construct validity. In the chapters to come the singular importance of construct validity will have important consequences for how we understand—or misunderstand—good quality of life.

---

[23] Streiner, D.L. & Norman, G.R. (2003) Health Measurement Scales a Practical Guide to their Development and Use, Oxford: Oxford University Press, p. 174.

I

Differences between the Standard Needs and Individualized Approaches

A. The Standard Needs Approach to Quality of Life

The standard needs approach and the individualized approach are the two main orientations to quality of life assessment.[24] But the standard needs approach is what we might call the primary approach to quality of life since it is more commonly used and more widely known. The standard needs approach represents an assortment of measurements which differ greatly in their content—the kinds of questions they ask—and their scaling—the format chosen for their questions and answers. In fact they differ so much that until the introduction of the individualized approach in the 1990's there was no recognized term or theoretical orientation with which to summarize or unify the characteristics embodied by these measures. As one researcher put it in 1995, '...the theoretical status of the QOL [quality of life] concept remains ambiguous.'[25] Even now the term 'standard needs' is the name given by proponents of the individualized approach to mark a contrast with their own. Nonetheless the term is useful not only because it identifies an important contrast, but also because it classifies a heterogeneous group of measures.

The term 'standard needs' refers to the fact that these measures are standardized. All quality of life measures following the standard needs model are written as questionnaires and share a similar form: they all have pre-determined, pre-weighted questions with a limited selection of potential answers. To be sure, as I said above these

---

[24] Browne, J.P. *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 737.
[25] Rosenberg, R. (1995) 'Health-Related Quality of Life between Naturalism and Hermeneutics', in *Social Science Medicine*, 41: 1411.

measures do vary in terms of content and scaling. For example, one measure may have questions that try to elicit information on physical health, mental health, and social well-being, while another measure may only focus on mental health or emotional well-being. Some measures like the Sickness Impact Profile (SIP), which I discuss in the next chapter, may ask mainly functional questions about one's ability to perform certain tasks while measures like the Short-Form 36 may include functional questions as well as questions regarding one's personal satisfaction or overall happiness.

Moreover, when trying to illicit information on the *same* aspect of quality of life measures may also differ in their scaling or format. For instance, one measure may ask respondents whether or not they can walk up one flight of stairs without shortness of breath and the available answers for this question may be yes/no. Another measure may be interested in the same topic, namely shortness of breath with respect to exertion, but instead of a yes/no format this measure might provide respondents with a vertical line on which to plot the ease with which they can walk up stairs.

There are hundreds of different standardized quality of life measures varying in content and scaling each purporting to measure different aspects of health and disease. Nonetheless, these measures are grouped into two basic categories: generic measures and specific measures. Some measures like the Nottingham Health Profile (NHP), which I discuss in Chapter 3 & 5 claim to measure *general* health across disease states; measures such as the NHP are generic measures applicable to anyone. Other quality of life measures, however, are specifically calibrated for a particular disease or population. One example of a disease-specific measure is the Patient Quality of Life Form (PQLF), which aims to assess the difficulties thought to affect deaf people. Some of the questions in this measure include how communication and hearing affects life, issues of isolation and one's dependence on others.[26]

Because there are so many ways that we can be ill and because there are so many ways that ill-health can affect our quality of life it may not be surprising that there are a substantial and growing number of standardized quality of life measures each of which attempt to capture a unique facet of this phenomenon. The sheer number of standardized measures is usually explained by the multidimensionality of the quality of life construct—there are many potential components that affect quality of life. But because as I said earlier these measures traditionally lack an explicit theoretical orientation critics have sometimes suggested that standardized quality of life measures represent an 'anything goes' approach.[27] In what follows, however, the contrast with the individualized approach suggests that standardized measures of quality of life do share certain features which mark out the standard needs concept of quality of life from just any concept of quality of life. Moreover, in the next chapter I will examine a philosophical analysis of standardized measures, which suggests that these measures also share a robust theoretical orientation.

Despite the variety of standardized measures these measures nonetheless share the paradigm of standardization. Put differently, each of these measures is designed to deal with a particular cohort of people and within this cohort standardized measures assume that the same questions and weightings are relevant. Thus the standard needs model, to a certain extent, conceptualizes quality of life as a universal construct: respondents from a particular cohort answer the same questions and each receives a quality of life score

---

[26] Birger, M. *et al.* (2005) 'Cochlear Implants and Quality of Life: A Prospective Study', in *Ear & Hearing*, 26: 189.

[27] Hunt, S.M. (1997) 'The Problem of Quality of Life', in *Quality of Life Research*, 6: 206.

independent of their individual understanding regarding the worth of their lives. As one proponent of the individualized approach puts it,

> 'The central assumption is that a standard set of life circumstances are required for optimal functioning: quality of life does not, therefore, depend on a subjective viewpoint but is an objective characteristic of the person concerned.'[28]

For this author standardized measures take quality of life to be objective because these measures do not take seriously idiosyncratic understandings regarding the worth of individual lives. On this view objective contrasts with subjective as universal contrasts with individual. Nonetheless, we might still understand standardized measures as *subjective* assessments if we expand our notion of subjectivity. On the standard needs model of quality of life subjectivity is understood in terms of what I will call 'embodiment' and *not* in terms of individual, idiosyncratic experience. On this view objective contrasts with subjective as observable contrasts with phenomenal.

The idea of subjectivity as embodiment is that we are all *in* similar bodies and measuring quality of life consists in measuring the similar experiences that result from our being in similar bodies. To explain, we might say that it is in virtue of being in a body that we have certain unobservable experiences like pain and well-being. Moreover, because we are all in similar bodies we all have these same kinds of similar unobservable experiences. When we experience the phenomena under investigation the only variables are meant to be how intensely we experience it. Thus quality of life measures can be standardized; we can use the same questions, the same scaling and response options for the same cohort of people. This particular understanding of quality of life differs from the individualized

---

[28] Browne, J.P. *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 737.

approach which takes quality of life to be purely idiosyncratic; nevertheless we can still understand standardized measures as subjective assessments.

As I will further explore in Chapter 3 standardized quality of life measures are explicitly designed to avoid idiosyncratic responses to individual questions. Idiosyncratic responses are problematic for standardized measures because when respondents answer in ways unintended by the construction of the question researchers may misunderstand their answers and as we will see in Chapter 3 this threatens the validity of a measure's outcomes. The standardization of quality of life measures—the understanding of subjectivity as embodiment—requires that respondents falling within the measure's cohort understand the questions in the same way.

B. The Individualized Approach to Quality of Life

Unlike proponents of the standard needs approach who design their measures in order to evade idiosyncrasy, proponents of the individualized approach embrace idiosyncratic responses. Because the individualized approach conceptualizes quality of life as purely idiosyncratic the validity of *these* measures requires that individuals provide personalized responses to questions about quality of life. I will have more to say about the validity of the individualized approach in Chapter 2, but it is important to note that on this view subjectivity and idiosyncrasy are synonymous. Their respective characterization of subjectivity is the fundamental difference between the individual and standard needs approaches to quality of life.

Unlike the standard needs model with its myriad of measures and its uncertain theoretical foundation the individualized approach is represented by relatively few measures and a relatively explicit theory. The individualized approach to quality of life begins from the perspective that quality of life is whatever an individual says it is.[29] This position is justified with reference to topics within phenomenology whereby phenomenology is taken to be the philosophy of individualism.[30] As one proponent puts it,

> 'Taken from a phenomenological perspective, every person is indeed an "island" and quality of life is inherently individual. Who else can experience the *n-of-*1 trial that life represents but the individual herself or himself? Who else can report the deepest aspirations and fears, the fondest goals and fulfillment, the meaning of the winnings and losings? These personal definitions and reports are individual quality of life, and they remain relentlessly personal and individual'[31]

According to the individualized approach we can only understand what counts as a good quality of life insofar as we understand what a good quality of life is for a particular person. To achieve this understanding and to evaluate quality of life we must develop measures which provide the maximum amount of latitude for individuals to describe and assess what is important to *them.*

Although proponents of the individualized approach claim phenomenology as their philosophical foundation they take their position to reflect the appropriate way to show respect for individual self-determination and the sacredness of human life.[32] On this view the only way to show regard for different ways of life and different understandings of the

---

[29] Hickey, A., *et al.* (1999) 'The Schedule for the Evaluation of Individual Quality of Life', in *Individual Quality of Life Approaches to Conceptualisation and Assessment,* C.R.B. Joyce, *et al.* eds., Australia: Harwood Academic Publishers, p. 121.

[30] Although phenomenology is taken, at least in part, to justify the individualized approach whether phenomenology *does* justify it is debatable. Phenomenology represents a large and diverse area of philosophy including different authors and different types of phenomenology. Proponents of the individualized approach do not specify whose phenomenology they take to support their theory and in light of this it is difficult to evaluate their claim. What is clear, however, is that for some so-called phenomenologists their support of the individualist position is questionable.

[31] Joyce, C.B.R., *et al.* (1999) Forward, *Quality of Life in Individual Quality of Life Approaches to Conceptualisation and Assessment,* Australia: Harwood Academic Publishers, pp. ix-x.

[32] Ibid.

'good' is to elicit the value system of the respondent and to quantify quality of life using this system.[33] This preoccupation with individual self-determination suggests that individualized quality of life not only rests on a particular understanding of phenomenology, but also a certain interpretation of liberal individualism.

Indeed individualized quality of life is an approach born in light of what proponents consider to be systematic failures of the standard needs approach to accommodate different understandings of the good and thus the failure to value sufficiently self-determination. According to proponents of the individualized approach there are two problems with standardized measures both of which unduly influence what can count as a good quality life. First, the dimensions into which the quality of life construct is deconstructed and which represent the sub-categories under which questions are posed are pre-determined on the standard needs approach. For proponents of individualized quality of life these pre-determined dimensions represent unwarranted assumptions as to the most important determinants of quality of life. They point out that these dimensions, dimensions such as dependence, family life and communication may not represent those areas of one's life which are most important or in fact areas that are even applicable to everyone in the cohort.[34] On this view measuring quality of life based on assumptions about what constitutes a good life bias standardized measures toward certain kinds of lives at the expense of others.

---

[33] Browne, J.P. *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 742.

[34] Hickey, A., *et al.* (1999) 'The Schedule for the Evaluation of Individual Quality of Life', in C.R.B. Joyce, *et al.* eds., *Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers, pp. 119-20.

Secondly, proponents of the individualized approach criticize the scoring techniques of the standard needs approach arguing that pre-determined weightings are biased. For example, measures whose weightings come from averaging the preferences of sample populations ignore the inter-individual variability which exists if individuals are asked to weight these items themselves.[35] For example, the Nottingham Health Profile (NHP) gives the weight of 39.2 to the question 'I'm tired all the time' if it is answered 'yes' whereas the question 'I soon run out of energy' contributes a weight of 24 if answered 'yes'.[36] These standard weightings contribute to the systematic bias for certain kinds of lives over others since someone answering yes to the above questions may not find these items significant contributors to poor quality of life. Moreover, even measures which carry de facto weights also indirectly affect quality of life scores because the number of questions contributing to the dimensions differ and thus indirectly weight phenomena differently.[37]

In order to redress the bias for certain kinds of lives inherent in standardized measures proponents of the individualized approach have developed their own quality of life measures. One of these measures is The Schedule for the Evaluation of Individual Quality of Life (SEIQoL). The SEIQoL is the first practical consequence of the individualized approach and is emblematic of these kinds of measures.[38] It is administered during an interview between the researcher and the respondent and in the first step of the

---

[35] Browne, J.P. *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 739-40.

[36] Ibid.

[37] Ibid. As I will discuss in Chapter 5 standard weights do not increase the sensitivity of quality of life measures like the NHP to changes in health state. See Jenkinson, C, *et al.* (1991) *International Journal of Health Sciences*, 2: 189-94.

[38] Hickey, A. *et al.* (1999) 'The Schedule for the Evaluation of Individual Quality of Life' in C.R.B. Joyce *et al.* eds., *Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers, pp. 120-21.

interview the respondent is asked to nominate five dimensions or components of life central to their quality of life and explain what they mean by each area.[39]

The second step of the interview is to determine how well each nominated dimension fares with respect to the respondent's current situation. For example, if socializing is one the nominated dimensions, but the respondent is currently bedridden this dimension's ranking may reflect this condition. Status is determined by providing respondents with vertical lines labeled at one end 'best possible' and at the other end 'worst possible'. Respondents are asked to imagine their best possible and their worst possible life with respect to each of their nominated dimensions. With these two mental anchors in place they then mark on the line where their current life falls with respect to each dimension. Respondents are also asked to mark on a horizontal line their current overall quality of life again imagining at one end the worst possible overall life and at the other end the best possible life. Researchers then transpose these ratings to a bar chart with a range of 1-100.[40]

The third step of the interview is to weight each nominated dimension. This step is meant to quantify the relative contribution of each area to the judgment of the individual's overall quality of life.[41] One increasingly popular way to do this is through a direct weighting method known as the SEIQoL-DW. The SEIQoL-DW uses a rotating, five-sliced pie chart to weight each dimension. Respondents are asked to manipulate the slices until each slice represents the particular distribution of importance of each dimension. On the back of the chart is an immovable disc which provides the increments by which each

---

[39] Ibid, p. 122.
[40] Ibid.
[41] Ibid, p. 123.

slice can be calculated.[42] Finally, scores are determined by multiplying each area weight by the corresponding self-rating as represented by the bar chart. Each of the five products can be summed to obtain an index score between 1-100.[43]

The standard needs approach and the individualized approach to quality of life provide two distinct answers to the question, 'What is quality of life?' On the one view quality of life is characterized, at least in part, to consist of unobservable, but nonetheless objective dimensions which exist in virtue of our embodiment. Thus quality of life measures can be standardized. On the other view quality of life is whatever an individual says it is; quality of life is idiosyncratic; it consists of unobservable and wholly subjective dimensions which exist in virtue of our individuality.

In the next section I turn to the processes by which we typically evaluate the accuracy or validity of our quality of life measures. A valid measure is an instrument that measures what it purports to measure and quality of life measures all purport to measure subjective experience in the context of different health states. As I will discuss in what follows if validity tests are successful, then over time they ought to provide evidence in favor of an answer to the question, 'What is quality of life?' and thus a mature theory for non-utility measurement.

---

[42] Browne, J. *et al.* (1997) 'Development of a Direct Weighting Procedure for Quality of Life Domains', in *Quality of Life Research*, 6: 301-10.

[43] The derivation of an index score allows us to compare SEIQoL results among respondents. But the real strength of the SEIQoL is not the index score, but its individual nature. Although scores from different respondents can be compared using the index scoring method this comparison hides the different weights and priorities of the participant. According to proponents of the individualized approach the advantage of individualized quality of life is not its comparability, but the ability to identify currently important areas of life with their relative weights. See Hickey, A. *et al.* (1999) 'The Schedule for the Evaluation of Individual Quality of Life' in C.R.B. Joyce *et al.* eds., *Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers, p. 128.

II

Validating Quality of Life Measures

As I discussed in the introduction to this chapter quality of life measures lack a gold standard; they lack an observable instance of their target construct. This predicament is in contrast to many clinical or biometric measures of health, for instance Magnetic Resonance Images (MRI) which at least in principle have observable occurrences against which we could test the accuracy of the MRI. In order to deal with the lack of a gold standard, quality of life researchers often use elements of classical test theory as a surrogate. Classical test theory understood as reliability and validity testing make up the psychometric properties of a quality of life measure. Together these two properties are meant to determine whether or not a measure fulfills its purpose; whether or not it reliably measures what it claims to measure. They are thus the critical apparatus of quality of life measurement telling us when a measure is 'good'; when we can reliably draw sound conclusions from the data. Although all quality of life measures are tested both for their reliability and validity, in this thesis I focus on the validity of quality of life measures.

Quality of life measures are typically tested in terms of what is sometimes referred to as internal and external validity. The former is comprised of both content and criterion validity and the latter consists of construct validity.[44] Content and criterion validity ascertain a measure's accuracy with respect to standards supposedly internal to the discipline at hand. For instance, content validity is an official peer-reviewed inspection to see if the measure in question is representative of its target construct. This involves a

---

[44] Streiner, D.L. & Norman, G.R. (2003) (3rd ed.) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 174.

formal review of the questions in the measure to see if they cover all aspects of the item we are trying to measure. Criterion validity, on the other hand, compares a new measure to an already established measure with a similar construct. If the new measure correlates appropriately with the established measure, than it is said to have good criterion validity.

Finally, construct validity assesses a measure's accuracy with respect to how well a measure's outcomes confirm hypotheses derived from the theory that underpins the measure. But as I will discuss below if the outcomes do not confirm the hypotheses, then the measure is not necessarily inaccurate for it might be the theory and not the measure which is wrong. In any case construct validity is important because theoretically it does not require a new measure to adhere to an already established internal standard, rather it tests to see if a new measure reacts as predicted and if it does not, then it is possible to learn something about our measure as well as our theory. Moreover, the hypotheses we make regarding the behavior of a measure can be drawn from multiple disciplines—from clinical diagnoses; economics; feminist theory and so on—thus construct validity is a test of how a measure relates to criteria external to the field in which it was created.

Although both standardized and individualized measures are tested for all three types of validity in some cases certain kinds of validity are taken to be more significant than the others. For example, content validity is usually understood as an inspection of the dimensions and questions in the measure to determine if they are representative of the target construct.[45] Content validity is meant to increase the number of valid inferences we

---

[45] Ibid, p. 119.

can draw about a person responding to questions in the measure because it is supposed to ensure that we have accounted for all aspects of the construct under investigation.[46]

Because the individualized approach to quality of life starts from the perspective that quality of life is whatever an individual says it is, the content validity of these measures is taken by developers of the SEIQoL and others to be good by definition.[47] Since no one but the respondent can tell us what quality of life amounts to the direct assessment of the content is thought to be unnecessary. To be sure respondents may misrepresent their view of quality of life, for instance moods, the accessibility of certain memories and information, as well as the time given to answer a question are all well-documented factors that can bias individual reports. To the extent that this happens the SEIQoL may be invalid. But given the idiosyncratic nature of quality of life proponents of the SEIQoL suggest that as long as individualized measures are reasonably reliable—individual respondents tend to nominate the same dimensions over time—we must trust the accuracy of what respondents report as most important to them.[48] As one proponent of individualized quality of life puts it, '...it must be borne in mind that the individual still remains the person with best access to their own quality of life...'[49]

---

[46] Ibid, p. 175.

[47] Ruta, D.A. & Garratt, A.M. (1994) 'Health Status to Quality of Life Measurement', in Crispin Jenkinson ed, *Measuring Health and Medical Outcomes*, London: University College Press, p. 147.

[48] The stability of domain elicitation in the SEIQoL was examined in a study of quality of life following hip replacement. The mean number of domain changes was 1.1 over 7.5 months and 1.3 over 24 months suggesting that the domains individuals take to be most important to their quality of life remain fairly constant over at least a 2 year period. See O'Boyle *et al.* (1992) 'Individual Quality of Life in Patients Undergoing Hip Replacement', in Lancet, 339: 1088-1091.

[49] Browne, J. (1997) *Individual Quality of Life in Older People, Conceptual and Methodological Challenges*, PhD Thesis in Psychology, Dublin University p. 242. Although proponents of the SEIQoL recommend that we *trust* the accuracy of individual reports regarding the *content* of the nominated dimensions the original version of the SEIQoL employed judgment analysis in order to ascertain respondents' implicit *weights* and compare these to the ones they actually reported. To determine a respondent's implicit weights a series of 30 cases representing variable health states are presented to the respondent and she is asked to rate the overall

The assessment of content validity, however, *is* central for the development of standardized measures. If, for instance, we intend to measure the quality of life of prelingually deafened individuals using pre-determined dimensions and questions, then we need to be sure that the dimensions we select and the questions we ask are indeed relevant and, moreover, that we have not overlooked anything.

The process of evaluating content validity typically includes three steps. The first is to elucidate the theory that underpins the measure's construct in order to verify which dimensions of the construct are important to the measure's purpose and which are irrelevant. This first step is accomplished by gathering knowledge of research previously executed within the relevant area of interest and completing an inventory of previous instruments designed to measure similar constructs. This review of the literature constitutes the second step in the process of content validity. Finally in the third step a preliminary instrument is presented to a variety of people for peer-review. These people might include individuals from the measure's target cohort, for example prelingually deafened individuals, as well as doctors, surgeons and other professionals such as epidemiologists and psychologists.[50]

---

quality of life for each case on a horizontal visual analogue scale. The relative weight assigned by the individual to different aspects of health embodied in the cases is calculated using a multiple regression analysis designed specifically for the SEIQoL. Nonetheless, this version of the SEIQoL which employs judgment analysis has gradually been replaced by the SEIQoL-DW which, as I discussed earlier, does not seek to determine implicit weights and rather uses the direct weightings that respondents report. See Hickey, A. *et al.* (1999) 'The Schedule for the Evaluation of Individual Quality of Life' in C.R.B. Joyce *et al.* eds., *Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers, pp. 121-4; Browne, J. *et al.* (1997) 'Development of a Direct Weighting Procedure for Quality of Life Domains', in *Quality of Life Research*, 6: 301-10.

[50] Streiner, D.L. & Norman, G.R. (2003) (3rd ed.) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 21.

Content validity, unlike criterion and construct validity, is often executed as part of the initial construction of a quality of life measure.[51] Researchers designing standardized measures look to maximize the content validity of their measures before they test it on sample populations. Even with respect to individualized measures the nomination of dimensions is part of the attempt to construct a respondent's value system and not part of the assessment of one's quality of life. Criterion and construct validity, on the other hand, require for their purposes the information from the outcomes data of sample populations. Thus these two kinds of validity are implemented only after a measure's construction is complete and they test whether the construction of the measure was successful.

Criterion validity is the validity of the gold standard.[52] In areas of measurement where a gold standard is available criterion validity tests the accuracy of the measure's outcomes against criteria external to the measure. For example, the criterion test of an MRI showing, say a torn Anterior Cruciate Ligament (ACL) would be the actual torn ACL. As I have said, however, quality of life measures lack a gold standard; in this context criterion validity requires the use of proxy standards. Proxy standards are usually other well-established quality of life or clinical measures.[53] Proxy standards act as gold standards in that if the new measure does not correlate with the established measure as expected, then the validity of the new measure is put into doubt.

---

[51] See Ibid, pp. 19-22.
[52] Ibid, p. 176.
[53] Ibid, p. 178.

One type of criterion validity is called concurrent criterion validity.[54] Concurrent criterion validity is used to correlate a new measure with an established measure whose construct is identical to the instrument under investigation.[55] For example, occasionally lengthy quality of life measures are developed which are not practical to administer to populations of people because they are time-consuming to complete. One example is the original Sickness Impact Profile (SIP) which had 149 questions.[56] In cases like the original SIP it is desirable to create a measure that provides the same essential information, but in a shorter and more practical version. To create the shorter version we can edit the longer version until it reaches a more convenient length. In doing this, however, it is important that questions essential to the measure's construct are not removed; it is important that the measure's content validity remains high. Using criterion validity we can test the content validity of the shorter measure by treating the longer measure as the criterion or gold standard. To achieve this validation both questionnaires are given to the same cohort of people at similar times. If the data from the shorter version correlates highly with the longer one, then we can feel confident that the shorter version measures the same thing as the longer version.

---

[54] Albrecht, G.L. (1994) 'Subjective Health Assessment', in Crispin Jenkinson ed, *Measuring Health and Medical Outcomes*, London: UCL Press, p. 21. There is another type of criterion validity which I will not discuss. It is called predictive validity. This type of criterion validity is most useful for the validation of intelligence tests. For example, tests meant to predict intelligence may be administered to students leaving secondary school. The results of these tests may then be compared to the students' actual achievements on leaving university four years later. Their achievements on leaving university act as the standard against which the validity of the original test is determined. See Streiner, D.L. & Norman, G.R. (2003) (3rd ed.) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 176.

[55] Streiner, D.L. & Norman, G.R. (2003) (3rd ed.) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 176.

[56] Wright, L. (1994) 'The Long and the Short of it: the Development of the SF-36 General Health Survey', in Crispin Jenkinson ed, *Measuring Health and Medical Outcomes*, London: University College London Press, p. 91.

But in the majority of cases when researchers create new measures they are not merely shortening already existing instruments, but developing genuinely new ones. In these latter cases a perfect criterion measure is unavailable. In this context it is possible, however, to test the new measure against established measures that are *similar* to it in terms of the construct or dimensions under investigation. What counts as a similar measure is at the discretion of those validating the new measure; sometimes, however, the relevant similarity is located in clinical or physician-based assessments as opposed to established quality of life measures.[57] This kind of validity is sometimes called convergent validity.[58]

Because a new measure is never identical to an established measure and because the relevant similarity between two such measures will vary, the extent to which the measures are expected to correlate will also vary. Thus to test a genuinely new measure for criterion validity researchers must develop hypotheses about how a new measure and its criterion measure will relate. In this context the assessment of good criterion validity requires that the correlation between the new measure and the established measure correspond to the hypothesis made about how the two measures should relate.[59] As I will discuss shortly there is good reason to think that this aspect of criterion validity is an extension of construct validity.

Construct validity unlike criterion validity does not require a gold standard or a proxy standard against which to compare the accuracy of a new measure. It does, however,

---

[57] Bowling, A. (1991) *Measuring Health A Review of Quality of Life Measurement Scales*, Milton Keynes: Open University Press, p. 172.

[58] Streiner, D.L. & Norman, G.R. (2003) (3rd ed.) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 183.

[59] Albrecht, G.L. (1994) 'Subjective Health Assessment', in Crispin Jenkinson ed, *Measuring Health and Medical Outcomes*, London: University College London Press, p. 21.

require a theory from which to draw hypotheses regarding the measure's outcomes.[60] For example, such a hypothesis might anticipate that individuals scoring high on a certain measure will have certain attributes and also differ from those with low scores in characteristic ways. Construct validity uses hypotheses like this to assess the extent to which a new measure is related to criteria derived from the theory that underpins a measure. To make this assessment it compares a new measure's outcomes data to the predictions or hypotheses made on behalf of the new measure and its guiding theoretical orientation.[61] If a measure has good reliability and if the predictions are confirmed in its outcomes, then greater confidence in the validity of the measure *and* the validity of the theory is justified.[62]

It is now possible to see how determining the criterion validity of a new measure with respect to its similarity to an established measure is a form of construct validity. As I discussed previously the criterion validity of genuinely new measures requires the development of hypotheses which indicate the way that certain established measures are similar to a new measure. And as I discussed with respect to content validity the identification of a new measure as *similar* to another is due to the theory which shapes how we understand the new measure's construct. Thus the hypotheses used to indicate criterion validity are derived from the theoretical orientation of the measure: if the correlation between the two measures bears out the hypothesis then greater confidence in the new measure's validity is warranted.

---

[60] Streiner, D.L. & Norman, G.R. (2003) (3rd ed.) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 181.
[61] Ibid, pp. 179-80.
[62] Ibid, p. 181.

But in fact it is not just criterion validity which manifests as a form of construct validity; content validity too depends on it. For as I also discussed earlier content validity requires an understanding of a new measure's theoretical orientation before one can verify which dimensions and questions ought to be included in the measure. When developing a new measure the decisions over what to include and what to exclude are, to a certain extent, based on hypotheses about what affects the quality of life of individuals within a certain cohort. Whether these hypotheses are accurate—whether a measure has good content validity—is determined with respect to positive criterion and construct validity. But as we have seen criterion validity of genuinely new measures is a form of construct validity. Thus to quote one researcher, 'All validity is at its base some form of construct validity...It is the basic meaning of validity'[63]

Construct validity is important not only because it supports claims regarding the validity of measures which lack a gold standard, but also because it is seen as substantiating a theory of quality of life; it is seen as providing an answer to the question 'What is quality of life?' It is important to emphasize, however, that no single confirming instance of one's hypothesis can prove a theoretical orientation and even one unconfirmed instance *can* put the theory into doubt. Construct validity, we might say, is an on-going process in which we continuously learn more about our theory, make new predictions and then test them.[64]

But because construct validity allows us to test at the same time both the validity of the theory and the validity of the new measure the source of trouble resulting from unconfirmed hypotheses is ambiguous. If our hypotheses are not confirmed in the

---

[63] Guion, R.M. (1977) 'Content Validity: Three Years of Talk—What's the Action?, in *Public Personnel Management*, 6: 407-414.
[64] Ibid. p. 180.

outcomes of the new measure, then one of *three* things is possible: the theory may be flawed; the measure may be wrong; or both could be inaccurate.[65] Although there is no way of knowing what is wrong without conducting further studies[66] once the problem is determined either the theory or the measure or both must be reconstructed.

As we saw in Section 1 of this chapter different measures of quality of life draw on different theoretical orientations and thus provide different answers to the question 'What is quality of life?': the standard needs approach characterizes quality of life in terms of what I have called 'embodiment'; the individualized approach understands it as idiosyncratic. If construct validity worked in the way that classical test theory describes it—we reject or reconstruct theories or measures or both when we fail to get confirming instances of our hypotheses—then it might be possible to capitalize on the work going on within these different orientations and eventually converge on a theory that accurately describes quality of life as well as a set of measures that quantify it.

But the logic of construct validity whereby unconfirmed predictions require the rejection or reconstruction of the measure or the theory or both is not always practiced. In fact when reviewing the psychometric properties of quality of life measures poor construct validity appears to be extremely rare.[67] Instead almost every measure—both measures within the standard needs and individualized approaches—report good construct validity. But significantly these reports of good validity are not always the result of confirming instances of hypothetical claims; rather hypotheses *not* born out in the outcomes are *still*

---

[65] Ibid.

[66] Ibid.

[67] See Bowling, A. (1991) *Measuring Health A Review of Quality of Life Measurement Scales*, Milton Keynes: Open University Press.

understood to support the measure's validity. This is usually due to auxiliary hypotheses that are brought in to explain the deviation between the original hypothesis and the outcomes data. For instance, when Spitzer's Quality of Life Index (QL), a standardized measure, did not correlate as highly as expected with the Karnofsky Performance Scale, another standardized measure, it was suggested that the QL's highly multidimensional nature, as opposed to Karnofsky's focus on physical functioning, was the reason.[68]

Moreover, one hypothesis based in the individualized approach claims that individualized measures will report low discriminative ability.[69] Put differently, the scores from individualized measures are unlikely to discriminate between, for example, individuals with different kinds or severity of disease. One reason for this hypothesis is founded on the idea that when individuals choose the dimensions of their life that represent areas of importance different people will choose different things and these choices may not be related to their illness. Furthermore, even if individuals do choose dimensions which vary with a particular illness the weighting and status of the items may be similar thus resulting in comparable index scores for the two illnesses.[70]

But in fact the SEIQoL does report the ability to discriminate between patients suffering from Peptic Ulcer Disease (PUD) and those suffering from Irritable Bowel Syndrome (IBS): in one study a group of IBS patients had significantly lower quality of life according to their index scores than did the PUD patient group. These results correspond not only to clinical perceptions that IBS disrupts quality of life to a greater degree than

---

[68] Ibid, pp. 86-87.

[69] Lacasse, Y., *et al.* (1999) 'Individualising Questionnaires', in C.R.B. Joyce, *et al.* eds., *Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers, pp. 98-99.

[70] Ibid.

PUD, but also correlates significantly with the NHP's findings within two of their dimensions.[71] These results perhaps suggest that at least in this case quality of life is not as idiosyncratic as proponents of the individualized approach suggest.

But proponents of individualized quality of life do not take this data to question their theory or their measure. Instead, on the one hand, they emphasize the individual differences that the SEIQoL does illuminate within the PUD and IBS patient groups—differences which clinical observations and the NHP cannot detect. For instance, the research team emphasize that the dimension of 'Health' was nominated less frequently by the PUD group than the IBS group. This is taken to suggest that the standard assumption that health is the most important component in a patient's quality of life may not be warranted.[72] On the other hand, proponents of individualized quality of life propose that the apparent similarities between individuals who *do* nominate the same dimensions—the IBS and PUD patient groups both nominated 'Family' as the most important life area—are irrelevant to general statements about quality of life since on their view what individuals understand family to *mean* is idiosyncratic.[73]

Rather than capitalizing on the work within different orientations to quality of life and thus building support for a mature theory and corresponding measures the validation of quality of life measures seems to support two different theoretical orientations—two different answers to the question 'What is quality of life?'—and hundreds of putatively valid measures. In Chapters 2 and 3 I continue to examine some of the practical

---

[71] Hickey, A., *et al* (1999) 'The Schedule for the Evaluation of Individual Quality of Life' in C.R.B. Joyce, *et al.* eds., *Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers, p. 125

[72] Ibid, p. 125.

[73] Personal correspondence.

consequences that come with the reliance on construct validity as the primary validation technique of quality of life measures.

As I indicated in the introduction to this thesis I will argue that construct validity and thus quantitative measures of question of life cannot provide us with independently valid outcomes; outcomes that can be understood clearly outside of a larger dialogic context. Rather we ought to conceptualize quality of life measures as one piece of a larger puzzle; as an important contribution to an ethical dialogue. In this chapter I identify one of the main questions which the on-going process of validation should be able to answer and which I will argue it cannot answer sufficiently, namely, 'What is quality of life?' In the next chapter I continue this argument and in doing so I focus on the theoretical orientation of the standard needs approach and explore some of the ways that this orientation as well as the individualized approach might negatively influence our understanding of quality of life.

Chapter 2

Construct Validity and Ethnocentricity

Introduction

As we saw in Chapter 1 construct validity is an important component to the validation process of health-related quality of life measures. In fact I argued in the previous chapter that it is the *most* important element. One of the reasons for its prominence is that it enables us to build up a theory while at the same time testing for the validity of our measure: if our hypotheses are born out in a measure's outcomes, then we can take this as evidence in favor of our theory and the validity of our measure. Thus construct validity requires a tentative theory—explicit or implicit—from which we can ground hypotheses and test the validity of our measures; our theory becomes less tentative as our hypotheses are better able to predict the outcomes of our measures.

Construct validity enables us to build a robust theory and test for the validity of our measures in part because it presupposes that the error in our theory or the inaccuracies in our measure will come to light as we make hypotheses and compare them to our outcomes. Inconsistencies between our hypotheses and our outcomes signal a problem with our theory, our measure or perhaps both. By investigating the origin and nature of such an inconsistency we further develop our theory and the ability of our measure to account for its construct. In the previous chapter, however, I discussed a couple of examples which suggest that the investigation into these inconsistencies is sometimes usurped by the use of auxiliary hypotheses. If this activity is widespread, then the validity of our measures as well as the strength of our theories is put into doubt.

In the present chapter, however, I turn to a more subversive problem with construct validity, a problem that is particularly significant when we apply construct validity to questions of quality of life. The logic of construct validity instructs us to question our theory or our measure only when inconsistencies between the two arise, on the other hand, when a measure's outcomes confirm our hypotheses we take this confirmation as evidence that both our theory and our measure are good. I suggest in this chapter that this latter aspect of construct validity is too conservative.

This conservative aspect of construct validity is a problem, firstly because it opens the possibility for the perpetuation of ethnocentric or biased understandings of what makes for a good quality of life. To be sure much depends on the specific content of our theory and the construction of our measures. Nonetheless a single set of outcomes can be understood to confirm different hypotheses derived from different theories, but the logic of construct validity leads us to favor one theory at the expense of others. Secondly, this conservative tendency is a problem if it leads to ethnocentric understandings of quality of life because quality of life outcomes are often used as empirical evidence to support public policy. These recommendations are often made without attention to or deliberation of the specific theory of quality of life that underpins the outcomes.

In response to this problem I suggest that we ought to continue to question our theoretical orientation of quality of life as well as the structure of our measures even when our outcomes confirm our hypotheses. For instance, the very fact that quality of life improves after a particular medical intervention might give us pause over our theoretical orientation and provide the impetus for further questions. For instance, we might ask what this outcome says about us—what does it say about our understanding of quality of life and the assumptions that this understanding embodies? Moreover, we might further ask if this understanding of quality of life is defensible?

In what follows I further develop some of the consequences that result from what I am calling the conservative quality of construct validity when it is applied to quality of life research. In the first half of this chapter I propose a theoretical orientation for the standard needs approach to quality of life. To begin this task I turn to Dan Brock's analysis of standardized health-related quality of life measures. I then use Ron Amundson's criticisms of Brock to argue that the standard needs model prioritizes certain biological functionings as the threshold for good quality of life. In the second half of this chapter I look at how at least in some cases construct validity works to maintain the theoretical assumptions built into the standard needs approach to quality of life. To illustrate this problem I turn to patient-reported outcomes of cochlear implant recipients and contrast them with Harlan Lane's account of deaf culture in *The Mask of Benevolence.* Finally in the last section I return to individualized quality of life measures and look at some of the difficulties and risks involved in assuming quality of life is individual, idiosyncratic and relative.

I

Theorizing about Standardized Measures of Quality of Life: The Importance of Self-determination

In his paper 'Quality of Life Measures in Health Care and Medical Ethics' Dan Brock examines two aspects of clinical ethics—informed consent and standardized health-related quality of life measures—in order to draw from these concrete practices a broad account of quality of life or as he sometimes calls it the 'good life' which might motivate them. This task is important for my purposes because as we saw in the previous chapter unlike the individualized approach the standard needs approach does not itself provide us with an explicit theory that explains the logic of standardized measures, instead these measures are often complex and lack a firm theoretical basis.[74] Moreover, unlike recent efforts to provide a theory on which we *should* model quality of life measures[75] Brock's project is descriptive not normative; he is concerned to identify an account of quality of life that these measures express in their present form and not with what they ought to express. Thus his work is helpful in providing an evaluative framework for current quality of life measures.

Although I am most interested in Brock's analysis of standardized health-related quality of life measures in order to fully appreciate the picture of quality of life which he takes to motivate them it is important to attend briefly to some of his comments on ethical frameworks for medical decision-making. Brock takes the dominant model of medical decision-making to be one of shared decision-making between the patient and the physician, but where the patient is ultimately responsible for the decision to accept or reject treatment.[76] This model of decision-making is usually referred to as the doctrine of informed consent. Informed consent reflects a relatively recent change in health care which Brock takes to be motivated by new developments in technology that provide for the

---

[74] Cummins, R.A. (2005) 'Moving from the Quality of Life Concept to a Theory', in *Journal of Intellectual Disability Research*, 49: 701.

[75] See Bramston, P. *et al.* (2005) 'Conceptual Principles of Quality of Life: an Empirical Exploration', in *Journal of Intellectual Disability Research*, 49: 728:-33; Cummins, R.A. (2005) 'Moving from the Quality of Life Concept to a Theory', in *Journal of Intellectual Disability Research*, 49: 699-706; Ferrans, C.E. *et al.* (2005) 'Conceptual Model of Health-Related Quality of Life', in *Journal of Nursing Scholarship*, 37: 336-42; and Hajiran, H. (2006) 'Toward a Quality of Life Theory: Net Domestic Product of Happiness', in *Social Indicators Research*, 74: 31-43.

[76] Brock, D. (1993) 'Quality of Life Measures in Health Care and Medical Ethics', in *The Quality of Life*, Martha C. Nussbaum & Amartya Sen eds, Clarendon Press: Oxford, p. 101.

possibility of extending lives indefinitely, but with uncertain benefit.[77] Informed consent reflects the view that what counts as "health" is no longer merely the purview of doctors—it is no longer simply a biological fact. New technologies which blur the line between health and mere existence mean that patients have a right to shape their treatment in line with particular goals, values and life plans.

But what grounds the right to informed consent? Even if health is no longer understood as simply a biological fact we still might see medics as our best guide to therapeutic practice. Why must we involve individuals and their broad concerns to bear on what are still essentially medical decisions? If we turn to the literature in bioethics informed consent is typically justified on the basis of two values: the protection from harm and the protection of individual autonomy.[78] Historically informed consent became essential to medical practice following public outrage over cases of patient exploitation for the purposes of medical experimentation. For instance, the Nuremberg Code laid out guidelines for informed consent after the experimentation on non-consenting individuals in concentration camps during World War II.[79] In this context voluntary informed consent was understood as indispensable to the protection of individual safety and well-being for it was thought unlikely that an individual would consent to procedures so contrary to her own self-interest. For those who understand protection from harm as the primary justification of informed consent an individual's informed and uncoerced decision ought to promote her well-being.

---

[77] Ibid.

[78] Beauchamp, T.L. & Childress, J. F. (1983) (2nd ed) Principles of Biomedical Ethics, New York: Oxford University Press, p. 67.

[79] Ibid, p. 66.

But it is sometimes the case—especially as technology increases our ability to prolong life—that the informed and voluntary consent given by a patient is considered to be contrary to their well-being, for instance in the case where a patient refuses treatment for a potentially curable illness. Instances such as these have led to legal cases which tend to justify informed consent on the basis of protection of individual autonomy. Take for instance, the landmark 1960 case of *Natanson* v. *Kline* where the judge found that,

> Anglo-American law starts with the premise of thoroughgoing self-determination. It follows that each man is considered to be master of his own body, and he may, if he be of sound mind, expressly prohibit the performance of lifesaving surgery, or other medical treatment.[80]

Subsequent court cases and current medical practice reinforce the protection of individual autonomy as the predominant justification of informed consent. Following Kant supporters of this justification often understand respect for persons to be the same as a respect for their choices and different life plans: to be a person is to develop a plan and choose a course of action.[81] On this view informed consent is usually understood as a necessary condition for respecting patients as persons. To withhold information from a patient or to make decisions in a patient's best interest is to treat them with something less than the respect owed to equals; it is to treat them paternalistically. Brock echoes this view when he writes that self-determination is what is required in order to recognize an

---

[80] Ibid, p. 69.

[81] Ibid, p. 59.

individual as a person.[82] Moreover, he writes in another paper that self-determination is, 'a central source of human dignity'.[83]

Given that self-determination practically and legally grounds informed consent we must respect an individual's decisions even though these decisions may or may not increase their well-being. But how should we understand the relationship between the value of self-determined choice and an individual's well-being?[84] According to Brock the typical view in medical ethics is that a patient's well-being is generally thought to be the same as a patient's good and self-determination is a value independent of one's well-being or good.[85] Sometimes in order to respect a patient as a person we must also respect decisions contrary to their good. On this account, a patient's individual well-being or good is seen as separate from the value of treating her as an autonomous person; the importance of the latter sometimes justifies the sacrifice of the former.

But for Brock if self-determination is a fundamental value necessary for treating patient's as persons, then our conception of an individual's good ought to be capable of encompassing this value instead of setting it off as a separate concern.[86] Brock suggests that unlike the conventional view above we should make a distinction between well-being and a patient's good. A patient's good should incorporate self-determination and well-being should be measured separately in terms of one's state of consciousness, activities and

---

[82] Brock, D. (1993) 'Quality of Life Measures in Health Care and Medical Ethics', in *The Quality of Life*, Martha C. Nussbaum & Amartya Sen eds, Oxford: Clarendon Press, p. 109.
[83] Brock, D. (2005) 'Preventing Genetically Transmitted Disabilities while Respecting Persons with Disabilities', in *Quality of Life and Human Difference Genetic Testing, Health Care, and Disability*, David Wasserman et al. eds, Cambridge: Cambridge University Press, p. 70.
[84] Brock, D. (1993) 'Quality of Life Measures in Health Care and Medical Ethics', in *The Quality of Life*, Martha C. Nussbaum & Amartya Sen eds, Clarendon Press: Oxford, p. 109.
[85] Ibid.
[86] Ibid.

capacities for functioning.[87] Moreover, Brock argues that the doctrine of informed consent implicitly accepts such a distinction: in acknowledging that we must respect choices like a patient's refusal for treatment the doctrine of informed consent implies that a life of choice is the best life even if those choices result in lower well-being.[88]

In turning from informed consent to standardized measures of health-related quality of life Brock further develops his position that self-determined choice is part of a good quality life and not something valued independently of it.[89] Brock begins to support this suggestion by directing our attention to the way standardized measures assess quality of life, namely with respect to what he calls primary functions.[90] Primary functions refer to centrally important activities which when missing from one's life significantly limit one's choices or opportunities in creating and pursuing different life plans.[91] These functions are represented by the different dimensions in a measure and the questions within the dimensions assess the impact of disease or illness on an individual's life by gathering information about how well she is able to perform the requisite function.[92]

To illustrate, the Sickness Impact Profile (SIP) asks respondents questions like, 'I am going out for entertainment less'; 'I laugh or cry suddenly'; 'I do not bath myself at all,

---

[87] Brock gives the example of a parent who foregoes expensive lifesaving treatment to save money for her child's education. See Ibid, p. 110.

[88] Ibid. p. 110.

[89] Brock's analysis of standardized measures of health-related quality of life has much in common with Amartya Sen and Martha Nussbaum's capability approach to quality of life. See for instance Sen, A. (1999) 'Capability and Well-Being' in Martha C. Nussbaum & Amartya Sen eds. *The Quality of Life*, Oxford: Clarendon Press, pp. 29-53. and Nussbaum, M.C. (1999) Sex and Social Justice, New York: Oxford Press, pp. 29-54.

[90] Brock, D. (1993) 'Quality of Life Measures in Health Care and Medical Ethics', in *The Quality of Life*, Martha C. Nussbaum & Amartya Sen eds, Clarendon Press: Oxford,, p. 117.

[91] Ibid, p. 124.

[92] Ibid, p. 119.

but am bathed by someone else'.[93] These questions relate to the dimensions or primary functions of 'Recreation and Pastimes', 'Emotional Behavior' and 'Body Care and Movement' respectively. Depending on respondents' answers to the questions in the different dimensions they are each given a score which determines their functioning level within each dimension. One's quality of life decreases as one's functioning decreases. To be sure, as I mentioned in Chapter 1 some standardized measures of health-related quality of life combine these sorts of functional assessments with questions regarding personal satisfaction and health perception. But in standardized measures answers to the latter kinds of questions cannot wholly override the results of the former although they are often taken into account. The Short Form-36 (SF-36), for instance, gives questions regarding satisfaction and health perception the same priority as functional ones.

Recall from the discussion on informed consent that the value of self-determination itself is grounded in an account of persons whereby respect for persons is procured when we respect their decisions in accord with a self-determined life plan. The value of self-determination highlights the value of personal choice, but the integrity of this value requires that individuals have a certain minimal array of valuable choices or opportunities from which to choose for without such an array self-determination is at best an empty promise and at worst a tool for manipulation.[94] For Brock the dimensions represented in standardized measures like the SIP capture the functionings which represent the minimal array of options one must have in order for self-determination to fulfill its promise of

---

[93] Ibid, p. 118.

[94] For instance, giving people the right to vote, but only placing one person or party on the ballot; giving women the right to make choices concerning their lives, but making it materially difficult for them to work outside the home.

treating patients as persons. Thus following the methodology of standardized measures Brock takes these primary functions to be objective—individuals with a loss of primary functioning have a lower quality of life even in light of individual accounts of satisfaction and happiness.

For Brock the objectivity of these judgments correctly redresses the danger of adaptive preferences. As Brock puts it,

> To be satisfied or happy with getting much less from life, because one has come to expect much less, is still to get *less* from life or to have a less good life.[95]

He formally justifies this position with reference to Norman Daniels' notion of a "normal opportunity range".[96] Echoing the idea that self-determination requires a certain array of valuable choices Brock uses Daniels to argue that individuals whose disability or illness restricts the activities that they would otherwise be able to perform have a lower quality of life than individuals without such restrictions because certain basic choices or normal opportunities are curtailed.

In Brock's analysis of standardized quality of life measures he emphasizes their focus on a person's functionings as opposed to, what he calls, a person's disability or illness, whereby he understands "disability or illness" to refer to a respondent's physical or mental impairment. Thus quality of life measures do not ask respondents to disclose their physical or mental circumstances—what we might think of as their biological condition. Instead quality of life measures ask respondents about what they can and cannot do, in other words, they ask about respondents' primary functions. It is the *loss* of primary

---

[95] Brock, D. (1993) 'Quality of Life Measures in Health Care and Medical Ethics', in *The Quality of Life*, Martha C. Nussbaum & Amartya Sen eds, Clarendon Press: Oxford, p. 125.
[96] Ibid, p. 124.

functions that is taken to reduce opportunity range and hence quality of life, not the *presence* of disability.

This emphasis on functions has the advantage that not all disabilities—physical or mental impairments—will affect functioning levels. Brock gives the example of a *60 Minutes* program from 1988 which interviewed individuals born to women who had taken Thalidomide during their pregnancy. The individuals were born with a variety of physical deformities, but they had so completely adjusted to their condition that they were able to perform all the primary functions albeit in unique and creative ways. In these cases quality of life may not be diminished and in fact Brock thinks we should question the use of "disability" as a correct representation of this situation.[97] We might understand Brock's suggestion to support the idea that term "disability" does not merely refer to physical or mental impairment, but impairment that results in a lack of functioning.

In a more recent paper Brock clarifies his position on this point: for a physical deformity or illness to have little or no impact on quality of life individuals must be able to perform the *same* primary functions as those without such problems. He gives the counterexample of deafness where in spite of claims that sign language and deaf culture allow individuals a rich and functional life he argues that there are valuable human activities like listening to music which they cannot chose to do. As a result their quality of life is objectively less than those who can hear.[98] Nonetheless, notice that according to Brock's analysis of standardized quality of life measures it is not the physical impairment

---

[97] Ibid, p. 123.

[98] Brock, D. (2005) 'Preventing Genetically Transmitted Disabilities while Respecting Persons with Disabilities', in Quality of Life and Human Difference Genetic Testing, Health Care, and Disability, David Wasserman *et al.* eds., Cambridge: Cambridge University Press, p. 75.

associated with being deaf—it is not their inability to hear—that makes their quality of life worse, rather it is their inability to participate in certain valuable functions.

If, as Brock suggested at the end of his discussion on informed consent, self-determined choice is a part of the good life and not merely independent of it, then significant limitations on an individual's ability to choose how they will live their life or limitations on the opportunities they have, for instance limitations due to disability or illness, will make for a worse life. The objectivity embodied by standardized health-related quality of life measures appears to operationalize this point and thus meshes with Brock's suggestion. On this account standardized measures of health-related quality of life aim to promote substantive choices and reflect negatively on disease or illness which limit these choices.

By looking at certain aspects of the practice of ethical medicine Dan Brock draws support for three conclusions about the account of a good quality life which undergirds standardized measures of health-related quality of life—what I referred to in the previous chapter as the standard needs approach. First, self-determination is central to quality of life, second, we need a sufficient number of valuable choices if we are to fulfill the goal of self-determination and third, an individual's primary functions can give us information about their quality of life. In the next section I want to supplement this account of the standard needs approach and examine the type of primary functions which are treated as conditions for the possibility of self-determination. To do this I look at a criticism of Brock's work on quality of life, which suggests that his approach—and by extension health-related quality of life measures—are biased to favor clinically or biologically sound bodies.

II

A Potential Bias in Standardized Measures of Quality of Life

As we have seen on Dan Brock's account of standardized quality of life measures disabled or ill individuals have an objectively lower quality of life if their disability or illness reduces their opportunity to partake in major life activities. In fact Brock uses the American Disabilities Act (ADA) to support this claim citing their definition of a disability as a physical or mental impairment that substantially limits at least one major activity.[99] But although this is the definition of disability which the ADA uses Ron Amundson questions Brock's interpretation of it in his paper 'Disability, Ideology, and Quality of Life: A Bias in Biomedical Ethics'. He begins by questioning Brock's judgment that if one has a limitation in a major activity, then one necessarily has a lower quality of life. Why, Amundson asks, must we assume that unlimited major life activities are necessary for a good quality of life?[100]

As we saw in the last section the link between unlimited major life activities and a good quality of life is due, at least in part, to the commitment to self-determination and the scope of valuable choices necessary to fulfill this value. Nonetheless, we might understand Amundson's question to point in a different direction. We might understand his question to challenge the notion of self-determination that Brock reads into quality of life measures,

---

[99] Brock, D. (2005) 'Preventing Genetically Transmitted Disabilities While Respecting Persons with Disabilities', in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman *et al.*, New York: Cambridge University Press, p. 72.

[100] Amundson, R. (2005) 'Disability, Ideology, and Quality of Life: A Bias in Biomedical Ethics' in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman *et al.*, New York: Cambridge University Press, p. 108.

namely that maximizing a set of valuable choices increases the scope for self-determination and thus quality of life.

Following this line of thought we might argue that in some cases restrictions on valuable opportunities enable certain kinds of life plans and identities which are otherwise unavailable and which also enhance quality of life. For instance, we may take oral communication to be a valuable functioning. But we might argue that it is only in virtue of the *inability* to communicate orally that the deaf community and the notion of a deaf identity exist. Moreover, we might argue that living a life as a member of the deaf community is a valuable life plan and it is so in virtue of the valuable opportunities open only to those who are deaf. For instance, deaf individuals have the opportunity to experience the world in a predominately visual and spatial framework; they have the opportunity to attend deaf schools and form friendships with other deaf individuals and be part of a community with a specific culture, history and political agenda.

Thus, we might say that the ability to communicate orally, to listen to music and so on—the ability to participate in certain valuable functionings—militate against a life lived as a member of the deaf community, a life which represents a different set of valuable functions. To be sure, it is often the case that having choices is important to self-determination and thus a good quality of life, but in arguing for the value of certain limitations in primary functions proponents of deaf culture need not take themselves to be limiting our opportunities, but rather protecting the opportunities that only some of us have.

One of the assumptions built into standardized quality of life measures is that one's quality of life is better if one has more valuable options from which to choose. But in light of what was said above regarding the incompatibility of some valuable choices we might think that the very idea of maximizing our valuable options is confused. Quality of life is not simply a matter of having more valuable options from which to choose, but first requires that we make decisions regarding what options are valuable.[101] These decisions, however, depend on a vision of the good life in which an option or set of options is deemed valuable. For instance, oral communication may not be taken as valuable given certain understandings of community, friendship, culture and bodily integrity. Standardized quality of life measures, however, do not attend to the variable conditions in which different choices are considered valuable, instead they assume that certain primary functionings are equally valuable for everyone regardless of their vision of the good life; they assume that certain functionings act as a threshold below which quality of life can only be worse.

From the point of view of the deaf community this assumption can have adverse consequences for deaf individuals and how people perceive their quality of life. For example, one of the dimensions or primary functions found in the Sickness Impact Profile (SIP) is 'Communication'. In this dimension respondents are asked questions regarding trouble writing or typing and whether they speak clearly when under stress. A yes answer to any of these questions is taken to reflect negatively on quality of life since difficulty writing, typing or speaking is taken to hinder communication and thus the ability to perform a primary function. But as I will examine in more detail in the next section deaf people may not be able to write or speak clearly—especially when stressed—and yet deaf advocates argue that their quality of life can be as good as hearing people's.

---

[101] Amaryta Sen has recognized this point with respect to the capabilities approach to quality of life. For his discussion of the topic see, for instance, Sen, A. (1999) 'Capability and Well-Being' in Martha C. Nussbaum & Amartya Sen eds. *The Quality of Life*, Oxford: Clarendon Press, pp. 29-53.

If we accept the point made by the deaf community and indeed other disability activists that some life plans are valuable in spite of, or even because of restrictions to certain valuable functions, then we might argue that standardized quality of life measures embody a bias against certain kinds of valuable lives. Indeed Amundson does make this argument and in doing so he argues that bioethics itself tends to embody a vision of the good which favors the lives of biologically normal individuals.[102]

In particular Amundson challenges Brock's commitment to primary functions by arguing that many of these functionings have biological prerequisites, for instance, being able to hear, see, walk and so on. For instance, in the SIP above 'Communication' is taken to be, in part, oral communication and oral communication almost always presupposes that one can hear. Moreover, recall from the previous section that the SIP also takes 'Body Care and Movement' to be a valuable functioning and asks respondents if they can bath themselves. This question, however, presupposes at least partial use of one's arms and legs. Although Brock emphasizes the fact that quality of life measures focus on what individuals are able to accomplish and not on their physical or mental impairment, if primary functions require certain biological 'normalities', then for practical purposes a good quality of life is only open to certain bodies. Indeed Brock comes close to endorsing this view, at least for some conditions, when he writes,

> ...serious disabilities...remain significant disadvantages for common human pursuits even after the goal of achieving reasonable and just social

---

[102] Amundson does not direct his argument specifically to quality of life measures, but rather to bioethics in general and Brock's theoretical work in particular. I will only attend to Amundson's arguments insofar as they relate to Brock's work and thus to quality of life measures.

> accommodation to disabilities has been reached; they are not "mere" or solely social constructions or socially constructed disadvantages.[103]

In this passage Brock's use of the term "serious disabilities" refers to physical or mental impairments which limit a major life activity no matter what kind of social adjustments are made, for instance, wheelchair access, Braille translations, sign language interpreters and the like. For Brock some disabilities remain disabilities and no amount of social resources can change the fact that some individuals just cannot participate in all of life's valuable functionings. Indeed, as I will discuss in the next section, no amount of resources will provide deaf individuals the level of oral communication which hearing people enjoy and I suspect that likewise there is very little we can do to enable the blind to watch a sunset. Moreover, it is because deaf people cannot hear and blind people cannot see that they are unable to do these things.

Nonetheless, we might argue that these "disabilities" *are* "socially constructed" for something can only count as a disadvantage given a certain view of a good life and we need not accept that a good quality life requires the ability to communicate orally or to view scenery. Put differently, we need not accept that a good quality of life requires certain biological prerequisites. To be sure, it is possible to define opportunity and thus a good quality life so narrowly that certain impairments by definition limit our opportunities. Amundson, however, calls this move a "trivialization"[104]—we can also say that men lack the opportunity to bear children due to their biology, but this fact is not typically taken to

---

[103] Brock, D. (2005) 'Preventing Genetically Transmitted Disabilities While Respecting Persons with Disabilities', in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman et al., New York: Cambridge University Press, p. 73.

[104] Amundson, R. (2005) 'Disability, Ideology, and Quality of Life: A Bias in Biomedical Ethics' in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman et al., New York: Cambridge University Press, p. 108.

objectively limit their quality of life.[105] Men, we might say, have a different, but nonetheless valuable set of opportunities as opposed to women and these differences need not reduce either their self-determination or their quality of life. If this is the case with men and women's different sets of opportunities, then why does the inability to communicate orally objectively decrease quality of life for the deaf?

This question is not merely rhetorical; I suggest that the answer to it is two-fold. The inability to communicate orally theoretically limits quality of life when, for instance, quality of life measures embody a vision of the good life that requires the ability to hear. In this case individuals who cannot speak clearly are understood to have a worse quality of life than those who can. Disability activists sometimes target their arguments at this level of the debate contending that in principle the deaf and disabled can live lives of good quality. But practically the inability to communicate orally begins to limit quality of life when social and environmental circumstances are such that the inability to speak clearly limits any form of communication or community. At this level of the problem activists argue that while quality of life may not be good, this assessment has nothing to do with hearing per se. Rather it reflects inequality of opportunity and a biased management of social resources. For disability and deaf activists we ought to recognize that some individuals live different, but nonetheless valuable lives and provide resources to help them achieve their unique vision of the good.

The theoretical and practical issues that affect the quality of life of the deaf work hand in hand, for if we assume a picture of the good life that requires the ability to hear,

---

[105] On the contrary, the inability to bear children is sometimes taken to improve men's quality of life and we could argue that this too is the result of certain notions of the good life that prioritize independence, detachment and individuality.

then our efforts to provide deaf individuals with a good life often dismantle the conditions in which an inability to hear might thrive. For example, consider the following. Health-related quality of life measurement is meant to be part of a progressive approach to health care; a commitment to health-related quality of life suggests that we ought to improve one's ability to function through the implementation of more effective medication, better hospitals and so on. But in the face of certain primary functions social reform cannot ameliorate all disabilities. Thus there are limitations to how much we can improve the lives of the deaf and disabled though social planning. From this perspective some individuals necessarily have fewer choices and thus they necessarily have a lower quality of life.

But there are other options besides social planning that *can* increase the quality of life of those lacking major life functions. These options involve medical or surgical interventions which aim to restore primary functions. For instance, surgical interventions like cochlear implants allow the deaf to hear and burgeoning genetic technology will soon make it possible to detect the presence of a variety of disabilities early in pregnancy and thus allow for selective abortions. Given a particular threshold of primary functions both interventions result in a population of fewer disabled people and therefore more people with more choices. It is not surprising perhaps that Brock supports the use of genetic technology and that outcome assessments based on standardized health-related quality of life measures indicate that cochlear implants indeed improve quality of life.[106]

---

[106] See for instance, Brock, D. (2005) 'Preventing Genetically Transmitted Disabilities While Respecting Persons with Disabilities', in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman *et al.*, New York: Cambridge University Press, pp. 67-100 and Harris, J.P. (1995) 'An Outcomes Study of Cochlear Implants in Deaf Patients', in *Archives of Otolaryngology Head Neck Surgery*, 121: 398-404.

Amundson's criticisms of Brock are not meant to argue that all disabilities make for a good quality of life nor that just any vision of the good is legitimate. Rather Amundson's criticism is directed at what he understands to be an *unjustified* assumption that the absence of certain biological norms necessarily reduces quality of life.[107] Nonetheless, even this allegation is not entirely true for quality of life researchers may reasonably take themselves to be *justified* in holding this assumption. As we saw in the previous chapter construct validity theoretically allows us to test both our measure and build up our theory. If our measures bear out our hypotheses, then greater confidence in our theory is justified. If the outcomes data from standardized measures of quality of life support the notion that valuable choices and hence quality of life improve as respondents gain the ability to hear, see, eat and so on, then according to the logic of construct validity we may be justified in assuming that certain biological differences reduce quality of life.

In the next section, however, I challenge the logic of construct validity and suggest that with respect to quality of life research even when—and sometimes especially because—our hypotheses are borne out in a measure's outcomes we still need to question our theory and our measure. In what follows I look at some of the past and present research that has guided public policies and opinion on the deaf and argue that claims of construct validity do not further the debate over what makes for a good quality life.

---

[107] Amundson, R. (2005) 'Disability, Ideology, and Quality of Life: A Bias in Biomedical Ethics' in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman *et al.*, New York: Cambridge University Press, p. 103.

III

The Quality of Life of the Deaf: A Case Study

A. A Brief History of Research on the Deaf

In Harlan Lane's book *The Mask of Benevolence* he argues that prelingually deafened individuals are part of an oppressed cultural minority. Moreover, like Amundson he argues that clinical assumptions about what makes for a healthy body has led medical professionals and educators to accept a pathological view of deafness whereby it is understood as something that needs to be cured. Confusing clinical criteria of health with ethical criteria for a good life he argues that the hearing establishment has for the last hundred years sought to give deaf individuals the "benefits" of hearing, whether this was through oral education, mainstreaming or surgery. In what follows I lay out parts of Lane's argument to provide the historical and cultural background against which the debate over the quality of life of the deaf has been, and continues to be, fought.

For Lane deaf individuals represent a *cultural* minority because they share a complex language replete with literature, jokes and customs; they share a common history and social structure.[108] Although the National Association of the Deaf (NAD) recognizes that the diversity within the deaf community makes the deaf experience difficult to generalize we might nonetheless acknowledge certain shared values.[109] For instance, the deaf community has traditionally identified deafness independent of decibel loss and in virtue of communication preference.[110] Moreover, they have expressed pride in this

---

[108] Lane, H. (1993) *The Mask of Benevolence*, New York: Vintage Books, pp. 16-7, 18.

[109] NAD Board of Directors (2000) 'Cochlear Implants NAD Position Statement', www.nad.org/site retrieved 5/7/2006.

[110] Cohen, J. (1998) 'The Deaf Identity Double-Bind: Culture Versus Disability', in *The Advocate's Forum*, 5.

identity: they have welcomed the arrival of deaf children[111], valued and maintained the friendships of deaf schoolmates and espoused the belief that a hearing person can never acquire a deaf identity, even those who grow up with deaf parents and are fluent in American Sign Language (ASL).[112]

This latter belief is important because it partly explains the reluctance of some members of the deaf community to accept the cochlear implantation of young children as well as the occasional prejudice against deaf adults who attempt to assimilate into hearing culture by way of lip reading, attempting to speak or receiving an implant.[113] But it also suggests that on this account "deaf identity" only refers to those who are prelingually deaf—those born deaf or who become deaf shortly thereafter—and not to those becoming deaf in later life. Thus Lane's account of the deaf does not necessarily apply to postlingually deafened individuals. This is different from Amundson's argument which was meant to concern the value of disabled life generally. Nonetheless, Lane's discussion reinforces Amundson's and helps us to see how contrary to Brock's account of quality of life giving people more choices by way of certain functionings may in fact illuminate bias and serve paternalistic interests.

'Audism' is the term that Lane and others use to describe the hearing-centered endeavor that attempts to help deaf people; audists are those who work to serve this

---

[111] Lane, H. (1993) *The Mask of Benevolence*, New York: Vintage Books, p. 18.

[112] Ibid, p. 17.

[113] Ibid, p. 6. Since 2000 NAD officially recognizes that one's choice of communication is personal and ought to be respected. With respect to cochlear implants see 'Cochlear Implant NAD Position Statement' at www.nad.org/site . Nonetheless, individuals receiving cochlear implants still report criticism from the deaf community. See Chee, G.H. *et al.* (2004) 'Benefits of Cochlear Implantation in Early-Deafened Adults: The Toronto Experience', in *The Journal of Otolaryngology*, 33:30.

cause.[114] Deaf people need help on the audist view, not only because they cannot hear and thus appreciate music, but because correlated with the inability to hear are a host of social, cognitive, behavioral and emotional problems. Although now over 10 years old Lane provides a list of these characteristics based on a review of over 350 articles and books representing the psychometric research done on deaf people at the time. It is useful to look at some of the items on his list since it is part of the history that defines the current work on the quality of life of the prelingually deaf. Some of these traits include: asocial; egocentric; clannish; aggressive; moody; unclear thinking; unintelligent; submissive; immature; morally undeveloped and isolated.[115] If these traits or characteristics came from reliable and valid studies, then this list purports to capture the ways that deaf individuals as a group differ from hearing individuals more than the ways these individuals differ amongst themselves. It is perhaps not surprising that children diagnosed as deaf or seriously hearing impaired were—and sometimes still are[116]—seen to be in danger of leading lives of poor quality.

B. The Controversy over Cochlear Implants

Cochlear implants are the latest development in biotechnology aimed to eliminate or reduce deafness and its associated problems. Cochlear implants are prosthetic replacements for a damaged ear and are surgically inserted into the inner ear through a depression in the skull. The implant itself is a very small wire which directly stimulates the auditory nerve

---

[114] Lane, H. (1993) *The Mask of Benevolence*, New York: Vintage Books, p. 43.

[115] Ibid, p. 36.

[116] See Montgomery, C. (2002) 'The Cochlear Implant Trial', in *The Ragged Edge Magazine*, at http://www.raggededgemagazine.com/extra/deaftrail; McClellan, T.D.(2002) 'Deaf Mom Fights to Keep Kids form Ear Implants', in *The Grand Rapids Press*, at http://www.bridges4kids.org/articles/9-02/GRPress9-6-02.html

with electrical current. A cochlear implant recipient must also wear an external earpiece which contains a microphone and a transmitter, and a speech processor which is connected to the earpiece by two thin cords. Sounds are picked up by the microphone and then carried via one of the cords to the processor which selects the sound waves most useful for speech and sends them to the transmitter via the other cord. The transmitter sends the coded signal as FM radio waves across the skin and the radio waves are picked up by the internal wire.[117] Cochlear implants do not cure deafness and although they do enable an individual to hear sound when the earpiece is inserted the sound that one hears is much less refined than that heard by a hearing person.[118]

In the literature it is now routinely acknowledged that cochlear implantation in *post*lingually deafened adults has significant benefit both in terms of hearing and quality of life. As one research team puts it, 'Cochlear implantation has become the method of choice for the treatment of postlingually deaf patients.'[119] But until recently it was thought that cochlear implants had little to offer *pre*lingually deafened adults. This is mainly because prelingually deafened adults who receive cochlear implants are unable to understand speech through the audition gained from the implant and the ability to understand speech was thought to be the main benefit from cochlear implants.[120] The inability of prelingually deafened adults to understand speech based on the sounds derived from the implant is usually thought to be due to profound hearing loss during the years of language learning,

---

[117] Washington University School of Medicine in St. Louis, Department of Otolaryngology homepage, *Cochlear Implant—What is it?* Found at http://wuphysicians.wustl.edu/dept

[118] NAD Board of Directors (2000) 'Cochlear Implants NAD Position Statement', www.nad.org/site retrieved 5/7/2006.

[119] Vermeire, K. *et al.* (2005) 'Quality-of-Life Benefit from Cochlear Implantation in the Elderly', in *Otology & Neurotology*, 26: 188.

[120] Kaplan, D.M. *et al.* (2003) 'Early-Deafened Adult Cochlear Implant Users: Assessment of Outcomes', in *The Journal of Otolaryngology*, 32: 246.

abnormal development of auditory pathways and the lack of auditory experience.[121] It is for these reasons that the early disagnosis of deafness and the subsequent implantation of young children was and still is thought to be important.[122]

But recent studies indicate that prelingually deafened adults, contrary to expectations, are not only seeking implantation, but also receiving some measure of benefit. In a recent study of 44 prelingually deafened cochlear implant recipients researchers at the University of Toronto found that independent of variable audiologic outcomes—the actual benefit to hearing in terms of speech perception—the quality of life data from these individuals approached that of the quality of life data from postlingually deafened implant recipients from the same program.[123]

In another study from the University of Toronto—the largest study of the subjective benefits of cochlear implantation in early-deafened adults—it was found that this improvement in quality of life is mainly due to more confidence in communication via lip reading ability, better environmental awareness and greater independence.[124] Moreover, when participants were asked about their personal satisfaction with the implant 66.7% were 'very satisfied' and when asked if they would go through the same process again 93.3% said they would.[125]

---

[121] Ibid.

[122] See the Alexander Graham Bell Association for the Deaf and Hard of Hearing at www.agbell.org to read about a campaign to diagnose and implant young children.

[123] Kaplan, D.M. *et al.* (2003) 'Early-Deafened Adult Cochlear Implant Users: Assessment of Outcomes', in *The Journal of Otolaryngology*, 32: 249. This study's conclusion is contrary to earlier findings which suggested that the longer one has been deaf the less one can expect improvement in quality of life. See Maillet, C.J. *et al.* (1995) 'Change in the Quality of Life of Adult Cochlear Implant Patients', in *Annuals of Otolaryngology Rhinology Laryngology Suppl.*, 165: 31-48.

[124] Chee, G.H. (2004) 'Benefits of Cochlear Implantation in Early-Deafened Adults: The Toronto Experience', in *The Journal of Otolaryngology*, 33: 29-8, 31.

[125] Ibid, p. 27.

That cochlear implants improve the quality of life of prelingually deafened adults was not one of the hypotheses with which auditory specialists or quality of life researchers began their inquiry. In fact researchers assumed that cochlear implants would *not* significantly increase their quality of life since it would not enable them to understand speech. As construct validity instructs us the Toronto team took their unexpected results as evidence that their original hypothesis was false and that the part of their theory regarding the relatively narrow importance of hearing was misguided.

But although the evidence from these studies suggested that part of their theory regarding the benefit acquired through cochlear implantation was false the researchers took this new information as evidence which further emphasized the significance of hearing to a good quality of life. The new evidence suggested to the researchers in Toronto yet another aspect of the importance of hearing to quality of life, an aspect that had been overlooked in the past. Hearing, it turns out, is important to quality of life not only because it allows for word discrimination and sentence recognition, but also when it simply enables one to hear indiscriminate sounds.

Although the working hypothesis from this study was not confirmed we might think, along with the Toronto team, that the outcomes from this study further validate a theory of quality of life which acknowledges the importance of certain biological prerequisites—the theory of the good life that, according to Amundson, standardized measures of quality of life embody. But in what follows I suggest that such a conclusion is too quick—the data from the Toronto study does not provide unequivocal support for this theory as the correct orientation to quality of life research. Rather even with these outcomes we are still left to grapple with the question of what makes for a good quality life.

Historically members of the deaf community have rejected the notion that the inability to hear decreases quality of life. They have rejected it because it suggests that deaf people have an objectively lower quality of life without implants or hearing aids; that quality of life and the ability to hear are synonymous. Instead, they argue that, on the one hand, the deaf life, at least in principle, represents a life of unique quality; put differently, a deaf life represents an alternative and legitimate vision of the good. On the other hand, they argue that practically speaking discrimination against deafness and attempts to "cure" it have to a large extent undermined the conditions for the possibility of this kind of life. Thus the predominate factor in the improved quality of life scores following cochlear implantation is not the ability to hear, but the consequences of over a century of discrimination.

Thus from this perspective when standardized measures report that cochlear implants increase the quality of life of prelingually deafened adults these outcomes might be understood as indications of areas where discrimination makes full participation in society difficult. For instance, as Lane points out mainstreamed deaf schoolchildren lacking a deaf community and deaf role models, and taught oral skills that they only poorly acquire become adults who *are* isolated, *are* unconfident, and *are* insecure.[126] On this view the findings from the quality of life studies from earlier represent a social critique: even given a relatively insignificant ability to hear prelingually deafened individuals find that life gets much better.

The idea that a deaf life represents a unique, but historically maligned vision of the good provides a competing interpretation of quality of life outcomes data than do quality of

---

[126] Lane, H. (1993) *The Mask of Benevolence*, Vintage Books: New York , pp. 129-54.

life researchers, the audist establishment or Brock. While for instance, these set of assumptions understand the outcomes from the cochlear implant study above as a social critique of equality, proponents who assume that the good life requires certain functionings which entail biological prerequisites understand the outcomes as evidence of the importance of these functionings. These two interpretations come from two different theoretical orientations to deafness: deafness as a pathology derived from the inability to hear; deafness as a biological difference that provides for a unique set of opportunities and challenges. From the former perspective prelingually deafened individuals receiving cochlear implants are liberated by their ability to hear and given a wider opportunity range from which to fashion the lives of their choice; from the latter perspective these individuals are the victims of discrimination and members of a paternalistic society which has insisted that a good quality of life is a hearing life.

But given these two different theoretical orientations how should we best understand the outcomes data from standardized quality of life studies? Although I won't attempt to answer this question until the end of the next chapter it is a question of central importance because it affects in what we take quality of life to consist. The important point here is that construct validity cannot provide the answer to this question. To be sure, construct validity promises to build confidence in our theory as it validates our measures and thus provide us with an answer to the question, 'What is quality of life?' But while construct validity can provide us with *an* answer to this question it risks providing us with an ethnocentric answer to this question.

Ethnocentric conceptions of quality of life are what deaf advocates worry about when they warn deaf adults and parents about the historical bias that effects the medical profession and the media. It might be true that given certain resource distributions, for instance those that favor surgical interventions over increased funding for deaf schools, quality of life increases after a cochlear implant. But this state of affairs does not tell us what quality of life would be like if deaf schools were plentiful and American Sign Language was regularly offered as a foreign language in public schools; it does not tell us what quality of life would be like if hearing was not seen as a prerequisite to a good life.

The problem with the application of construct validity to standardized quality of life measures is that it does not always invite us to challenge these measures' preconceptions—what I have been calling their vision of the good—that defines what counts as a life of quality. As the example of deaf culture shows this shortcoming of standardized measures tends to restrict not only our conversations about quality of life, but also the different kinds of quality lives we might lead. Thus, part of the significance of this problem is the way standardized measures might perpetuate, for instance through public policy recommendations, limitations on the choices and opportunities that individuals can appreciate as valuable. As I discussed earlier, Brock suggests self-determined choice is a part of a good quality life and that significant limitations on the opportunities individuals have make for a worse life. After my discussion we might think that at least in some cases the application of construct validity to standardized measures of quality of life can contribute to a decrease in valuable opportunities; a decrease in self-determination and quality of life.

The risk of perpetuating ethnocentric understandings of quality of life and thus potentially doing damage to individuals' self-determination is the result, I suggest, of a conservative bent to construct validity which encourages researchers to discontinue the questioning of our theory and our measures when our outcomes confirm our hypotheses. We might, I suggest, go some way toward resolving this problem if we understood our outcomes as a point of departure for further questions—even when they confirm our hypotheses. For instance, with regard to the Toronto study we might learn to see how the confirmation of our hypothesis itself raises certain questions. For instance, what does the very confirmation of our hypothesis say about our understanding of quality of life in this context? We might ask what social, cultural, historical or economic factors contribute to the importance of hearing to quality of life?; Do these factors indicate a lack of social and political support for certain types of diversity?; If so, is this the kind of society we wish to cultivate?; Does a lack of support for diverse experiences contribute to our understanding of the good?

These questions illustrate the importance of questioning our outcomes for they begin to clarify the assumptions that our measures take for granted as well as suggesting a possible alternative to these assumptions. To be sure, merely learning to question our theory does not ensure that we will once and for all rid our quality of life measures from ethnocentric bias. Firstly, we can always ask further questions and thus we can always overlook areas of bias. Secondly, as I will discuss throughout this thesis there is no method of learning to ask questions, we simply learn to ask better questions as we attempt to understand the things around us; we learn to question different aspects of our outcomes as we learn to better understand quality of life. Nonetheless, questioning our outcomes is itself an essential part of coming to understand quality of life. Eliminating the bias from our measures is a process and it is on-going, but even if questioning our outcomes does not guarantee our measures against ethnocentricity it is a place to begin.

If we must question the outcomes from standardized quality of life measures, then what about individualized measures? Do they escape ethnocentricity? In what follows I discuss some of the problems with understanding quality of life as idiosyncratic.

IV

A Potential Bias in Individualized Measures of Quality of Life

One of the reoccurring themes in quality of life research is that people should be free to determine for themselves what counts as a good quality life. As we have seen both the standard needs and the individualized approaches profess to promote self-determination: the standard needs approach focuses on the conditions for the possibility of self-determination; the individualized approach focuses directly on the individual's assessment of their experience. In this way both approaches take themselves to be liberal approaches, both recognize that individuals differ in their understanding of the good life and both aim to make room for these differences within their measurements.[127]

But as we saw in Chapter 1 the individualized approach criticizes the standard needs approach for not living up to the liberal ideal; it criticizes it for being paternalistic. Proponents of the individualized approach claim that the standard needs approach systematically affects what counts as a good quality of life insofar as it determines in advance which dimensions, or primary functionings to use Brock's terminology, are most important and insofar as it pre-weights individual questions. In this chapter Amundson's

---

[127] Proponents of the individualized approach have sometimes misunderstood the liberal intent of the standard needs approach and criticized it for what at least one researcher has referred to as its 'communitarian' orientation, meaning that the standard needs approach promoted a substantial view of the good life. See Browne, J. (1997) *Individual Quality of Life in Older People, Conceptual and Methodological Challenges*, PhD Thesis in Psychology, Dublin University pp.270-5.

criticism of Brock was similar to the individualized approach's claim that pre-determined dimensions unduly dictate what counts as a good life.

As you will recall from Chapter 1 the individualized approach takes quality of life to be whatever an individual says it is. Quality of life is individual and idiosyncratic; we can only measure someone's quality of life by learning to see the world through their eyes and with their values. Measurements like the SEIQoL-DW provide individuals with the opportunity to choose the dimensions that are important to *them*, to determine how their current life fares in each of these areas and to weight *themselves* the relative contribution of each area to one's overall quality of life. Each measure is thus uniquely tailored to an individual's subjective experience; the judgment of a poor quality of life is only made when individuals rate their current lives as not living up to their own expectations.

Unlike standardized measures of quality of life the individualized approach does not nominate certain primary functionings as thresholds for good quality of life and thus individualized measures cannot be criticized for the kind of systematic bias for which Amundson criticizes certain standardized measures. Moreover, although individualized measures undergo a validation process, validation is recognized as having limited value. Nonetheless, I suggest that individualized measures suffer from a similar problem as do standardized measures: they do not encourage—indeed they discourage—researchers to question a measure's outcomes. Given the nature of the theoretical orientation of the individualized approach confirmations of *as well as* deviations from researcher hypotheses are explained in terms of the idiosyncratic nature of quality of life. Thus, as the user's manual for the SEIQoL states individualized measures are self-validating.[128] This methodology is not seen as a problem for individualized measures, but as a virtue of its particular theory.

The self-validation of individualized measures is seen as a virtue because it emphasizes that quality of life is free from the paternalism that affects standardized measures: quality of life is whatever an individual says it is. In the words of one proponent of the individualized approach, 'It seems clear that the method [of the SEIQoL] is culture-free.'[129] But are the SEIQoL and other individualized measures 'culture-free'? In closing I want to suggest that the individualized approach is not culture-free and perhaps it is even more deeply embedded in cultural bias than the standard needs approach.

To make my point about the bias inherent in individualized measures I turn to the practice of Female Genital Mutilation (FGM). FGM is the practice of cutting or removing the clitoris and sometimes part of the labia minora. Although condemned by the UN Commission on Human Rights, UNICEF and others it remains a common practice in Africa, and it can be found in smaller numbers in other countries like Australia, France, the United Kingdom and the United States.[130] It is preformed on young girls usually aged four to seven for reasons of beauty, purity and social continuity. FGM is linked to a variety of medical conditions some of which include decreased sexual functioning due to pain during intercourse; infertility; and insensitivity around the scar tissue.[131]

---

[128] See O'Boyle, C.A. *et al.*, (1993) *The Schedule for the Evaluation of Individual Quality of Life User Manual*, Dublin: Royal College of Surgeons in Ireland.

[129] Hickey, A. *et al.* (1999) 'The Schedule for the Evaluation of Individual Quality of Life', in *Individual Quality of Life Approaches to Conceptualisation and Assessment*, C. R. B. Joyce, Ciaran A. O'Boyle & Hannah McGee eds, Sidney: Harwood Academic Publishers, p. 128.

[130] Nussbaum, M.C. (1999) *Sex and Social Justice*, New York: Oxford University Press, p. 120.

[131] Ibid.

Imagine that we give the SEIQoL to a genitally mutilated woman. Because the intervention occurs when girls are quite young we would have to give it to her after the operation, let us say we give it to her at some point in her twenties. If FGM does affect quality of life, we would expect her to nominate at least some area of life associated with sexual functioning and its status to reflect her disability. Now let us imagine that despite the areas of life she nominates—whether they include sexual functioning or not—her quality of life is quite high. Are we justified in concluding that for this woman, FGM is unproblematic?

The difficulty here is that human beings *can* get used to a huge variety of conditions and learn to live rich and meaningful lives within them. As Brock recognized our preferences *are* adaptive. This woman, for example, may have learned to understand FGM as an honor or a rite of passage. Perhaps she came to understand it as part of her body and the way it functions just as a man may come to understand his circumcised penis. But just because we *can* get used to a lot of things modern democracies tend to support certain thresholds below which no one should function. This idea is captured in Brock's analysis of standardized measures of quality of life and it reflects the understanding that to be a self-determining agent certain material conditions must be met. The United Nations' condemnation of FGM suggests that the freedom to fashion a life for oneself presupposes the ability to participate in certain sexual functionings without undue pain or insensitivity.

To be sure the support for a ban on FGM does imply that this practice and the social values it expresses are not part of a good quality life just as the standard needs approach stipulates that a life without certain primary functionings is not a good quality life. But while in some circumstances we might find these thresholds paternalistic is individualism the appropriate response? As Martha Nussbaum notes in *Sex and Social Justice,* victims of FGM tend to live in communities which promulgate beliefs about female impurity and the second class status of women; as a result these women often lack education and basic literacy skills.[132] Can conditions like these lead to self-determined choices about what is and is not a good quality life? Far from self-determination situations like these begin to look like manipulation; they look like a systematic bias in favor of internalized gender hierarchies.

The individualized approach is not culture-free. When individuals express their quality of life these expressions are always grounded in cultural norms, norms which we may well wish to challenge. The individualized approach tells us that quality of life is whatever an individual says it is, but we can be wrong in our assessment of our lives. After all our vision in these matters is always only partial and the orientation of the individualized approach, by insisting that our appraisal is always legitimate, limits our opportunities to expand that vision and create perhaps better lives for ourselves. But as we have also seen, to define the dimensions of quality of life in advance, to pre-determine the conditions for the possibility of self-determination with respect to a particular vision of the good also potentially limits our opportunities because it can blind us to the genuine quality of different lives.

We might say that while the individualized approach overemphasizes our differences, the standard needs approach under appreciates them. Nonetheless, my criticisms of these two approaches stem from a common problem: in neither case can we

---

[132] Ibid, p. 127.

adequately challenge the assumptions—the vision of the good—against which quality of life is measured.

In the next chapter I will turn from issues of ethnocentricity and how we understand respondent answers—our measure's outcomes—to respondent bias and how we should understand questions. Respondent bias occurs when individuals answer questions inaccurately. These inaccuracies often occur when respondents understand the questions in a measure differently than researchers anticipated they would. In the next chapter, however, I suggest that alternative understandings of our questions often provide insight into our measures—not bias. Moreover, developing ways to incorporate this insight into the analysis of our measures might be one way to begin to challenge some of the assumptions that underpin our measures while at the same time laying the groundwork for quality of life research that honors its commitment to self-determination.

Chapter 3

Construct Validity and Respondent Bias

Introduction

In Chapter Two I argued that both standardized and individualized quality of life measures tend to perpetuate ethnocentric understandings of quality of life. In the case of standardized measures I argued that this bias takes the shape of a partiality for functionings that require certain biological prerequisites; in the case of individualized measures I argued that this takes the shape of a bias in favor of atomistic individualism. Moreover, I argued that these assumptions can result in conclusions about quality of life that are at odds with the value of self-determination—a value that both proponents of the standard needs and individualized approaches take quality of life measures to promote.

In the previous chapter I was at pains to show that the bias that infects quality of life research is not in the specific *construction* of the measure—both standardized and individualized measures are subject to it. Nor does the problem lie in a particular *theory* of quality of life—neither standardized nor individualized measures can escape criticism. Rather I argued that the difficulty resides in our tendency to stop short of questioning or challenging our measures and our theory in certain cases. With regard to standardized measures we stop asking further questions when our hypotheses are confirmed in our outcomes; with regard to individualized measures we stop asking further questions once an individual has completed a measure. I suggested, however, that we ought to continue to question or challenge both our theoretical orientation and the construction of our measures.

Moreover, I argued that to do otherwise risks the promulgation of ethnocentric understandings of quality of life and policies which support them.

The worry over bias is not unknown in the quality of life literature. But in this literature bias typically refers to the situation where respondent answers are taken to be inaccurate. In standardized measures some of the factors that are understood to contribute to this problem include the following: the influence of social norms may tempt respondents to lie rather than admit to something in opposition to expectations; the particular wording of a question may mislead respondents; time constraints may result in the misunderstanding of questions or arbitrary answers; a question's placement in a measure may suggest an understanding of it that researchers did not wish to convey; and a question's response alternatives may not include the answer that respondents would like to make.[133] Individualized measures also face problems of respondent bias insofar as respondents may misrepresent their true feelings regarding their quality of life. For instance, respondents may answer the questions put to them out of habit, they may forget to consider something important to their answer or they may feel bound by outside influences to respond in certain ways.

These issues not only affect the accuracy of respondents' answers to the questions in a quality of life measure, but they also affect the validity of that measure's outcomes. For instance, when respondents lie about their answers or when respondents understand the questions differently than researchers imagined they would then their answers can confound the measure's outcomes. Instead of measuring, say, quality of life the instrument

---

[133] Clark H. H. & Schober, M.F. (1992) 'Asking questions and influencing answers', in J. M. Tanur (ed.) *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, New York: Russell Sage Foundation, pp. 25-6.

in question may be assessing the population's adherence to conventional norms, their risk aversion or their concern for consistency. To make matters more complicated respondent bias does not always show up in tests of construct validity. Instances of respondent bias can be detected by construct validity only when respondents answer contrary to established hypotheses. But respondent bias also occurs when respondents answer *in line* with established hypotheses, but for *different reasons* and sometimes in response to alternatively interpreted questions.

Sometimes respondent bias occurs because respondents understand the questions in a measure differently than researchers envisioned. Although respondent bias can occur for other reasons this is the case on which I will focus my attention in this chapter. In response to and in anticipation of this type of respondent bias researchers who design standardized measures attempt to create questions that are both as clear as possible in terms of the understanding that they want to convey to respondents and as uncontroversial as possible so that respondents will answer truthfully. Put differently, researchers attempt to limit the different ways that respondents will understand the questions posed in the measure. There are many strategies to achieve this goal. With regard to the importance for clarity sometimes the approach is as intuitive as keeping the measure short and simple.

With regard to certain socially sensitive topics researchers sometimes hide the purpose of a question. For example, a question asking you to rate your expected enjoyment in racing motorcycles may appear to be asking you about your preferences for spending your leisure time, but in fact it may be an index of risk-taking.[134] Another tactic is to

---

[134] Streiner, D.L. & Norman, G. R. (2003) *Health Measurement Scales a Practical Guide to their Development and Use*, Oxford: Oxford University Press, p. 90.

change the scaling of a question. Instead of asking parents to rate on a *linear scale* whether they would spank, do nothing or cuddle their child when she won't sleep measurement developers may use the paired-comparison technique in which parents are asked which of two behaviors they were more likely to do in a particular circumstance. The idea is that on a linear rating scale it is too obvious which end of the scale constitutes the "wrong" answer. Paired-comparisons are thought to hide their normativity better.[135]

The stress on the reduction of respondent bias is not surprisingly motivated by its connection to validity: valid measures are measures with relatively little scope for bias. Moreover, because construct validity cannot always recognize respondent bias, claims regarding construct validity presuppose that respondents in fact understand questions and answers as researchers imagined they would. Thus it would seem as though there is an added incentive to constraining how respondents understand and respond to the questions in a measure, namely the accuracy of our validity claims. But while it is true that valid measurement requires researchers and respondents to understand one another, in this chapter I will challenge the idea that validity requires us to narrow the margin for bias by *limiting* the ways that respondents understand questions to those that researchers imagine. Alternatively, I will suggest that the notion of validity compels us to acknowledge and indeed encourage different understandings of these questions.

The concept "respondent bias" as it is used in this literature presupposes that researchers generally know in advance how to understand the questions in a measure and the notion of validity in this context depends on the possibility of conveying this understanding to respondents. In the beginning of this chapter, however, I look at some

---

[135] Ibid, p. 43.

claims of respondent bias and argue that we need not understand these respondents' answers as biased. Instead we might take their alternative understandings of certain questions as opening up new avenues in our conception of quality of life.

To make this argument I look at one theoretical attempt to articulate how respondents understand the questions in quality of life measures. I argue that understanding a question's purpose is essential to our understanding a question and, moreover, when we misunderstand a question or a statement it is often because we misunderstand its purpose. Many instances of respondent bias are the result of difficulties understanding a question's purpose. But I will argue that we—respondents *and* researchers—can only understand or misunderstand the purpose of a question insofar as we understand the question and we understand a question only insofar as we ask it. I suggest that quality of life researchers cannot know in advance how best to understand the questions they ask; rather they must simply ask their questions.

If we cannot know the best understanding of our questions in advance of asking them, then the validity of our measures cannot depend on the possibility of conveying this understanding to respondents. Nonetheless, if we do not know how to understand our questions in advance, then construct validity cannot determine the accuracy of our measures. I propose instead that to ensure the validity of these measures we ought to situate them within a dialogic framework in order to *utilize* the different understandings of our questions and to better grasp the limitations and opportunities of the constructs that our measures embody. In the end I suggest that a measure's validity is dynamic, that while we can incorporate the lessons from different studies into a particular measure, one measure can never substitute for the entire dialogic framework. While measures can be more or less valid the question of their validity if never finished.

I

The NHP and SF-36: Respondent Bias or Respondent Insight?

A. The Nottingham Health Profile

In 1993 Donovan *et al.* published a paper entitled, 'Assessing the Need for Health Status Measures' which examines the propriety of using quality of life measures to determine the health needs of local populations. Donovan and her team query the validity of the Nottingham Health Profile (NHP) by pointing to instances of respondent bias—instances that were not identified in the course of normal validity testing. At the same time, they are keen to emphasize that the issues they raise are not exclusive to the NHP 'but apply to the genre'.[136] Their study proceeds by comparing respondent answers to questions about health taken from interviews with the same respondents' answers in the NHP. Their aim is to discover whether the participants' answers in the measure were accurate given what they said in the interviews.[137]

In this study respondents are asked to fill in part of the General Household Survey, two measures of health and health care developed by the Rand Corporation and the NHP. They are then asked to participate in an audio-taped, semi-structured interview about both their health care and any issues that the standardized measures may have raised.[138]

---

[136] Donovan, J. L. *et al.* (1993) 'Assessing the need for health status measures', in *Journal of Epidemiology and Community Health*, 47: 159.
[137] Ibid.
[138] Ibid.

Donovan *et al.* focus on the NHP because at the time it was the leading generic, standardized quality of life measure on the market. As they remark it replaced the Sickness Impact Profile (SIP) as the leading generic measure because it was perceived to be better in terms of validity and more sensitive in terms of responsiveness to change.[139] The NHP is scaled with categorical judgments in the form of "yes/no" responses to short statements and it measures three dimensions: perceived physical, social and emotional health.

During the interview section of the study Donovan *et al.* find three areas of discrepancy between respondents' answers on the NHP and their interview responses. These discrepancies suggest to Donovan *et al.* that the NHP scores are biased. In the first instance, Donovan and her team propose that this bias is due the NHP's categorical response options: the yes/no alternatives are too limiting for respondent answers.[140] Donovan *et al.* offer the following examples to support their diagnosis:[141]

Things are getting me down: yes/no
I have pain at night: yes/no
I have unbearable pain: yes/no
I take tablets to help me sleep: yes/no

To the first question, one interviewed respondent answers, 'I won't let them if I can. Can I put sometimes?'[142] Here the categorical scaling of the question does appear to be limiting. The respondent's answer perhaps reflects the need to implement some kind of continuous judgment scale in the NHP which would allow for a larger range of response options.

---

[139] Ibid, 161.

[140] Ibid, p. 159.

[141] Although the 'questions' on the NHP, and many other standardized questionnaires, are not explicitly in the form of an interrogative I will refer to them as 'questions'.

[142] Donovan, J. L. *et al.* (1993) 'Assessing the need for health status measures', in *Journal of Epidemiology and Community Health*, 47: 159.

But changing the response scale will not solve all the NHP's problems. In the other three questions from above respondents have rather different sorts of difficulties; difficulties which reflect more than a problem with the categorical response options. For instance, to the question about pain at night one respondent replies that she doesn't so much have pain, but discomfort.[143] Here her confusion seems to be whether discomfort counts as pain. Similarly, with respect to the question about unbearable pain one respondent replies that the pain is only unbearable when she has a backache.[144] We might say that it is unclear to her if an occasional backache is sufficient to count as unbearable pain in general.

Finally, to the question about taking tablets to help with sleep, one respondent answers, 'I take tablets at night for the cramp and they help me sleep. What do I put there?'[145] In this case the respondent appears to be unsure what the question is asking her: is the question a causal one—does she take tablets *because* they help her sleep—or is it interested in *whether* she takes them, for whatever reason? In this situation we might say that the respondent does not understand the question's purpose or meaning—*why* is she asked this question; what does the NHP want to know?

According to Donovan *et al*:

> The statements [in the NHP] made people think about aspects of their health, but then constrained their responses...The forcing of responses into predefined categories negated people's desires to negotiate the meanings of health and illness.[146]

Earlier they comment that although the yes/no dichotomy of the NHP allows for straightforward analysis, it 'does not allow people to express what they really feel'. Thus if

---

[143] Ibid.
[144] Ibid.
[145] Ibid.
[146] Ibid.

respondents are forced to answer simply yes or no their answers may be inaccurate.[147] I agree with Donovan and her team that categorical scaling constrains responses and I also agree that it does not allow respondents to negotiate what, for instance, pain might mean in certain contexts. Moreover, to the extent that respondents are typically left to answer these questions on their own and researchers analyze these answers on *their* own the NHP might not allow individuals to express what they really feel.

But I disagree with the conclusion that Donovan *et al.* draw from these difficulties, namely that the categorical scaling of the questions is at fault and, moreover, that a solution to these problems might just require an adjustment to the scaling options. I suggest that the problem in these examples is not solely with the yes/no answers. To be sure, scaling is an important aspect of measurement construction, but the primary problem in these cases is that respondents are unsure about what various terms or, in some cases, entire questions in the NHP mean. Once it is clear how we understand, say, unbearable pain, then the yes/no format of the NHP need not be a problem or at least not a major problem.

The more major problem that these examples illustrate is the problem that all standardized questions have, namely, confusion over how to understand the questions. To this end the respondents' confusion regarding how they ought to understand the questions in the NHP may be instructive: in this context *should* discomfort count as pain?; *should* occasional pain count as unbearable pain?; *is* the sleeping pill question a causal question or a whether question?

The second difficulty with the NHP is what Donovan and her team refer to as a contradictory or arbitrary filling out of the questions. They provide two examples. In the

---

[147] Ibid.

first case, one respondent replies to the question, 'I find it hard to bend' by saying 'I do find it hard to bend, but I'm not ticking yes there.' This response is interpreted by Donovan *et al.* as a contradictory answer—in other words, they think the respondent's answer to the question is a lie. This result, they suggest, is due to the pressure that respondents feel to give socially acceptable responses.[148] But while this respondent's answer *may* represent a contradiction it does not necessarily do so. To see why not consider Donovan *et al.*'s second example in which they cite the following question and response: 'Worry is keeping me awake at night'; 'Well yes, but it's only stupid things. I lie awake thinking. I'll put no because I'm just being stupid.'[149] Unlike the bending example, this respondent explains why she has decided to mark 'no': yes she worries, but she only worries about stupid things because she's 'being stupid'.

Although Donovan and her team interpret this second example as yet another instance of contradiction we might understand it differently. We might take this respondent to understand her worry as different from the sort of worry in which the NHP is interested. The worry the respondent experiences is petty and 'stupid'; the worry in which she understands the measure to be interested involves, say, existential doubt and acute torment. To be sure, the respondent could be wrong about what the NHP is after, but that is not the point. The point is rather that we could paraphrase this respondent's answer in terms of Donovan *et al.*'s bending example: 'I do lie awake worrying at night, but I'm not ticking yes.' and nonetheless understand it as an appropriate response given a particular understanding of what she took the question to mean. In these cases the respondents'

---

[148] Ibid.
[149] Ibid.

answers are contradictory only if we cannot imagine ways of understanding the questions other than our own. Conversely, we might assume that these answers are not contradictory and, indeed, we might learn something about quality of life if we try to figure out what questions the respondents take themselves to be answering.

The point in trying to imagine different ways of understanding questions is clear in the third area that Donovan *et al.* examine. Here they think respondents misunderstand the purpose or meaning of a question and thus respond according to their misunderstanding. For instance, to the question, 'I find it hard to reach for things', one respondent answered, 'I do find it hard to reach for things, yes, because I am short.'[150] This question however, was not geared to eliciting facts about the respondent's height; rather it was meant to elicit facts about her health. But, nonetheless, given this respondent's answer we might ask where and when height does become a health issue?

Donovan make several important points. For instance, if respondents understand questions differently from the researchers who analyze them, then the conclusions of a study can suggest morbidity or a certain kind of morbidity that does not actually exist—as in the example of the respondent who answered positively to the statement about whether she found it hard to reach things. Conversely, a study can also underreport morbidity. For example, respondents might mistakenly take 'unbearable pain' to mean 'unbearable pain all the time' and tick no despite having it occasionally. Although I agree with Donovan *et al.* that the validity of the NHP is threatened when these kinds of misunderstandings are generalized in the population under study, I disagree over why it is threatened. I suggest that these examples do not illustrate instances of bias so much as they illustrate

---

[150] Ibid.

opportunities where we might learn more about quality of life in the context of the NHP. The validity of the NHP is threatened, however, when we fail to recognize these opportunities.

Towards the end of their article Donovan and her team observe that just as the NHP replaced the Sickness Impact Profile (SIP) because of its supposed increase in validity, at the time of their study the Short Form-36 (SF-36) was beginning to replace the NHP for the same reason.[151] But Donovan *et al.* suggest that the assumption of the SF-36's increase in validity over the NHP is unwarranted.[152] With this warning in mind I now turn to a study published in 2002 by Sara Mallinson entitled, 'Listening to Respondents: A Qualitative Assessment of the Short-Form 36 Health Status Questionnaire'.

B. The Short-Form 36

Like Donovan and her team Mallinson is worried about the validity of the SF-36 in spite of its respectable showing according to classical validity testing. Specifically she is worried that people do not understand the questions posed to them in the manner that researchers anticipated. As a result she worries that their answers are misinterpreted by researchers analyzing the data.[153] To make her point Mallinson conducted a study of the SF-36 with a very similar design to the study done by Donovan *et al.*

In Mallinson's project the respondents are all 65 years of age or older and reside in the Northwest of England. Moreover, at the time of the interview they had all been recently

---

[151] Ibid, p. 161.

[152] Ibid.

[153] Mallinson, S. (2002) 'Listening to respondents: a Qualitative Assessment of the Short-Form 36 Health Status Questionnaire', in *Social Science & Medicine*, 54: 11.

referred to a community physiotherapy or occupational rehabilitation program. All of the respondents in the study were asked questions that covered socio-demographic characteristics, self-reported health problems and expectations of treatment; they were also asked to complete the SF-36 and two global health evaluation questions. Mallinson conducted two audio-taped interviews with the respondents as they answered the questions from the SF-36 with the aim of better understanding the ways in which the respondents interpreted the measure's questions. One interview was conducted before their physiotherapy began and one was conducted six months later.[154]

The SF-36 is a generic quality of life measure with 36 questions scored in eight dimensions. It is now the most widely used measure of its kind in the United Kingdom and it has been translated into 45 languages for use around the world. Moreover, there is an extensive literature that suggests it is one of the most reliable and valid short-form measures available.[155] For this study Mallinson focuses on the responses from two of the eight dimensions: the 'Physical Functioning Scale' and 'General Health Perceptions'.[156] In the Physical Functioning dimension respondents are asked how their health limits them in a variety of everyday activities. They are asked to respond to each question with one of three answers: 'Yes, limited a lot'; 'Yes, limited a little'; 'No, not limited at all'. In the General Health Perceptions dimension respondents are first asked to assess their general health as 'excellent'; 'very good'; 'good'; 'fair'; or 'poor'. They are then given four statements

---

[154] Ibid, p. 13.
[155] Ibid.
[156] Ibid, p. 14.

about their general health and asked to rate them on a Likert scale as 'definitely true'; 'mostly true'; 'not sure'; 'mostly false'; or 'definitely false'.[157]

As Mallinson makes clear in the beginning of this paper it is the problem of understanding meaning—how researchers hope the question is understood and how the respondents in fact understand it—that drives her research.[158] In what follows I will therefore focus on those findings that make this problem most clear. In the Physical Functioning dimension Mallinson found what she calls 'vague questions'. In these cases when respondents were asked if they could walk half a mile they wanted to know whether that half mile was up-hill or on the flat; whether it was at a fast or leisurely pace. Consider, for example, what one respondent had to say:

> I can walk down to the garden centre but there's no way I could get back because it's up-hill, and as soon as I, I can't walk up that hill so it depends which, if you're talking about on the flat, slowly, not talking or carrying anything...I can walk around the shopping precinct and round the supermarket because you're going slowly and you're stopping and looking at things and you're not talking to anybody.[159]

We have seen this problem of 'vague questions' before in Donovan *et al.*'s study. Recall the questions about pain in which respondents, although familiar with pain were unsure what counted as pain in that context. In the SF-36 Mallinson found this problem in questions not only about walking, but also in questions about lifting and bathing.[160]

Turning now to the General Health Perceptions dimension the first question respondents encounter is a question asking them to rate their general health. Mallinson

---

[157] Ibid, p. 14. Likert scales are bipolar scales in which the neutral option is located in the middle and the extremes are at either end. See Streiner, D.L. & Norman, G.R. (2003) Health Measurement Scales a Practical Guide to their Development and Use, Oxford: Oxford University Press, p. 36.

[158] Ibid, p. 12.

[159] Ibid.

[160] Ibid.

discovered that respondents understood the meaning of health in a myriad of ways and attributes these differences to what she calls 'comparative' problems.[161] For instance, with regard to the question, 'In general would you say your health is: excellent; very good; good; fair; poor', some respondents understood this question as asking 'would you say your health *as opposed to others in the population* is....'; some respondents understood it as asking 'would you say your health *now as opposed to other times in your life* is...'; and still others understood it as asking 'would you say your health *as opposed to others your age* is...'[162] Typically, when respondents read questions they must decide for themselves what it means. Thus, asking a question like this one is tantamount to asking at least three different questions without any idea which one a particular respondent is answering.

Finally, Mallinson looks at problems involving what she calls 'inclusions and exclusions' in conceptualizing health.[163] These problems resemble the problems found in some of the examples from Donovan *et al.*'s study, namely the examples about bending, worrying and reaching things. Recall that in those examples I suggested that respondents' interpretations of questions sometimes lead to surprising and at times confusing answers which only begin to make sense given the context of the respondents' understanding. In Mallinson's example, the SF-36 asks 'In general how would you say your health is?' and in response one individual answers, 'My health is good. It's the spinal atrophy that's the problem'. As Mallinson notes, this man separated his chronic health problem from his sense of being a "healthy" person, which is what he thought the surveyor was interested

---

[161] Ibid, p. 18.
[162] Ibid, pp.18-19.
[163] Mallinson, S. (2002) 'Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire', *Social Science & Medicine*, 54: 19

in.[164] Another way to put this is that he understood the purpose of the question differently than perhaps the researcher imagined.

A similar problem occurs with another question from the SF-36 which asks respondents whether they think they get ill more easily than other people. In response to this question individuals in Mallinson's study differed in how they interpreted the term "ill". For instance, many thought it referred to everyday problems and thus excluded chronic health problems as the following two responses illustrate: 'I'm not bodily ill, its mobility more than anything isn't it?'; 'I don't think so—apart from my diabetes and it doesn't bother me that, I've got used to it.'[165] These answers Mallinson suggests are not the result of social desirability or denial, but rather logical responses to different interpretations of the question. Mallinson sums up the problem well:

> ...they include and exclude problems in different ways and may unexpectedly weigh all kinds of information before answering. In effect, people are responding from different premises to each other and from the surveyor. This inevitably affects respondents' intentions/meanings in selecting a particular response option and makes it difficult for the surveyor to interpret their answers.[166]

In many ways Mallinson's observations of respondents' answers to the SF-36 resemble my interpretation of Donovan *et al.*'s observations of the NHP. In both cases we recognize the inherent difficulty respondents have in understanding what researchers might mean with particular terms in the questions. In an interview setting respondents tend to ask questions about the questions, but without an interlocutor they are left to guess. As I do, Mallinson also emphasize cases where respondents do not understand the purpose of a

---

[164] Ibid.
[165] Ibid.
[166] Ibid.

question in the same way as the researchers do. In such cases respondents might ask for clarification or they might provide answers that at first do not make sense or seem unlikely. In these latter instances Mallinson suggests, and I agree, that we need to understand the context of the respondents' answers before we dismiss them.

Unlike Donovan *et al.* Mallinson is more willingly to take account of the respondent's point of view when judging the accuracy of their answers. Although she is worried about the validity of the SF-36 she often characterizes this problem as one where researchers misinterpret respondent answers as opposed to Donovan *et al.* who tend to suggest that the problem with validity resides in respondents' interpretation of the measure's questions. Perhaps as a result of her sensitivity to the way respondents interpret questions Mallinson is also more willing to learn from respondents and we might even say that she recognizes—as Donovan *et al.* do not—some of the opportunities that respondents present for the improvement of our measures.

But what lessons or opportunities does Mallinson draw from her work with respondents and the SF-36? At the end of her article Mallinson concludes that the issue of understanding meaning is 'absolutely central to understanding subjective views' and that without further research into how respondents' understand survey questions it will be difficult to establish the validity of subjective assessment.[167] We might say that what Mallinson takes from her work on the SF-36 is that our current claims regarding the validity of subjective assessments are premature: respondents often understand questions differently than researchers imagined they would. Unless we have a better grasp of how respondents do understand the questions in our measures we cannot confirm the measures'

---

[167] Ibid, p. 20.

validity—and perhaps when we do better understand this process we will find that our understanding of these measures and our expectations regarding their validity will have changed. In any case it would appear that we need a better account of how respondents interpret questions.[168] In the next section I turn to one theoretical attempt within the quality of life literature that tries to provide one.

II

Understanding Questions and the Principles of Discourse

In J. Tanur's book *Questions about Questions* Herbert Clark & Michael Schober try to explain how respondents understand questions in a chapter entitled 'Asking Questions and Influencing Answers'. Their task is to outline some of the ways we typically come to understand our interlocutors in normal discourse and to show how the lack of these normal conversational cues affects the answers respondents give in standardized measures.

Clark & Schober begin their discussion with the principle of 'common ground'.[169] Here they discuss how words become meaningful against a shared context; change the context and you often change what we understand a word to mean. Making a similar claim, Larry Wright asks us to consider the sentence, 'The cat is on the mat.' Typically we conjure ideas of an animal on a rug. The context that provides this interpretation is perhaps a domestic setting. But we can also imagine a different context in which this statement

---

[168] Although it is difficult to know how often the kind of misunderstanding illustrated in the above studies occurs, the frequency with which it occurred in these cases indicates that it may happen often enough to affect the overall results of a measure. Certainly the authors of these studies believe the problem to be widespread and significant.
[169] Ibid, pp. 17-18.

might arise—for instance, a construction site. In this new context the sentence above takes on a new meaning, namely that a piece of machinery is parked on a blasting mat.[170]

It is the awareness of the common ground that we share with others that allows us to use ambiguous language unconsciously, but with meanings we expect our interlocutors to grasp. Perhaps this is most obvious in the frequent use of indexicals like "here" and "now". Common ground, Clark & Schober write, '...is essential in interpreting everything people say'.[171] In some of the examples of questions and answers from the previous section we can see the consequences of an absence of common ground. For instance, when respondents were unsure about the context of a question they often registered confusion over what a term in the question meant. Is unbearable pain constant pain or occasional pain? Is the mile uphill or on the flat?

Clark & Schober associate two further principles with common ground: the 'accumulation of common ground' and 'grounding'.[172] As Clark & Schober suggest some kind of common ground is necessary for every conversation, but as a conversation progresses conversation partners typically accumulate more common ground through what Clark & Schober call the process of grounding. Grounding occurs as we assure one another that we understand what has been said. For instance, a speaker is encouraged that she has been understood when her interlocutor nods her head, displays positive facial expressions,

---

[170] Wright, L. (2001) *Critical Thinking an Introduction to Analytical Reading and Reasoning*, Oxford: Oxford University Press, p. 5.

[171] Clark, H. H. & Schober, M. F. (1992) 'Asking questions and influencing answers', in J. M. Tanur (ed.) *Questions about questions: Inquiries into the cognitive bases of surveys*, New York: Russell Sage Foundation, p. 18. This is the case not only where "common ground" leads us to unconsciously expect that our meaning is understood, but also in cases where we recognize that a "common ground" is missing. This recognition often explains what may seem to others who share our "common ground" as unnecessary or lengthy description.

[172] Ibid, pp. 19-20 &24-5.

paraphrases what was said or shows her understanding in what she says next.[173] Insofar as interlocutors understand one another they establish new common ground—an *intersubjective* ground

Because standardized measures lack the mechanisms interlocutors use to ground new information respondents must make assumptions about the meaning of the questions in the measure without the give and take of normal conversation. Clark & Schober write that respondents assume that the researcher chose her wording so that the meaning would be obvious and as a result respondents impute whatever understanding seems right to them.[174] Respondents, however, can differ from one another and from researchers in their understanding of questions. Such a difference was illustrated nicely in the example from earlier when the respondent understood, 'I find it hard to reach things' as 'I find it hard to reach things with respect to my height'. In everyday dialogue the respondent's answer 'yes, because I am short' would have been met with a clarification perhaps something like, 'No, I meant do you find it even harder to reach things after your by-pass surgery'.

In addition to clarifying questions for respondents the process of grounding might also illuminate areas of ambiguity or vagueness in places that the *researcher* did not anticipate. For instance, recall that in the SF-36 Mallinson finds that the question on general health gives rise to at least three different interpretations depending on the contrast respondents use to understand it. It may be the case that the measure's developers did not recognize the ambiguity of "health" in this question. Nonetheless, the respondents' different understandings of "health" might not only fruitfully help to bring this ambiguity to

---

[173] Ibid, p. 25.
[174] Ibid, p. 23.

light, but also help researchers learn new ways of understanding health. Particularly suggestive in this regard, is the respondent's answer, 'My health is good. It's the spinal atrophy that's the problem'.

In normal conversation different contexts or points of view can enrich a discussion. These differences are capable of *enhancing* a conversation in part because the process of grounding helps to prevent interlocutors from talking at cross-purposes. When individuals are at cross-purposes they misunderstand each other because the aim or motivation of their respective discourse is opaque to their interlocutor(s). To be sure, even the most careful and well grounded of discussions are subject to what we might call the problem of purpose. Nonetheless, this problem is exacerbated when individuals are not able to confirm their understanding with a nod or a paraphrase; when they are left to understand and then answer questions on their own and others determine the meaning of these answers on *their* own. In fact I suggest that the problem of purpose is the general difficulty with the NHP and the SF-36's validity: both researchers and respondents are at cross-purposes in their understanding of the questions and answers in the measures.

We might say that in normal conversation the process of grounding helps us to avoid the difficulties associated with cross-purposes by aiding in the establishment of a common purpose. Indeed in their attempt to make more precise what is involved in their understanding of common ground Clark and Schober stress the importance of a 'common purpose'.[175] They maintain that common purposes are important to understanding because purposes shape the direction of a conversation as well as what people mean by what they

---

[175] Clark, H. H. & Schober, M. F. (1992) 'Asking questions and influencing answers', in J. M. Tanur (ed.) *Questions about questions: Inquiries into the cognitive bases of surveys*, New York: Russell Sage Foundation, p. 22.

say. Clark & Schober write that the evidence suggests that common purpose is essential in figuring out a speaker's meaning.[176] We might say that without understanding the purpose of a statement or a question an individual cannot understand what a speaker means. If this is the case, then the general task for respondents to is to determine the purpose of the questions in our measures.

To make clearer what exactly is involved in the idea of common purpose, recall Wright's example: 'The cat is on the mat.' Here, as I discussed earlier, the context helps to determine whether the sentence refers to an animal or a machine or something else. But the context should also give us insight into *the aim* someone has in uttering this sentence—otherwise the sentence is still ambiguous. For example, if you're in the kitchen and your partner walks in and suddenly says 'the cat is on the mat' you may know that he's referring to your pet Cooper, but you may not understand the point of what he's saying. Was he looking for the cat? Is the mat off-limits and he is expressing frustration? Is his statement code for, 'now's a good time to give Cooper his meds'?

Wright puts the importance of purpose a bit differently than Clark & Schober when he writes that, '...the main thing a context supplies to help us understand each other is *motivation*'.[177] Moreover, Wright suggests that we understand the motivation or purpose of a conversation when we understand the *question* to which the discussion is the *answer.*[178] Thus we understand the significance of 'the cat is on the mat' when we understand it as the answer to one of the following questions: 'Where is the cat?'; 'What is Cooper doing?';

---

[176] Ibid, p. 22.

[177] Wright, L. (2001) *Critical Thinking an Introduction to Analytical Reading and Reasoning*, Oxford: Oxford University Press, p. 142.

[178] Ibid. Although here Wright is specifically referring to how we understand the motivation of an argument, the idea applies generally. For a discussion of its more general application see, Gadamer, H. (2003) [1975] *Truth and Method*, 2nd edn, trans. J. Weinsheimer & D. G. Marshall, New York: Continuum Press, pp. 370-1.

'When should we give him his meds? Depending on the implicit question posed your reaction will be different. If Cooper was lost and now he's been found, you might express relief; if he's lying on your Persian rug, you might run off with the squirt bottle; or if he's half asleep and now is a good time for the meds, you might take them from the kitchen counter and walk to the mat.

But Wright's suggestion that we understand motivation when we understand the question to which a sentence or discussion is the answer is misleading if we then think that the question which articulates the motivation (Where is the cat? What is Cooper doing? Should we give him his meds?) can stand alone; that we can understand this question without further contextual support. We need look no further than one of the examples from Donovan *et al.*'s study to see that this is not the case. Recall the example in which the woman is unsure whether the question about taking tablets to sleep is interested in *why* she takes the tablets or simply *whether* she takes them. In this example and others like it the yes/no answers are not made clear even in light of their respective questions. For these answers to make sense I suggested earlier that respondents need to better understand the *question* itself. If we apply Wright's question/answer structure to this example, then in order to understand what the yes/no answers signify the respondent needs to understand the question about taking tablets to sleep as the *answer* to a further question. But what might this question be?

The idea that we understand the purpose of individual questions only when we understand these questions as answers to another question follows the cascading design of standardized measures nicely. Recall from Chapter 1 that we begin the construction of standardized measures with an idea of what we want to measure—the research construct.

For instance, say we want to measure depression and its relation to quality of life. This construct represents the broad question that the measure aims to answer: 'What is the quality of life of those suffering from depression?' In order to design a measure that will answer this question researchers generally use factor analysis or some other method to determine the relevant dimensions of a construct. For example, in the case of quality of life and depression we might find that social and emotional functioning are two of the relevant dimensions. Each one of these dimensions is then broken down into further questions which are then posed to respondents, for example the question from earlier about taking tablets to sleep

I suggest that it is the question that the research construct implies that provides the purpose with respect to which the individual questions in a measure ought to be understood. If understanding a question's purpose is essential to understanding questions and validity requires that researchers and respondents understand these questions in the same way, then we—researchers and respondents—need to understand the research construct and we need to understand *it* in the same way.

To be sure, recognizing the point of a question is something we do often and do with ease. But standardized health assessments present at least two obstacles which make the recognition of a question's purpose more difficult. First the quality of life literature acknowledges that constructs such as quality of life and well-being are contentious and difficult to define.[179] Hence, not only may researchers and respondents not share a common ground; they may also have trouble establishing one. Second, as I discussed in the

---

[179] Mallinson, S. (2002) 'Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire', *Social Science & Medicine*, 54: 18.

introduction of this chapter, measurement developers sometimes try to hide or neutralize the purpose of their questions in order to deal with the confounding that comes from social norms. Hence, researchers may not want respondents to connect, for instance, the question of disturbances in sleep to the possibility of depression in part because they wish to assess psychotic disturbance. Moreover, sometimes researchers conceal the purpose of a question because they want to use the data from one measure for purposes other than those for which it was designed. For example, the United Kingdom's National Audit of Nasal Polyposis and Rhinosinusitis was originally designed to distinguish between the levels of performance among different hospitals, but once the data was collected it was also used for prognostic modeling.[180]

Thus it is sometimes the case that researchers have good reasons to discourage respondents from making the connection between a particular question and its relation to the measure's construct. But if Clark & Schober are right and understanding the purpose of a question is essential to understanding its meaning and if understanding meaning is essential to validity, then we need to find a way to overcome these difficulties.

Clark & Schober end their discussion with the following remark,

> To understand surveys and the data they produce, we must see survey interviews as a type of discourse...Only then will we resolve many of the puzzles of survey design.[181]

But Clark & Schober do not tell us how understanding the content of a measure as a type of discourse will lead us to these resolutions. In their discussion they are clear that the

---

[180] Personnel correspondence with John Browne, 28 November 2005.

[181] Clark, H. H. & Schober, M. F. (1992) 'Asking questions and influencing answers', in J. M. Tanur (ed.) *Questions about questions: Inquiries into the cognitive bases of surveys*, New York: Russell Sage Foundation, p. 43.

principles people use in ordinary language do not get left behind when answering the questions in a quality of life measure and they illustrate how an understanding of these principles can lead us to explain and in some cases predict the kinds of respondent bias that typically plague researchers. Still they do not tell us how this knowledge should help us to improve our measures.

Toward the end of this chapter I will begin to suggest my own way of thinking about how we might best understand the outcomes data from quality of life measures. But in order to help make a case for it I now turn to develop further one of the obstacles we face when we attempt to understand questions in our measures: the contentious nature of quality of life constructs. In this next section I look at a recent quality of life study on cataract patients to highlight one of the consequences that a contentious research construct has for the validity of quality of life measures.

III

VF-14 and the Problem of Purpose

In the June 2005 issue of the *Canadian Journal of Ophthalmology* Lorne Bellan attempts to answer the following question, 'Why are patients with no visual symptoms on cataract waiting lists?' This question is motivated by a case in Winnipeg, Canada where 30% of patients placed on waiting lists for cataract surgery had a score on the Visual Function Index (VF-14) of 91or more out of 100. A score of 100 on the VF-14 indicates no visual complaints. Moreover, there are multiple studies that indicate that measures of functional impairment are the best indicators of the degree of impairment and the potential gain from surgery.[182] But if the VF-14 is a powerful indicator of impairment and potential gain from surgery, then why do 30% of the patients placed on cataract waiting lists in Winnipeg have scores from this measure which indicate that they do not need the surgery?[183] This is the question that the Winnipeg media posed to tax payers and Bellan's study tried to provide an answer to their question.[184]

The VF-14 is required of all individuals scheduled to undergo cataract surgery in Winnipeg. It comprises a series of questions asking respondents about the degree of difficulty they have performing common visual tasks such as watching TV or driving a car. Respondents answer the questions on a continuous scale with four responses ranging from no difficulty to unable to perform. Between January and May 2002 prospective cataract patients who completed the VF-14 and whose scores reported *no* functional impairment were asked to participate in Bellan's study. 149 individuals agreed and were then asked three questions: 1) Are there any other problems with your vision that you are experiencing that I haven't asked you about?; 2) Please tell me the reason, as you understand it, why you have been scheduled to have cataract surgery?; 3) What activities do you think will be easier for you after your surgery? Of the 149 patients, 108 were having surgery because of symptoms not specified on the questionnaire, 28 were doing it purely based on the doctor's advice and 13 were asymptomatic.[185]

In January 2003 the same patients were contacted again to assess their satisfaction with their surgery. They were asked four questions: 1) How satisfied were they with their

---

[182] Bellan, L. (2005) 'Why are patients with no visual symptoms on cataract waiting lists?', *Canadian Journal of Ophthalmology*, 40: 434.
[183] Ibid.
[184] Ibid, p. 434.
[185] Ibid, pp. 434-5.

vision in the eye that had undergone surgery? 2) Had they found that their vision had been more impaired than they thought before surgery? 3) Did they feel that their vision had improved after cataract surgery?; 4) Would they be willing to repeat this type of surgery again?[186] Out of the 149 original participants 105 had completed their surgery at the time of this second round of questions. Of these 105 participants 85% were very or extremely satisfied with their surgery; 75% felt that their vision had markedly improved; and 94% were willing to repeat the procedure. Only 9% reported being unsatisfied or said their vision had not improved or that they would not repeat the surgery.[187]

Bellan concludes from these results that most of the patients who had no functional impairment according to the VF-14 did in fact have a large enough a degree of impairment that surgical correction created significant benefit. Thus the VF-14—a measure of subjective assessment—underreports morbidity; we might say that the VF-14 gives us false negatives when it comes to determining an individual's need for cataract surgery. In fact Bellan concludes that, 'the VF-14 cannot reliably and accurately identify all patients who are likely to benefit from cataract surgery'.[188]

But notably he goes on to say that this conclusion regarding the VF-14's deficiency with respect to cataract surgery is consistent with the findings of the original study in which the VF-14 was developed and validated: no one ever suggested that the VF-14 should be the only measure used to determine the need for cataract surgery.[189] Thus, the VF-14's inability to identify all patients in Winnipeg likely to benefit from cataract surgery is not

---

[186] Ibid, p. 435.
[187] Ibid.
[188] Ibid, p. 437.
[189] Ibid.

evidence of *its* invalidity. Rather the measure underreports morbidity because it was not designed to arbitrate in these kinds of situations. In this example we see another way in which quality of life measures can provide invalid results, namely when the measure is applied to the wrong context. The problem in the Winnipeg case is not with the VF-14 itself, but with the decision to use it to determine benefit from cataract surgery.

Bellan, however, does more than suggest that the sole use of the VF-14 in the Winnipeg study was inappropriate. In addition, he suggests that anyone familiar with the original design of the VF-14 should have known in advance of the media story that the VF-14 would not accurately discriminate among cataract patients. But how could they have known? Before Bellan's study we did not know that this cohort's disease state would significantly affect our assessment of functional impairment. I suggest that the contentious and difficult nature of quality of life constructs stems not only from disagreement about what a particular construct entails—does a good quality of life entail certain biological prerequisites? —but also from the fact that before we apply them to a particular situation these constructs are inherently vague.

Here we might draw an analogy between quality of life constructs and moral concepts such as courage. Such moral notions are traditionally difficult to define in part because the actions that constitute instances of courage differ depending on the practical situation in which one finds oneself: sometimes the measure of one's courage involves standing and fighting; sometimes it involves walking away. Indeed, we might say that the more ways we learn to measure courage correctly the more nuanced and sophisticated our understanding of it becomes. But, we cannot always determine in advance the features of a situation that might affect that measure; it is only in a concrete context that "courage" takes on a definite meaning and only through multiple measures that its meaning is refined.

Likewise, we understand what constitutes quality of life constructs differently depending on the concrete situation to which they are applied. 'Functional impairment' like 'quality of life' and courage is a multidimensional construct. Moreover, we learn more about it by learning to measure it in different ways. The proliferation of disease and culture specific measures in the quality of life literature attests to some of the different ways we have learned to understand some of these constructs.

In deciding to use the VF-14 in the Winnipeg study a charitable interpretation would suggest that the best evidence at the time of the design suggested that as a measure of functional impairment the VF-14 was the best indicator of impairment and gain from cataract surgery. It was not unreasonable to suspect that it might be able to discriminate among those who would or would not benefit from cataract surgery. But in the course of its use and through Bellan's research some of the limitations of the VF-14 as a measure of functional impairment crystallized. Although these precise limitations are consistent with the developmental study of the VF-14 they were not prescribed by it. On this interpretation the study designers had no way of knowing in advance that the VF-14 would overlook 30% of the people with visual symptoms. They *may* have had reason to be cautious since they were using it in a slightly different context than it had been used before, but it seems that they *were* cautious in that the 30% of people who did not show any impairment in terms of the VF-14 still received surgery.

I suggest that what we ought to take from the study on the VF-14 is not that researchers made an obvious mistake in applying it to the Winnipeg situation, but that in applying it to this context not only did they learn more about the limitations of the VF-14, they also learned more about the features of functional impairment; put differently, they made our understanding of the construct more determinate. The vague nature of quality of life constructs suggests that we can always learn more about these constructs and we learn more about them by applying them to new situations. This is essentially the lesson of construct validity—a construct is never valid; we can always be surprised; we can always learn more about it and we learn more by applying or testing it in new contexts.

IV

Quality of Life Measures and the Hermeneutic Circle

The vague character of the research question or construct may not be particularly surprising. Much of the quality of life literature acknowledges the contentious and difficult nature of the constructs and the literature on construct validity appears to appreciate at least some of these difficulties. However, what is surprising is the notion that the questions in standardized measures of quality of life become clearer as we break them down into smaller chunks. In other words, even if it is not clear what functional impairment means, the questions in various dimensions of functional impairment *are* supposed to be clear. But as I discussed earlier an important aspect of understanding a question is our understanding of its purpose. When a question's purpose is as contentious and vague as the constructs of quality of life, health and functional impairment then we cannot expect a measure's questions to be completely clear—and by clear I mean we cannot expect everyone to understand them the same way and we cannot assume that alternative understandings of these questions are inaccurate.

We are now in a position to see that the validation of quality of life measures faces a challenge in regard to the understanding of their questions—a challenge with the structure of a circle. For as I have developed the problem, in order to provide valid information in a quality of life measure respondents and researchers must understand the questions in the same way, but to understand these questions we need to understand their purpose. I have suggested that the purpose of these questions is given by the research construct (functional impairment), and I argued that the research construct is understood only insofar as the measure (VF-14) is applied to various contexts—insofar as respondents in different contexts complete the measure and provide important information.

The question facing quality of life researchers regarding the validity of these measures is not only how we can ensure that everyone involved understands the questions in the same way—how we can avoid being at cross-purposes—but also how we should best understand these questions; how should we best understand their purpose? Put differently, What *is* quality of life?; What *is* functional impairment?

I suggest that the resolution of this dilemma first requires that we reconceptualize our measures. To this end I propose that we adopt one solution to the hermeneutic circle found within studies of textual analysis and apply it to quality of life research. This solution situates a text or text-analogue—in this case a measure—within a dialogic framework. Although I will discuss this framework at more length in the next chapter, in a dialogic framework a text is understood as a partner in a dialogue regarding how we ought to understand the text's subject matter—in this case the subject matter is whatever construct the measure is supposed to assess. In a dialogic context a text or in our case a particular kind of measure can be one of many partners in the dialogue regarding the subject matter. These partners can include other studies on quality of life as well as the researchers who analyze measurement outcomes. Different measures as well as different researchers illuminate different aspects of a construct and thus enhance our understanding of it. A dialogic framework aids in the validity of our measures because by better understanding a measure's construct we in turn better understand the meaning of the questions and respondent answers in our measure.

Within a dialogic framework quantitative measures of quality of life are understood to provide information which *contributes* to our questions about, and our understanding of quality of life; they are not understood to provide unambiguous answers to questions about quality of life. One of the consequences of understanding quality of life measures within this larger context is that we can take other sources of information, for instance the studies I discussed by Donovan *et al.* and Mallinson, not as evidence that our quantitative quality of life measures are invalid and thus in need of methodological improvement, but as contributions to our continued understanding of the questions and answers in a measure and thus quality of life itself. Moreover, we can understand Bellan's study of the VF-14 not as an anomalous exercise, which was required in order to defend spending on healthcare. But rather we can understand it as part of the essential dialogue needed in order to better understand our measures and thus the meaning of our constructs.

As with any progressive research agenda the results of a dialogue between a variety of studies and researchers should be incorporated into future research in quantitative measurement. For instance, certain questions within a measure may be expanded to include a contrast class; other questions may be eliminated or redirected. Moreover, we may understand these revisions as improvements on the *validity* of the measures. But it is important to recognize that given the circle which the validation of quality of life measures presents a measure's outcomes can never be properly understood outside the dialogic framework. The changes we make to the questions in a measure and our increasingly enhanced understanding of a particular construct are not steps towards an independent measure, but rather part of a better and always evolving understanding of quality of life. This recognition means that although we may improve on the validity of our measures the process of validation is dynamic.

Furthermore, the fact that the validation of quality of life measures involves a circle means that construct validity is unable to judge the accuracy of our measures. As I discussed earlier claims of construct validity presuppose that respondents and researchers understand a measure's questions and answers in the same way. Not only do respondents and researchers often not understand the questions and answers in the same way, but researchers cannot be sure how best to understand these questions or their constructs until respondents answer questions about a construct. This indeterminacy of our questions is also a problem for construct validity. For if we rely on construct validity in spite of our inability to determine the meaning of questions in advance, then answers which appear to bias a measure may in fact be the result of alternatively interpreted questions which shed light on our understanding of the construct; answers which appear to confirm our hypotheses may, however, be the result of questions understood in such a way as to undermine our them.

Toward the end of the next chapter I explore further how a dialogic context can aid in the improvement of the validity of our measures and thus how we might evaluate different understandings of the questions and answers in a measure. To do this I turn to some of the work of Hans-Georg Gadamer. But before my discussion of Gadamer's contribution I will situate some of my concerns regarding quality of life research within well-known debates in the philosophy of science and social science. To this end I consider Bas van Frassen's work on the pragmatics of explanation with a specific interest in his work on questions. Here I will defend van Fraassen against criticisms by Wesley Salmon & Phillip Kitcher in order to maintain and broaden one of my arguments in this chapter, namely that we should not attempt to limit in advance the questions we take to be legitimate.

Chapter 4

An Argument in Favor of a Dialogic Framework

Introduction

In the last three chapters I looked at how different understandings of questions and answers can affect the validity of quality of life measures. One of my questions has been how should we best understand the outcomes data from standardized quality of life measures? This question about outcomes has in turn led me to question our understanding of quality of life itself. Does the inability to communicate orally preclude good quality of life?; Do all unexpected interpretations of questions result in inaccurate answers? At the end of Chapter 3 I suggested that our quality of life measures are best understood in terms of a dialogic framework.

My suggestion that we understand quality of life measures as part of a dialogic framework is the result of my conclusion that the structure of these measures constitutes a circle: to understand the questions in a measure we must understand the measure's construct, to understand the construct we must complete the questions in the measure. My point here is that quality of life researchers do not know in advance of asking their questions how best to understand them. But even if we do not know in advance how best to understand the questions and answers in a measure we are not without any orientation for if we situate our measures within a dialogic framework we can use information from other studies to begin to understand them.

Nonetheless, if we do not know in advance how to understand the questions and answers in a measure how can we evaluate different understandings of these questions and answers and remain critical of possible misunderstandings—both misunderstandings on behalf of respondents and on behalf of researchers? In this chapter I will discuss some of the opportunities that a dialogic approach offers for critical evaluation, but first I turn to the philosophical literature regarding what is involved in asking and understanding questions. I look at this literature so that we might better grasp some of the limitations all questions encounter, namely we cannot determine in advance the questions that will prove fruitful to an inquiry as opposed to those that will not; quality of life researchers are not alone in their inability to know in advance how to best understand their questions.

I begin my discussion of this literature with a defense of Bas van Fraassen's theory of pragmatic explanation. I begin here not because I am necessarily interested in explanation, but because van Fraassen thinks that explanations are *answers* to why-questions and he thus develops a theory of questions, which is meant to help us recognize good explanations. In examining van Fraassen's work I am not as interested in why-questions per se as I am in the lessons we can draw from them about questions and answers generally.[190] My defense of van Fraassen will focus on his position that we cannot put limits on a question's relevance relation—we cannot limit in advance the kinds of questions we might legitimately ask—and on Wesley Salmon & Philip Kitcher's criticism of this position. Salmon & Kitcher are worried that without a priori limits on questions we are in danger of adulterating our scientific knowledge with bad questions and thus misleading answers. I will argue, however, that in limiting the questions we ask we unduly limit what

---

[190] Van Fraassen says that 'why-questions introduce genuinely new elements into the theory of questions' and also that some of these elements are peculiar to why-questions. I will suggest that although some these elements may have been new to the theory of questions they are not necessarily peculiar to why-questions. For van Fraassen on the unique character of why-questions see van Fraassen, B. (1980) *The Scientific Image*, Clarendon Press: Oxford, pp. 141-6 & 151-3.

we can discover and, moreover, that such limits are unnecessary. To make this latter argument I turn to the work of Hans-Georg Gadamer.

At first van Fraassen and Gadamer may seem like an odd choice of interlocutors to help me make my point about the dialogic nature of quality of life, coming as they do from two different philosophical traditions. But what they share is an almost unique interest in questions; for with only a few exceptions the art of questioning has generally been ignored in philosophy in favor of what comes after: its answer. Moreover, although their points of departure are somewhat different both van Fraassen and Gadamer are interested in the contextual factors that allow certain questions to arise. Furthermore, both are at pains to illuminate these contextual factors and this process of illumination takes up much of their theoretical work in this area.

To be sure there are differences between these philosophers. For instance, in Gadamer's work on questions the process of illuminating contextual factors or as he calls it giving "them full play"[191] is primarily progressive—to evaluate the contextual factors which contribute towards our understanding—but for van Fraassen the process of illumination is mainly conservative. He is interested in contextual factors only because they determine the particular meaning of a question or as he puts it 'which question a particular interrogative expresses'.[192] Despite this difference, however, both agree that the context of a question is central to the meaning of the question and that we cannot put a priori restraints on that context in order to limit what questions we might legitimately ask.

---

[191] Gadamer, H. (2003) (2nd ed) *Truth and Method*, trans Joel Weinsheimer & Donald G. Marshall, Continuum Press: NY, p. 299.

[192] Van Fraassen, B. (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 141.

Nevertheless, when designing their measures quality of life researchers do try to limit the questions we ask respondents. For quality of life researchers aim to create questions which are open—they allow for genuinely new information—but foreclose possibilities not relevant to the measure's purpose—they discount information that comes from misunderstanding or other forms of bias. As we will see Salmon & Kitcher have a similar aim. On the one hand, they want to keep scientific questions sufficiently open so as to allow for new or unpredictable information. On the other hand, they want to foreclose the possibility of questions and answers antithetical to science. Their solution is to demand a priori limitations on van Fraassen's relevance relation, or as I will suggest, a priori limitations on the purposes to which different questions can legitimately aim. I will argue, however, that this ambition is not only unnecessary to maintain the integrity of our questions, but it is also damaging to their integrity. Thus I suggest we have reason to rethink not only our conceptualization of quality of life research, but also what it is we do when we ask questions.

I

Bas van Fraassen on Understanding Questions

A. Introduction

In chapter 5 of *The Scientific Image* van Fraassen explores what he calls 'The Pragmatics of Explanation'. In what follows I want to look at the three contextual factors which he identifies as seminal to our understanding of why-questions: the topic, the contrast class and the relevance relation.[193] In this section I will suggest a particular reading of the relationship among these contextual factors, viz. one which emphasizes the dependency of the topic on the contrast class and relevance relation for its meaning or, put differently, our understanding of it. I refer to this as a *diachronic reading* of van Fraassen's contextual factors as opposed to a synchronic reading which would take our understanding of the topic to be independent of the contrast class and relevance relation.[194] The significance of this reading is to emphasize just how dependent a question's meaning is on the context in which it arises, as well as to make clearer a limitation in van Fraassen's ability to evaluate questions.

For van Fraassen, if we want to understand the relationship between A and B that makes B an explanation of A, we need to take seriously the subordinating conjunction: A *because* B.[195] For some time most philosophers have recognized that the explanatory 'because' points to some nodes in the causal net rather than others and, moreover, those to which it directs our attention vary with the context of inquiry.[196] Nonetheless, many attempts to clarify the explanatory relation between A and B have tried to develop a priori, context *independent* criteria that connect B to A. But for van Fraassen taking the subordinating conjunction seriously means recognizing that 'because B' cannot be determined independently of the context 'Why A?' As a result he insists that we think of

---

[193] Ibid, pp. 141-3.
[194] I would like the thank Josh Rush for helping me to name this distinction.
[195] Wright, L. work in progress
[196] Van Fraassen, B (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 115.

explanations as answers to why-questions.[197] This means that the explanatory force of 'because B' is relative to its status as an answer to the question 'Why A?'

After my discussion in the previous chapter it should not be surprising that evaluating 'because B' as an answer to 'Why A?' requires that we understand just what question 'Why A?' expresses. In fact it is the difficulty in determining just what question 'Why A?' expresses that leads van Fraassen to turn to the erotetic literature in search of a theory of questions. As Larry Wright notes, perhaps the most significant result of this endeavor is van Fraassen's utilization of the role contextual factors—the topic; the contrast class; the relevance relation—play in both the asking and the understanding of questions.[198] In van Fraassen's analysis these factors have three main roles to play in the determination of why-questions and I will discuss each in turn.

B. Contextual Factors: The Topic

The first point to notice is that when we ask a why-question we presuppose that the topic of the question is true.[199] The topic of a question is a proposition which states that for which an answer is sought. For example, the question, 'Why do cochlear implants improve quality of life?' implies that cochlear implants do indeed improve quality of life. If they do not, then the appropriate response is to reject the question rather than answer it.[200] For van Fraassen a question's topic is true when the accepted background theory and factual information against which the question is asked implies that it is true.[201] Thus depending

---

[197] Ibid, p. 134.
[198] Wright, L. work in progress
[199] Van Fraassen, B (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 143.
[200] Ibid, p. 127.
[201] Ibid p. 145.

on the accepted background theory and factual information the same topic may, in one instance, be true and in another be false.

But despite, or perhaps because of the contextual sensitivity of van Fraassen's view coming to an agreement on the truth or falsity of a question's topic is not always straightforward—even when individuals share a common theoretical orientation and the same factual information. For example, say that a psychologist rejects the question 'Why do cochlear implants improve quality of life?' in the context of a guidelines development group designed to determine the conditions under which cochlear implants should be recommended. Cochlear implants, she argues, do not improve quality of life because they do not make it easier to learn; they do not make it easier to learn because the standardized test scores of those receiving cochlear implants are about the same as those not receiving them. Nonetheless, imagine that this question was first posed because the speaker—another psychologist from the same group—implicitly assumed that an improvement in learning should be measured in terms of self-reports and according to *those* results recipients indicated an easier time learning. The issue in these two interpretations is what counts as or measures an *improvement* in quality of life qua learning: test scores or self-reports?

In cases like the one above where the truth of the topic is at issue how do we decide which interpretation is the appropriate one? Van Fraassen is reticent about this particular matter and he chooses examples in which the truth or falsity of the topic is relatively uncontested. For instance, one of his examples is the question, 'Why did the conductor become warped during the short circuit?' He says that this question arises only if the conductor did indeed become warped during the short circuit; if it did not, then we would say that the speaker was under a false impression.[202] Van Fraassen seems to assume in this example and others like it, that the appropriate understanding of the question's topic is not an issue; that what counts as 'warping' and 'short circuit' is already sufficiently settled—at least for those asking and answering the question. This would seem to imply a conversational situation in which discussants already acknowledge the conditions that make the topic true. This observation leads Salmon & Kitcher to remark that van Fraassen assumes that interlocutors operate within "...a common context with a common body of background knowledge..."[203]

But it is sometimes the case that although interlocutors operate within a common context and a common body of background knowledge they still may disagree about the appropriate understanding of a question's topic. When this happens disputes can occur not only over whether the topic is or is not true, but also over different understandings that make the topic true. In cases of dispute it is unclear on van Fraassen's account what decides the matter. We will have reason to come back to this difficulty later, for now I turn to van Fraassen's second contextual presupposition: the contrast class.

C. Contextual Factors: The Contrast Class

For van Fraassen a contrast class is a set of contrasting propositions which includes the topic of the question where only the topic is true.[204] For example, we might ask, 'Why did the conductor become warped *as opposed to melting* during the short circuit?' The contrast class would include: the conductor became warped during the short circuit; the

---

[202] Van Fraassen, B (1980) *The Scientific Image*, Clarendon Press: Oxford, pp. 126-7

[203] Salmon, W. & Kitcher P. (1987) 'Van Fraassen on Explanation', in *The Journal of Philosophy*, 87: 318.

[204] Van Fraassen, B (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 145.

conductor melted during the short circuit. According to van Fraassen all why-questions require an implicit contrast and thus the underlying structure of why-questions is: Why (is it the case that) *P in contrast to* (other members of) X?[205] Van Fraassen refers to the contrast class as the central presupposition of a why-question.[206] Moreover, he says that the background theory and factual information of the context must imply the central presupposition if why-questions are to properly arise.[207] Put differently, for van Fraassen a why-question *arises* only if the context implies that the topic is true *and* the contrast(s) is false.

But as Charles Cross has argued in his 'Explanation and the Theory of Questions' not all properly arising why-questions presuppose that their topic is the only true member of the contrast class. In fact some why-questions both arise and presuppose that their contrast class contains only true propositions.[208] Cross provides the following example, 'I know why chemicals A, B, and C dissolve the cell walls of amoebae, but why does *chemical D* do that?' [209] Presumably the members of the contrast class for this question are: A, B and C dissolve the cell walls of amoebae; D dissolves the cell walls of amoebae. But unlike the example of the conductor none of the members of this contrast class are false. In light of Cross' argument, I will assume that a why-question properly arises just as long as its topic is true regardless of the truth-values of the other members of the contrast class.

When we ask questions we generally do so because there is a particular gap in our understanding. Following Larry Wright's work on van Fraassen we can say that it is the

---

[205] Ibid, p. 127.
[206] Ibid, p. 145.
[207] Ibid.
[208] See Cross, C.B (1991) 'Explanation and the Theory of Questions', *Erkenntnis*, 34:253.
[209] Ibid.

job of the contrast class to locate wherein that gap lies. It does this by directing our attention to the relevant class of items from which the topic of the question departs.[210] Notice that this way of understanding the relationship between the topic of the question and other members of the contrast class avoids over-generalizations regarding the truth values of members of the contrast class with respect to the topic and instead focuses our attention on the particular way the topic *differs* from the other members of the contrast class; here the notion of the contrast class is primarily comparative. Notice also that this understanding of a contrast class is not limited to why-questions, but applies to all questions: we begin to locate just where the gap in our understanding lies when we understand with what the topic is contrasted.[211] Put a bit differently we can say that the contrast class helps to locate the gap in our understanding by narrowing down what Wright calls the 'infinitude of relevance'.[212] This is the infinitude of potentially relevant answers one could give to any particular question depending on how we vary the contrast class.

To see how a contrast class limits the infinitude of relevance consider again the example from earlier, 'Why do cochlear implants improve quality of life?' Relevant answers to this question include, but are not limited to: because they amplify the nerve cells; because they allow deaf people to hear; because they make learning easier; because

---

[210] Wright, L. work in progress.

[211] Take an example similar to one given in Chapter 3: 'How is your general health?' Recall that in this question from the previous chapter respondents replied with a myriad of answers because they weren't sure what contrast class was implied. Indeed as Cross has also argued how-questions have a similar structure to van Fraassen's why-questions in part because they require an implied contrast. See Cross, C. B. (1991) 'Explanation and the Theory of Questions', *Erkenntnis*, 34: 253. Other examples of questions include: Who called *now as opposed earlier*? Do you want breakfast *as opposed to lunch*? Can you *as opposed to another group member* carry the groceries? Did you go to the museum *as opposed to the gallery*? What is justice *as opposed to injustice*? Moreover, for an account of the importance of contrasts to the meaning of assertions see Austin, J. L. (1980) (2nd ed.) *How to do Things with Words*, eds. J. O. Urmson & M. Sbisa, Oxford: Oxford University Press.

[212] Ibid.

now they can listen to music; because schools do not have the money to invest in special education programs; because recipients say they do; because hearing aids do not work as well; because people are less depressed; because hearing is important for language development...

Although this list could go on, if the context in which the question arises is rich enough, then it will limit these answers to those that address the specific gap in our understanding which the contrast class locates. For instance, to the question, 'Why do cochlear implants improve quality of life?' there are at least three different contrasts corresponding to three different places in the topic where a contrast could be implied[213]: 'Why do cochlear implants *as opposed to non-surgical interventions which aim to aid language development* improve quality of life?'; 'Why do cochlear implants improve *as opposed to decrease or maintain levels of* quality of life?'; 'Why do cochlear implants improve quality of life *as opposed to physical functioning*?'

Although each of these interrogatives represents a different question they each have the same topic: cochlear implants improve quality of life. Nonetheless the particular answers we might make in reply to these questions vary with the contrast. For example, to the question, 'Why do cochlear implants improve quality of life *as opposed to physical functioning*?' we might reply, 'because now those with them can hear sound' or 'because they are less depressed'. But we would not reply, 'because they amplify the nerve cell'. This response does not answer the question 'Why do cochlear implants improve *quality of*

---

[213] There are, however, other contrasts that we could make by varying the contrast class itself and not just where we place it in the topic. For instance cochlear implants *as opposed to other **surgical** interventions which aim to aid language development*, quality of life *as opposed to **quantity** of life*...etc.

life?' Rather it is a potential answer for the alternative question, 'Why do *cochlear implants* improve quality of life?'

Notice in support of my diachronic interpretation of van Fraassen's contextual factors that it is not only the answers that we give to the above questions that vary with the contrast class, but also that our understanding of the topic as a true assertion alters. For instance, when we ask the above question from the perspective of quality of life *as opposed to physical functioning* our answers reflect an understanding of the topic that emphasizes the qualitative aspects of our life at the expense of physical functioning—some non-physical factor makes cochlear implants an improvement on quality of life. But when we look at the second question, which contrasts cochlear implants with, say non-surgical interventions, than our understanding of the topic emphasizes the fact that cochlear implants are surgical interventions—something about the fact that the intervention was invasive means it improves quality of life.

D. Contextual Factors: The Relevance Relation

The contrast class helps us to limit the infinitude of relevant answers we might give in reply by clarifying just where the gap in our understanding lies, both in terms of directing our attention to a particular part of the question's topic and, on my diachronic reading of van Fraassen, in terms of providing us with a contrast against which to begin to understand what the topic means. Nonetheless, notice that although a contrast class gives us a class of items against which to compare the topic of the question, it does not tell us *how* to make that comparison; it doesn't tell us what counts as a significant difference between the topic and its contrast. For instance, take one of the examples from above, 'Why do cochlear implants *as opposed to non-surgical interventions which aim to aid language development* improve quality of life?' In this question we are asked to provide an answer that names an important difference between cochlear implants and non-surgical interventions that aim to aid in language development. We know that the relevant answer has to be one way in which implants positively differ from non-surgical interventions, but what counts as an important and positive difference?

The difficulty here is that there is an infinity of differences between cochlear implants and non-surgical interventions and each difference is potentially important. Some of these differences include: cochlear implants can allow for speech recognition if implanted early in a deaf person's life—non-surgical interventions very rarely allow for speech recognition no matter when they are implemented; the implant procedure is physically invasive; there is a chance of infection with surgeries; implants are cheaper than creating more effective special education programs; implantation requires a stay at the hospital; implants stimulates the auditory nerve'; implants are more effective than hearing aids; implantation is a faster procedure than learning sign language; implants work via a prosthetic replacement; the list could go on.

The main point here is that with a little imagination any one of these differences, and any number of others, could answer the question, 'Why do cochlear implants *as opposed to non-surgical interventions which aim to aid language development* improve quality of life?' For example, 'because cochlear implants can allow for speech recognition' may answer the question in a context where the ability to understand speech is synonymous with quality of life. And 'because it stimulates the auditory nerve' makes perfect sense in a situation where the relevant concern is the mechanism by which cochlear implants work.

Furthermore, 'because there's a chance of infection' may conceivably answer the question in a context where perceived risk only makes an intervention more valuable.

For van Fraassen none of these answers contradict one another because they are all answers to different questions[214]—questions which differ in terms of what sort of thing is being requested as an answer.[215] Moreover, notice that once again in favor of my diachronic reading of van Fraassen each of these answers further differentiates our understanding of the question's topic—even though the questions have the same contrast class. For instance, the reply 'because cochlear implants can allow for speech recognition' construes cochlear implants as a tool for oral communication; 'because it amplifies the nerve cell of the inner ear' emphasizes the biomechanical functioning of implants; and finally 'because there's a chance of infection' makes implants a risky operation. The different understandings of the topic that *these* replies provide are due to various interests and concerns that *in light* of the contrast class further limit the relevant answers we might give. Van Fraassen calls this third contextual presupposition the relevance relation: the respect with which an answer is given.[216]

For van Fraassen a reply to a question can only be considered an answer properly speaking insofar as it bears the right relevance relation to the topic and the contrast class.[217] Since we are dealing with why-questions the relevance relations in which van Fraassen is interested are relations of explanatory relevance, namely relations that request a reason or

---

[214] Van Fraassen, B (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 117 & 125-6.
[215] Ibid, p. 144.
[216] Ibid, p. 142.
[217] Ibid, p. 143.

cause as an answer.[218] But requests for reasons or causes are not the only kinds of interesting requests we can make when asking questions. Indeed as Charles Cross argues they are not even the only kind of explanatory requests we can make.[219] As I hope to make clearer in the next section I take all questions to require some kind of relevance relation whether they are explanatory requests or not. Thus I do not think there is anything unique about why-questions in their need for a relevance relation.

In light of the discussion so far the question, 'Why do cochlear implants *as opposed to non-surgical interventions which aim to aid language development* improve quality of life?' can be further specified as 'Why do cochlear implants *as opposed to non-surgical interventions which aim to aid language development* improve quality of life *with respect to oral communication skills*?; or *with respect to the factors that make cochlear implants work*?; or *with respect to the risk involved in different interventions*? Moreover, the addition of a relevance relation helps us to further understand the topic of a question by telling us what might count as an important difference when the topic is compared to its contrast.

E. A Neglected Contextual Factor: The Relevance of the Relevance Relation

Before moving on to the next section I want to look at some of the consequences that occur when different relevance relations affect our understanding of the topic. To begin, recall my earlier discussion of the topic. At the time I gave an example that

---

[218] Ibid, p. 142.

[219] Cross, C. B. (1991) 'Explanation and the Theory of Questions', in *Erkenntnis*, 34: 257. Cross suggests that how-questions can be explanatory and that when they are they request an answer that specifies *a way* (as opposed to a reason or cause) for the topic to be the case.

illuminated two different ways we might understand the topic 'cochlear implants improve quality of life' within the same context of inquiry. After my discussion of relevance relations we are in a position to see *how* these different understandings create a legitimate dispute over whether or not the question arises.

Return once again to the interrogative: 'Why do cochlear implants improve quality of life?' Suppose that our interlocutors from earlier understand the question as implying the same contrast namely, 'Why do cochlear implants improve *as opposed to decrease or maintain levels of* quality of life? In my earlier example, one interlocutor rejects this question because as she understands the topic there are no significant differences between individuals with or without cochlear implants. Her understanding of the topic is *with respect* to standardized test scores and it is because there is no difference in scores between the two cohorts that she rejects the question; the topic is false. Recall, however, that the second interlocutor's interests were *with respect* to the data from self-reports. With respect to this concern there is a difference between quality of life before and after the intervention. Thus in this latter case, but not in the former, the topic is accepted as true and the question arises.

Van Fraassen's sensitivity to contextual factors in determining just *what* a question is requesting for explanation is a significant improvement on much of the earlier explanation literature, which tended to overlook how contextual subtleties can sometimes change the meaning of why-questions and thus increase the variety of acceptable answers. But because a relevance relation affects how we understand the topic of the question it is possible to question the 'relevance' of the relevance relation. In other words, as in the example above regarding what counts as an 'improvement', sometimes we question or challenge whether the respect for which a question is asked *correctly* articulates the topic of the question.

Van Fraassen's interest in contextual presuppositions such as the contrast class and the relevance relation is not to debate which combination best articulates the understanding of a topic, but rather to differentiate the question a particular interrogative expresses from all the other questions it might express. Nonetheless he does give us criteria with which to evaluate whether a question properly arises in a particular context—whether we ought to accept or reject a question. In fact I have already discussed these criteria. Recall from my earlier discussion that rejected questions have false topics; questions which arise have true topics.[220]

Within a highly shared context, by which I mean a context in which the interlocutors generally understand key concepts and mechanisms in the same way, background information can work to illustrate misplaced questions in the form of false topics. But these mistakes or misunderstandings must always be somewhat superficial if a shared background is enough to reveal them. Consider again the example of the warped conductor: 'Why did the conductor become warped during the short circuit?' Van Fraassen writes that this question only arises if the conductor did actually become warped otherwise the interlocutor is under 'a false impression'. In the case of a highly shared context what counts as 'warped' or 'conductor' is not at issue thus the interlocutors agree at least in principle about what constitutes a warped conductor. If an interlocutor thinks that the

---

[220] It is important to note that van Fraassen also allows us to reject questions with true topics if the background theory and factual information of the context imply that all answers to the question is false, but my focus in this chapter will not be on these questions. See Van Fraassen, B (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 145-6.

conductor is warped, then she might ask the question why it is so. But if she is wrong—say she is looking at the conductor upside-down—then reference to the shared context should correct her mistake, for instance, we might flip the conductor around.

But deeper disagreements occur in the absence of such a *highly* shared context. By this I do not mean the *absence* of a shared context, but rather the absence of sufficiently shared *understanding* of the shared context. This is the problem in the cochlear implant example where the tension is over what counts as an 'improvement'. In this example the disagreement is over the appropriate employment of a particular relevance relation.

Note in this example reference to a topic's truth-value as a criterion for accepting or rejecting a question will only deepen the disagreement and not, as in the case of the conductor, illuminate error since here the interlocutors approach the question with different interests, interests which in turn suggest different understandings of the presiding context. But if van Fraassen cannot rely on the truth or falsity of the topic to decide whether a question properly arises, neither can he merely fall back on his understanding of context-sensitivity and simply notice how all the relevance relations are legitimate given their respective interests and background theory. The tension in this case is that the interlocutors take themselves to *share* a context of inquiry—they are questioning the *same* thing—but they disagree over its appropriate understanding.

For van Fraassen when we are confronted with a why-question it arises when its topic is true and otherwise we reject the question. But the dichotomy of merely accepting or rejecting a question based on its truth-value conceals a further contextual factor: the relevance of the relevance relation. Although van Fraassen's recognition of the importance of contextual factors when asking questions is important his process for determining when we accept or reject a question is insufficient because he does not allow for disagreement over the understanding of a topic among individuals who share a context. As I will discuss in the next section, this oversight deprives van Fraassen of the appropriate counterbalance to his insistence that relevance relations remain unlimited.

II

Wesley Salmon & Phillip Kitcher: An Objection to van Fraassen

A question's relevance relation—the respect with which we ask a question—affects how we understand a question's topic. That a question's relevance relation plays such an important role in making clear just what question an interrogative expresses should not surprise us. The relevance relation, understood as the interests or motivations that prompt a question, is very similar to what I referred to as the purpose of a question in the previous chapter. There we saw how essential the purpose of both assertions and questions was to understanding what they meant.

One of the issues in the last chapter concerned how we ought to understand of the questions in the measures. Outcomes researchers, concerned about misunderstanding and other forms of respondent bias, attempt to design measures in such a way that the interrogative answered by the respondent corresponds in some way to the question that the researchers had in mind. Inevitably, however, respondents sometimes understand questions differently. I argued in Chapter 3 that these differences are, in the first instance, often the result of a different understanding of a question's purpose and secondly that it is not necessarily a misunderstanding since it is not clear in advance what the question's purpose ought to be.

I now want to rephrase and strengthen the second part of this argument in the following way: we cannot determine *in advance* the sort of relevance relation or purpose that will make a question fruitful and penetrating as opposed to those that will result in misunderstanding or futility. This claim applies equally to scientific questions and to questions about quality of life.

Wesley Salmon & Phillip Kitcher deny this claim in 'Van Fraassen on Explanation'. They argue that if we do not put a priori limits on which relevance relations can count as legitimate, then 'almost anything can explain almost anything'.[221] If almost anything can explain almost anything, then they conclude that we will not have any criteria for which to distinguish scientific questions and explanations from nonscientific ones. Put slightly differently, we will have an "anything goes" theory of explanation.[222] If Salmon & Kitcher are right and if we extend van Fraassen's account further afield, then we also will not be able to distinguish good from bad questions and answers in other disciplines like quality of life.

In the rest of this chapter I shall argue that Salmon & Kitcher are mistaken. We do not need a set of a priori relevance relations in order to distinguish between fruitful unfruitful questions. Nonetheless they are right to think that on van Fraassen's account 'almost anything can explain almost anything'. This consequence however, is the virtue, not the downfall, of van Fraassen's pragmatic theory of explanation.

---

[221] Salmon, W & Kitcher, P. (1987) 'Van Fraassen on Explanation', in *The Journal of Philosophy*, 87: 322.
[222] Ibid, p. 329.

In a variety of places in his chapter 'The Pragmatics of Explanation' van Fraassen more or less argues that we cannot limit the scope of appropriate relevance relations. He does so by first recognizing, as I demonstrated in the previous section, how an almost infinite variety of responses can properly answer a single interrogative. In the car crash example he takes from Norwood Russell Hanson there are as many answers to the question, 'Why did he die?' as there are interests and orientations to the question.[223]

Thus depending on the circumstances we could say he died from a brain hemorrhage or from the overgrown shrub in the road or from bad brakes. But if we say that he died from the hemorrhage it is not that we think the overgrown shrub was actually quite trim or that the brakes were in good order, rather we accept that the tree might have been overgrown and the brakes bad, but nonetheless keep these factors fixed—ceteris paribus—and only allow physical factors to compete as the answer.

Van Fraassen writes that there is nothing objectively right or wrong about keeping one sort of thing fixed as opposed to another.[224] That is, there is nothing in principle right or wrong about asking a question with respect to one sort of interest or relevance relation as opposed to another. We cannot reject a question merely on the basis of the interests or motivations it seeks to satisfy because there is nothing in our picture of the natural world that limits the sort of questions that can count as valid. Gaps in our understanding occur in all sorts of places depending on our background, our training and the current context.

As we have already seen it is true that on van Fraassen's account just about any reply can count as an answer to just about any question given the right context. In the

---

[223] Van Fraassen, B. (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 125.
[224] Ibid, p. 116.

example I used in the last section, 'Why do cochlear implants *as opposed to non-surgical interventions which aim to aid language development* improve quality of life?' some of the seemingly irrelevant differences such as, 'it requires a stay at the hospital' may be relevant if in asking this question we are interested in cochlear implants with respect to quality of care and we have reason to think that a short stay in the hospital and the required follow-up visits provides us with that. If we consider the variety of contrasting cases and interests with which we can ask the question, 'Why do cochlear implants improve quality of life?' we begin to see how almost anything could explain it.

Moreover, just as almost any reply can be the right answer to a particular question, as we saw in the last chapter any declarative sentence can be the answer to many different questions. Recall the assertion, 'the cat is on the mat'. This could be the answer to any of the following questions: Where is the cat? What's the password? Why didn't you hoover the carpet? Can we make a fire? Where's the bulldozer? Who took my yoga mat? What are you laughing about? And so on. Here we see how something can explain almost anything.

My point here is that it is part of our explanatory practice that almost anything can explain almost anything. In other words it is a valuable aspect of van Fraassen's theory of explanation that it allows for this flexibility. Van Fraassen is right that gaps in our understanding can occur in any number of unlikely places and the variety of interests with respect to which a question arises can mean that seemingly bizarre questions are sometimes met with understanding and acceptance.

The problem with van Fraassen's account is not that an almost infinite variety of questions and answers *can* be legitimate given enough imagination or the right context. The problem is how do we know when a question *is* legitimate given a *particular* context? How do we know when we have asked a *good* question—a question with a relevance relation, which is itself relevant to the context at hand? How do we know if our question is valid? This issue is over whether van Fraassen gives us an 'anything goes' theory of explanation.

Salmon & Kitcher argue that van Fraassen does give us an 'anything goes' theory of explanation. To illustrate this they provide the following example: 'Why John F. Kennedy died on Nov. 22, 1963 as opposed to other dates, with respect to astral influence?' They suggest that astral influence is not an appropriate relevance relation and in the context of twentieth-century science we ought to reject the question.[225] If van Fraassen's theory does not provide the tools to reject this question, then they argue that the answer will serve to adulterate the integrity of scientific knowledge.

Although there is general agreement in the literature that the JFK question ought to be rejected there is controversy over whether it would properly arise on van Fraassen's account. Salmon & Kitcher argue that it would arise since, on their reading, it fulfills van Fraassen's criteria for when we ought to accept questions as properly arising. Firstly, it is true with respect to scientific theory and the factual information of the case that JFK did indeed die on Nov. 22, 1963 and secondly, they argue that there is nothing inconsistent with an answer to this question—a certain configuration of the planets and stars—and current scientific knowledge. In fact the configuration of the planets and stars would be an application of scientific knowledge.[226]

---

[225] Salmon, W & Kitcher, P. (1987) "Van Fraassen on Explanation" in *The Journal of Philosophy*, 87: 322.

[226] Ibid, p. 324.

On the other hand, Elisabeth Lloyd & Carl Anderson along with Allen Richardson argue that the question would not arise. They rely on an often quoted passage in 'The Pragmatics of Explanation' where van Fraassen appears to limit the possible relevance relations with which one can request explanations to just those that are scientifically relevant.[227] They argue that our scientific background knowledge serves to limit what counts as scientifically relevant and thus on van Fraassen's account we can, 'reject the JFK question on the grounds that astrology is not an accepted scientific theory...'[228]

We might reinterpret Lloyd & Anderson's point in light of my diachronic reading of how we understand a question's topic. This reading would suggest that the JFK question should be rejected because its topic, when understood in light of its relevance relation, *is* false with respect to current scientific theory. The reasoning would go something like this: it is not true given our background information and current scientific theory that JFK died on Nov. 22, 1963 due to the alignment of the planets and stars on this day. But even if we accept this interpretation of the topic as the correct one are Lloyd & Anderson right to reject this question tout court? In other words, is it the case that there is not *any* current scientific context in which the date of JFK's death is due to the alignment of the planets and stars?

In suggesting that we reject the JFK question both Salmon & Kitcher and Lloyd & Anderson tacitly appeal to a scientific context in which we are interested in the causal features of JFK's assassination. But this is not the only scientific context in which the question could arise. For example, perhaps one of the assassins was a devout follower of

---

[227] Van Fraassen, B. (1980) *The Scientific Image*, Clarendon Press: Oxford, p. 126.

[228] Lloyd, E. A. & Anderson, C. G. (1993) 'Empiricism, Objectivity, and Explanation', in *Midwest Studies in Philosophy*, XVIII: 124.

astrology and the advantageous planetary alignment on Nov 22, 1963 caused him to pick this particular day for the assassination. We can even imagine the coroner incredulously asking Salmon & Kitcher's question to an investigator.

My point is that "astral influence" is not always an unscientific relevance relation. Moreover, the fact that "astral influence" is not *always* unscientific is similar to an example I discussed in the previous chapter. Recall the example from the NHP that asked 'I find it hard to reach for things'. In this example the respondent answered 'yes, because I am short'. The respondent interpreted the question *with respect* to her height rather than with respect to her health. But "height" like "astral influence" is not *always* a bad relevance relation. For instance, in certain circumstances we might think that not being able to reach things due to one's height is relevant to depression. Moreover, I suggested in Chapter 3 that "height", even in the context of the NHP, may help us to better understand this measure's construct by having us reflect on when height does become a health issue.

That these relevance relations are sometimes, but not always inappropriate even within their respective disciplines is an important point. First, it is important because it emphasizes van Fraassen's position that anything can explain anything given the right circumstances. Second, it is important because if we eliminate "astral influence" from a whole field of interest, then we also potentially limit our arsenal of scientific knowledge in our attempt to preserve it.

But if Lloyd & Anderson's suggestion fails, if we cannot simply reference the current context of science in order to limit relevance relations and thus reject certain questions out of hand, then on what grounds can we criticize questions? Notice that this problem is very similar to the problem in the previous chapter where I argued that quality of life constructs themselves cannot fruitfully limit in advance the purposes with which questions are posed in quality of life measures. The problem is that in both situations unexpected purposes or relevance relations can sometimes expand our understanding of scientific subjects or quality of life constructs in novel and fruitful ways.

My analysis of the JFK examples is a reminder that in science, like quality of life research, we cannot solve disputes regarding the validity of certain questions from the top down; we cannot take the context of science or the research construct in quality of life as our limiting factor because van Fraassen is essentially right—anything can explain anything and limiting what might count as explanatory limits what we might come to know.

In the next and final section I will argue that although Salmon & Kitcher are right to recognize the need for a critique of relevance relations, they are nonetheless wrong to suggest that such a critique must manifest itself as an a priori delimitation of relevance relations. I will put forward an alternative solution and in doing so I turn to Hans-Georg Gadamer.

III

Hans-Georg Gadamer: Textual Understanding and Critical Evaluation

Salmon & Kitcher argue in their article, 'Van Fraassen on Explanation' that if anything can explain anything, then it follows that we have an "anything goes" theory of explanation. Since an "anything goes" theory of explanation is unacceptable they deduce that the antecedent is false: anything cannot explain anything. Instead, they argue that we need to put limits on our requests for information—we have to limit what counts as a scientific relevance relation.

To make this point about scientific inquiry they provide the following analogy with literary interpretation:

> Just as pluralists about literary works will insist that there are many interpretations of Hamlet while denying that any reader's fancy counts as an interpretation, so too, relativists should concede that there are some relations that are not genuine relevance relations at any historical stage of any science.[229]

I agree with Salmon & Kitcher that "relativists"—by whom they mean van Fraassen—and literary pluralists should concede that not just any interest will provide us with a valid understanding of nature or a text. But I disagree that in order to be critical about our understanding of nature, we must concede in advance that some interests or relevance relations are inappropriate '...at any historical stage of any science'.

Recall from the last section my point that although "astral influence" is certainly a dubious relevance relation in some scientific contexts, it need not be problematic in *all* scientific contexts. In fact, as I noted at the time, if we eliminate "astral influence" as a legitimate interest in scientific inquiry, then we may also unduly limit scientific knowledge. Salmon & Kitcher's mistake is to think that the way to deal with the threat that certain questions pose is to limit in advance what counts as a scientific interest. This solution, however, goes too far and ignores van Fraassen's positive contribution, namely, that there is nothing in principle right or wrong about asking a question with respect to one sort of interest or relevance relation as opposed to another.

---

[229] Salmon, W & Kitcher, P. (1987) 'Van Fraassen on Explanation' in *The Journal of Philosophy*, 87: 326.

To avoid Salmon & Kitcher's mistake and preserve van Fraassen's insight I suggest a third way. This "third way" takes its point of departure from the very literary pluralists that Salmon & Kitcher draw on for support in their analogy. For literary pluralists need not determine in advance what interests count as 'literary' in order to maintain a critical stance and neither do scientists. To help make this argument I turn to part of Gadamer's claim in *Truth and Method.*

When we approach a text, or text-analogue we come to it with certain implicit and explicit assumptions regarding what it might say about the subject matter that it addresses. These assumptions in turn affect how we understand what a text says about its subject matter. The same holds for text-analogues. For instance, in Chapter 2 I discussed how researchers and deaf advocates might understand the same set of outcomes differently given their different orientations to deafness. If one assumes that deafness is a pathology derived from the inability to hear, then one understands the increase in quality of life that cochlear implant recipients receive in one way. If one takes deafness to be a biological difference that provides for a unique set of opportunities and challenges, then one understands the reported increase in quality of life a different way.

Likewise, Gadamer recognizes that our interests or assumptions about a subject matter guide our understanding of a text. Moreover, he thinks that these assumptions pose the danger of appropriating a text in terms of these interests and assumptions and thus being unable to distinguish between, 'our customary usage and that of the text'.[230] We might then wonder how it is possible to avoid misunderstanding a text from the very start as I suggest

---

[230] Gadamer, H. G. (2003) (2nd ed) *Truth and Method,* trans revised by Joel Weinsheimer & Donald G. Marshall, Continuum Press: New York, p. 268.

outcomes researchers or even deaf advocates are in danger of doing when they interpret outcomes data.[231] To solve this problem Gadamer introduces two criteria for understanding texts and text-analogues, which act as a counterbalance to these kinds of misunderstanding. First, we must assume that texts and their analogues are coherent and second, as we do in communicative speech[232], we must assume that what is presented to us for consideration is at least possibly true.[233]

Both of these criteria are meant to provide a standard that can separate valid from invalid interpretations. For instance, valid interpretations must find a way to unify the parts of a text into a coherent whole; conversely invalid interpretations cannot make the various parts of the text cohere. But notice that this criterion alone is not enough to guard against misunderstanding: it is possible to make a text coherent in light of certain assumptions and yet also potentially misunderstand it.[234] In fact this is the situation we find with respect to the outcomes data, for researchers or deaf advocates understand the data as forming a coherent whole in light of their assumptions about what makes for a good quality life and yet I argue that they still potentially misunderstand it. It is Gadamer's second criterion—that we assume a text or its analogue is possibly true—which helps to overcome these kinds of misunderstanding.

In assuming that what a text says is true we presume that we might learn from a text; that our understanding of its subject matter is potentially incomplete and might be

---

[231] Ibid.

[232] Herman, B (2006) "Ends Justifying Means" University of California, Riverside 16th Annual Conference *Normativity & Universality from a Kantian Perspective*, 24 February.

[233] By possibly 'true' Gadamer does not mean that we must assume that the events in the text *actually happened* nor must we agree with what the text expresses. Rather for Gadamer assuming that the contents of a text are possibly true involves recognizing the text as an experience from which we might learn something.

[234] See Warnke, G. (1987) *Gadamer Hermeneutics, Tradition and Reason*, Stanford: Stanford University Press, p. 84.

enhanced or altered by our interaction with the text. In assuming that a text is true we take any differences between the expectations that arise from our assumptions and what we find in a text as potential challenges to those assumptions; we neither take the anomalies as instances to be explained in terms of our assumptions nor as views to which we must acquiesce.[235] On the other hand, when a text meets our expectations, then Gadamer enjoins us to relate the text back to ourselves; to ask what significance this fact might have for us; what does it say about us and our assumptions that the text plays into our hands? The job of the interpreter then is not to appropriate a text as mine or yours, but rather to become clearer about the subject matter at hand and doing so is a matter of trying to see the significance of the text's truth claim in relation to our previous assumptions about that subject matter.[236]

It is through the process of applying our own assumptions to a text and considering what the text has to say about them that Gadamer, as I said in the introduction to this chapter, illuminates our contextual presuppositions about a subject matter or as he says, gives them 'full play' [237]. At the same time, however, he puts them 'at risk'[238] for although we may end up maintaining our presuppositions and disagreeing with a text about its subject matter we also might alter our presuppositions or sometimes replace them. Thus in trying to understand the meaning of a text we sometimes come to understand its subject

---

[235] Notice that what we take to be discrepancies between what we expect and what we find in the text is necessarily predicated on our initial assumptions, nonetheless in trying to give a valid interpretation we do not merely explain away these discrepancies in terms of our assumptions, but rather allow the discrepancies to act as interruptions which both highlight some of our assumptions and put them into question.

[236] See Warnke, G. (1987) *Gadamer Hermeneutics, Tradition and Reason*, Stanford: Stanford University Press, p. 104.

[237] Gadamer, H. (2003) (2nd ed) *Truth and Method*, trans Joel Weinsheimer & Donald G. Marshall, Continuum Press: New York, p. 299.

[238] Ibid, p. 388.

matter differently. Moreover, even if we do not change our presuppositions a coherent and truth-sensitive interpretation of a text can nonetheless serve to educate us about a subject matter not only by making our own presuppositions about it more transparent, but also by making them better—that is less reliant on unexamined and inappropriate interests. It is in this sense that Gadamer's interest in understanding is primarily progressive for on his account by interpreting a text we also evaluate our assumptions about a subject matter and come to understand *it* better.

In contrast to Salmon & Kitcher, Gadamer does not require literary pluralists to determine in advance which interests in or assumptions about a subject might give a valid as opposed to an invalid reading of a text. Valid interpretations of a text are coherent and truth-sensitive. Conversely, when our interests and assumptions do not allow us an interpretation that can make sense of a text or see its point then they may be invalid. In other words, if we are to understand a text, then we might need to shift our perspective on it and hence the framework from which we perceive it. Yet we do not do so prior to putting our assumptions in play and at risk and in this sense Gadamer mirrors van Fraassen's claim that anything can "explain" anything given the right context. To follow the connection between Gadamer and van Fraassen further I want to look at Gadamer's account of interpretations as answers to questions.[239]

For Gadamer we understand what a text says about its subject matter when we see the text as an answer to a question. This process of understanding begins, as we have seen, when we approach a text with certain implicit and explicit interests, concerns and assumptions regarding a particular subject. If we assume that a text has something to tell us

---

[239] Ibid. pp. 369-70.

about the subject, then we characterize any differences between the text and our expectations as potential challenges to our assumptions. But notice that these differences first present themselves to us as questions. For instance, 'Why do John McGahern's books lack a plot?'; 'Is Shakespeare a racist?'; 'Was Anne Boleyn really so ruthless?' The text properly challenges our assumptions when in answering these questions we start to reconsider what we mean by "plot" and what might count as being a "racist" or "ruthless". As this dialogue between the text and the interpreter continues we build up an understanding of a text in terms of answers to our questions and in doing so we enrich and clarify our understanding of the subject matter.[240]

Our particular experiences, history, culture, knowledge and training all shape our prevailing interests in a subject matter. These interests in turn affect the kinds of questions we find in a text and the answers we eventually give these questions ultimately affect not only how we understand the text, but also how we understand its subject matter. But because our experiences, history, culture, knowledge and training change over time and among individuals so will the questions we ask a text and thus the understanding we glean from it. For instance, the assumptions 21st century readers bring to Shakespearean plays are vastly different from those with which a 16th century audience would have approached it. Today we might ask if Shakespeare was a racist, but in the 16th century this was not a relevant category of interest.

The upshot of both the fact that experiences, history, culture, knowledge and training affect our individual interests, concerns and assumptions about a subject, and the

---

[240] This dialogue between a text and an interlocutor may also be a dialogue among different text*s* and an interlocutor. Some of the questions we raise may be due to other things we have read; some of the answers we give to these questions may be the result of other texts.

fact that we need an orientation with which to begin to understand a text is that we cannot determine in advance the sorts of interpretations that make for valid readings of a text. We cannot provide these criteria because we cannot anticipate all the different ways that a text might be fruitfully understood. If we did want to say in advance what sorts of interpretations were valid we would have to know in advance the limits of what the particular text could say, but this is a problem since what a text can say in part depends on the interests we bring to it and the questions we ask it.

The idea from the last section that anything can explain anything becomes for Gadamer the fact that any set of interests, concerns and assumptions can potentially illuminate any text. But as we have seen, for Gadamer, this fact does not translate into Salmon & Kitcher's worry in science that we then have an "anything goes" theory of explanation. For Gadamer just because any interest *can* illuminate any text does not mean that just any interest *does* illuminate a particular text: some interests will not be able to get various parts of the text to cohere and some interests will not be able to make the truth claim of the text significant. Indeed, as we have seen it is in trying to understand what the text means that we may recognize the inadequacy of some of our interests and assumptions and thus alter or replace them.

If we look at Salmon & Kitcher's JKF example in light of Gadamer's work on valid and invalid interpretations, then we see that they are partially right: under certain conditions we should reject this question. For example, if the text-analogue in their example is the causal features of JFK's assassination, then an individual is going to have a very difficult, if not impossible job showing the efficacy of astral causation. Yet, rather than eliminating the interest in advance, we would expect its irrelevance to come out in the wash, as it were.

We would expect our interlocutor to question the relevance of astral influence by actually pursuing the inquiry and in doing so perhaps learn more about western science.

With respect to quality of life measures our critical evaluation of different understandings of questions also depends on whether or not a particular understanding succeeds in making a text-analogue—a measure—coherent and whether or not it respects a measure as possibly true. In quality of life measurement problems understanding questions apply both to researchers and respondents. Although respondents can misunderstand questions and thus their understandings can be criticized—for instance, they can fail to understand how individual questions in the measure cohere with the whole; they can fail to see the significance of the text's truth claim in relation to their previous assumptions about the subject matter—this fact does not take away from the researchers' responsibility to try to understand how respondent answers render a measure coherent and possibly true.

We improve on the validity of our measures when we take respondents' answers seriously and often this means learning to see the subject matter—the measure's construct—differently. We also improve on the validity of our measures when we strive to construct them so that respondents can make sense of them and can treat them as possibly true and sometimes this means altering our measures in light of information from other studies. But doing these things does not mean that we merely acquiesce to respondents' understanding about quality of life. For Gadamer a dialogic framework allows us to engage with our interlocutors—different measures, studies, researchers—about the subject matter at hand and while this process requires us to give our assumptions 'full play' and put them 'at risk' it does not require us to agree with our interlocutors about the subject matter.

Questions like Salmon & Kitcher's JFK example are similar to questions we might ask today about the morality of slavery: at one time they were questions we took very seriously—Do the plants and stars cause events on earth? Are black people inferior?—although we no longer take them seriously. This is mainly because after extended debate we settled on answers to these questions and the answers so changed the way we understand science or, with respect to the slavery question, how we understand humanity, that the questions now seem unscientific or racist.[241] One difference between these kind of questions and questions about quality of life is that the latter questions are, for the most part, unsettled. Much less is built into the assumption that a quality of life measure is true—especially for respondents who have multiple answers with which to understand it—than is built into the assumption of the causal features of JFK's assassination. Thus there is greater leeway in the kinds of questions it makes sense to ask in regard to these measures.

The fact that the JFK question can now be used as an obviously unscientific question is educational because it shows how the answers we give to questions can transform our understanding of things. The fact that the configuration of the planets and stars do not cause events on earth helped us to distinguish science from mysticism, the fact that slavery is wrong helped us to distinguish humans from animals. But dead questions like those about astral influence and slavery should not confuse us into thinking that they were always uninteresting. A long history of debate surrounds both of these questions in which women and men vied to give ever more complete interpretations of the data and in

---

[241] See Taylor, C. (1989) *Sources of the Self the Making of the Modern Identity*, Cambridge: Cambridge University Press, pp. 64-5 for a short discussion on the historical similarities between the evolution of human equality and scientific progress.

doing so they transformed our scientific and social/political practices respectively.[242] In time we might expect the same transformation in certain areas of quality of life research and in fact we might take the debates around Female Genital Mutilation (FGM) as following in the footsteps of astrology and slavery questions.

The fact that we do not take the configuration of the planets and stars to impinge on events on earth contributes to the kinds of questions which properly arise: questions about the assassin's motivation can arise, but questions about the planets' causal effect on the date of the assassination do not. But the answers we give to the questions that do arise in turn affect how we understand the text or text-analogue in question and thus its subject matter. These changes to our understanding further affect the kinds of questions which it makes sense to ask. Moreover, sometimes it happens that previously 'dead' questions are reconceptualized and take on a new life. My point here is that although a question like Salmon & Kitcher's appears to be dead it was not always dead and it may live again. We cannot determine in advance the kinds of questions which will illuminate a text or a text-analogue, but we have the ability to be critical of those that do not.

Gadamer's criteria for valid and invalid understandings give us a solution to Salmon & Kitcher's worry that without objective limits on relevance relations or the purpose with which we ask questions we shall end up with bad questions and misleading answers. It also provides a solution for my own worry that if we do objectively limit relevance relations, then we seriously limit what we might come to know about our subject of inquiry. But Gadamer's solution comes at a price and that price is that we can only test our assumptions and try out a particular understanding of a subject matter through the process of questioning

---

[242] I am grateful to Georgia Warnke for her help with this argument.

a text or its analogue. And as Gadamer quite explicitly says, 'There is no such thing as a method of learning to ask questions, of learning to see what is questionable'.[243] We cannot know in advance what might be a fruitful question and we cannot teach someone to recognize what might be a valuable interest in a subject.

Scientific practice is in fact rife with the very kind of questioning that Gadamer promotes. We see the evidence for this not only in the historical debate over astral influence, but also in the broad literature in Science Studies that has accumulated since Thomas Kuhn's work on scientific revolutions.[244] Yet, in the area in which this thesis is concerned the price that Salmon & Kitcher exact in the limiting of questions is a steep one. In quality of life research researchers strive to limit in advance just what question an interrogative can express in order to create valid measures. However, when we attempt to put a priori limits on how the questions in quality of life measures should be understood we limit the answers we might give to the question, 'What is quality of life?' But the limits put on questions and indeed the assumptions about quality of life are often premature: when we examine medical histories, such as the interventions aimed to 'cure' deafness, as well as qualitative and other complimentary studies of quality of life we find that researchers may have an incomplete understanding of quality of life and that respondents and others can help us to understand it better.

If we want a valid and sufficiently broad understanding of quality of life, then we need to participate in something like a dialogue on the subject matter, i.e. a measure's

---

[243] Ibid, p. 365.

[244] See Kuhn, T. S. (1996) (3rd ed) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press. And for another example see Harding, S. (1998) *Is Science Multi-Cultural?*, Bloomington: Indiana University Press.

construct. In this dialogue respondents are given a voice through the different kinds of measure's and histories that target subjective experience. The researchers who analyze this data are like interpreters who question a text in order to understand it. Researchers question the outcomes from different studies with an eye to better understanding quality of life and constructing better, more accurate measures. This process does not mean that we need to turn quantitative research into a series of interviews; this is certainly not a call for the elimination of measurement. But it is a call to understand quality of life measurement somewhat differently; it is call to think of them not as determinate assessments of quality of life, but rather tools for enhancing communication about quality of life.

In the next and final chapter I return once again to the topic of construct validity. Having argued in the previous chapters that it is flawed, I now turn to examine what might encourage its use. Drawing on the work of Charles Taylor I offer two different pictures of the human subject: the weak and strong evaluator. I argue that construct validity presupposes that respondents are weak evaluators and I suggest that such a picture of persons might promote the use of construct validity. Nonetheless, I argue that this picture is an inadequate portrayal of persons. The notion of the strong evaluator provides a better conception of the human subject and allows me to introduce an ethical dimension to our questions about quality of life. If Gadamer supports the view that our investigations into quality of life are dialogic, then Taylor shows us that these investigations are *ethical* dialogues.

Chapter 5

The Ethics of Quality of Life and the Logic of Construct Validity

Introduction

In the previous two chapters I have argued that the continued improvement and indeed validity of quality of life measures rests on our ability to reconceptualize these measures as elements within a larger dialogic framework of inquiry. In other words, we ought to use quality of life measures as tools to enhance communication between patients, doctors, health service researchers and other health care professionals and care-givers about quality of life. Understood as tools to enhance communication about quality of life neither the questions nor the outcomes from these measures are taken at face value, but rather they are understood as a point of departure for further inquiry.

Part of the reason for the connection between a dialogic framework and the validity of our measures lies in the way it allows us to take respondents' answers seriously, by which I mean that we need neither acquiesce to their responses nor fall into an opportunistic interpretation of their answers. Taking respondent answers seriously is important to the inquiry into quality of life in part because the development of quality of life research is grounded to a significant degree in the promotion of self-determination and an antipaternalistic ethos. Valuing self-determination requires something more from our measures than the opportunism as sometimes embodied by standardized measures when we dismiss respondent answers as biased or the quick acquiescence of both standardized and individualized measures when we take a measure's outcomes to confirm our hypotheses.

But besides being grounded in the promotion of self-determination, quality of life research is also grounded in psychometric methodology where construct validity prominently figures. Construct validity, however, is *itself* is one of the obstacles we face when we attempt to take respondent's answers seriously for as I discussed in Chapter 3 it presupposes that we know in advance how best to understand the questions in our measures and as I discussed in Chapter 2 it is too quick to assume that we know the significance of respondent answers. For instance, assumptions built into our use of construct validity regarding how to understand the questions in a measure can unduly affect our understanding of the outcomes—we may incorrectly understand certain answers to be the result of social pressure or misunderstanding. On the other hand, even if respondents understand our questions and their answers as indeed we understand them construct validity can sometimes encourage us to overlook their significance.

The use to which construct validity is often put in quality of life research suggests the assumption that questions and answers are much less open to interpretation than I suggest. For it is in part the indeterminate nature of our questions and answers that makes construct validity ineffective, for instance, answers may be taken to confirm an assumption when in fact they might also be understood as undermining it; alternatively interpreted questions may yield answers that appear to bias our measure when in fact they might deepen our understanding of it. In this final chapter I draw on Andrew Edgar's work on quality of life measures to propose—and reject—a picture of the human subject that conceptualizes our responses to many questions as resistant to further interpretation and thus might motivate the current use of construct validity.

Much of Edgar's work relies on Charles Taylor's distinction between weak and strong evaluations to argue that quality of life researchers mistakenly assume that respondents are simple weighers of weak evaluations. As I will discuss in the next section weak and strong evaluations mark a difference in how desires or outcomes are appraised: when we weakly evaluate we judge something to be good simply on the basis that it is desired; when we strongly evaluate we judge something to be good on the basis that our desire itself is worthy.

For Taylor the assumption of the simple weigher marks a particular, and ultimately untenable, understanding of the human subject. I will suggest that it is this conception of the human subject which lends support to the use of construct validity in quality of life research and in light of the previous chapters I suggest that it is an untenable account of respondents. If quality of life research is to take respondent answers seriously, then it must relinquish the notion of the subject as a simple weigher of alternatives and embrace them as equal partners in ethical discourse.

In what follows I begin with Taylor's distinction between strong and weak evaluations in order to lay the groundwork for Edgar's claim that quality of life measures assume that respondents are simple weighers of alternatives. I then end with the suggestion that the conceptualization of respondents as simple weighers makes the use of construct validity appear natural.

I

Charles Taylor on Strong and Weak Evaluations

Taylor begins his essay 'What is Human Agency?' in agreement with Harry Frankfurt that the capacity to evaluate our desires is bound up with our capacity for self-evaluation which in turn is essential to human agency.[245] For Frankfurt it is our ability to form second-order desires that is responsible for our particularly human form of self-evaluation. Taylor, on the other hand, argues that we come closer to understanding human self-evaluation if we focus not on desires themselves, but rather on two analytically distinct categories of evaluation, namely weak and strong evaluation.[246]

Taylor's distinction between these two types of evaluation turns on whether or not desires or outcomes—motivations or choices—are distinguished as to their qualitative worth.[247] Weak evaluations deem that something is good just insofar as it is desired—it is an evaluation of objects; strong evaluations determine that something is good insofar as the desire itself is worthy—it is an evaluation of desires.[248] For the latter type of evaluation our motivations or choices are deemed worthy in terms of the quality of life they express relative to the life I want to lead. For instance, I may consider a desire for philanthropy to be worthy because I take it to express a life of selflessness and compassion, a life to which I aspire. Some further examples will help to make the distinction between strong and weak evaluations clearer as well as to bring out some further implications.

---

[245] Taylor, C. (1985) 'What is Human Agency', in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, p. 16.
[246] Ibid.
[247] Ibid, p.18.
[248] Ibid.

(1) I get ready to go for a run and I face the choice of turning left for the 6 mile version or turning right for the 4 mile one. I decide on the long run, my legs feel fresh and I had a big dinner the night before so the extra miles would do me good.

(2) I am offered a job in a hospital to do applied ethics. I have the choice of accepting the position or declining it. I am inclined to accept the position because it's a job and I need a job to pay my bills and this allows me to feel like an adult—a welcome change after my long years of student-induced poverty. But I worry that simply needing a job is not a good enough reason to take one. This position would include work in organizational ethics and the standardization of ethical care, which I worry, would violate some of my deepest beliefs about patient autonomy and ethical care.

In both of these examples I face a choice, but in the second example my desire to have a job and pay my bills—to be an adult—is not enough to make taking the job a good decision. In fact given a certain understanding of integrity and authenticity I may understand the desire to take the job as a disturbing weakness, an indication perhaps that I am too concerned with material interests and thus susceptible to inauthenticity and alienation. In this example and in strong evaluations more generally I evaluate my motivations and choices in terms of what they say about me given my understanding of their place in the good life. Notice that when I evaluate my motivations and choices in terms of what they mean given my particular ethical framework I distinguish them based on their qualitative worth. The desire to accept the job is a threat to my integrity; it is unworthy of the kind of person I am or wish to become.

In contrast to strong evaluation the motivations and choices involved in weak evaluation are not distinguished as to their qualitative worth. As I have construed the running example above my choice to run 6 miles instead of 4 was not based on my estimation that it was a more worthy choice. Rather I ran the longer route because it was the best course of action given my circumstances and goals. Moreover, my decision to run 6 miles does not reflect negatively on the 4 mile run. My decision to run 6 miles is contingent on the twin facts that my legs feel good and I had eaten a lot of food the night before. If I had not eaten a big dinner or my legs were feeling heavy, than I could just as easily have turned in the other direction for the shorter run.

In my running example the important difference between the two runs is their length and so we might be tempted to say that weak evaluation deals with quantitative assessment whereas strong evaluation has dominion over qualitative assessment, but this would be wrong. For Taylor the distinction between strong and weak evaluation does not turn on quantity vs. quality, but rather whether or not our motivations and choices are distinguished by their *qualitative worth.* Consider another example of weak evaluation. I am flipping through the cookbook trying to decide what to make for dinner. I narrow it down to lentil loaf with mash or couscous salad with carrot and ginger soup. I opt for the lentil loaf because I like the way if feels in my mouth and I'm "in the mood for it". While there is a qualitative difference between lentil loaf and couscous salad and indeed it is this qualitative difference that helps to make up my mind about dinner there is not a qualitative difference as to the worth of my desire for lentil loaf or couscous salad. Similarly in my example there is not a qualitative difference between the worth of the desire for a 6 mile or 4 mile run.[249] As Taylor says, there is 'nothing to choose' between the different desires.[250]

This lack of difference in the qualitative worth of competing motivations means that for weak evaluations my contingent circumstances, such as the state of my legs or what I am hungry for, settle which of my competing desires or outcomes, on reflection, is preferred; there is nothing about my motivations that give me pause for reflection. It is for this reason that Taylor claims that for weak evaluation it is enough that something is desired for it to be judged good; weak evaluation as I said in the beginning of this section is an evaluation of objects. I want the lentil loaf because I am hungry for it, so I think it is good if we have it for dinner; I want to run the 6 miles because I ate too much lentil loaf, so I consider it a good thing to do.

A further consequence of this lack of a difference in the qualitative worth of my motivations is that with respect to my weak evaluations I cannot articulate within the framework of considerations offered why one of my motivations or choices—why one of my desires and their outcome—is more worth having. For example, I cannot say why the choice to run 6 miles is inherently superior to the choice to run 4 miles. This fact should not be surprising since I have already said that when we make weak evaluations our considerations do not include those that imply that the alternative is inferior to our selection. Nonetheless, for Taylor this inability to express why one motivation is more worth having than another is an important limitation on weak evaluations.

---

[249] Although given a different context it is possible to construe the desire to run 6 miles as opposed to 4 miles as more worthy. I will discuss this possibility towards the end of this section.

[250] Taylor, C. (1985) 'What is Human Agency', in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, p. 16.

To see how we might consider the inability to discuss the qualitative worth of our desires as a limitation, compare the situation of the weak evaluator with that of the strong evaluator. In making weak evaluations I am, as Taylor says, a simple weigher of alternatives.[251] As a simple weigher I understand my options according to my goals and circumstances. For instance, given that I want to go for a run, the state of my fitness, the time I have available and the terrain at my disposal I decide on two options: the 4 mile run or the 6 mile run. As a simple weigher I reflect on these choices by looking for the circumstantial evidence which will support one option over the other. I might consider the state of my legs and how much I've been eating before I decide on the longer run. But notice that as a person or a self I am removed from my decision. In other words, there is a space between my evaluation or choice and the kind of person I am or aspire to become. As a weak evaluator I understand my ends independent from my identity and seek the best means to achieve those ends. In an importance sense the simple weigher is *herself* impassive between her choices.

But in the context of strong evaluation a person's identity is necessarily bound up with her choices. For instance, in my earlier example I articulate the qualitative worth of declining the job as an expression of integrity as opposed to inauthenticity. As we saw I come to this understanding of what it means to decline the job when I reflect on the sort of life I take it to express. But the sort of life I take it to express will be a function of my vision of the good life and the sorts of fundamental evaluations I take to be central to the kind of person I am or to which I aspire. If I hold myself to the standards of integrity and authenticity and if I understand them to include a certain fidelity to principles, one of which

---

[251] Ibid, p. 23.

includes a kind of transparency between my working life and my contemplative self—and I understand taking this job to be a violation of this principle—then I cannot but see declining the position as an expression of integrity and authenticity. Unlike the weak evaluator the decision to decline the job intimately reflects how I understand myself: I am someone with this kind of integrity or I am not.

Because strong evaluations allow for expressions of qualitative worth which are bound up with one's identity, Taylor argues that they give us depth in the articulation of our motivation, a depth which is lacking in weak evaluation.[252] When we make weak evaluations we can sometimes evaluate our motivations and choices as good or bad in terms of their success at achieving certain goals and ambitions. But in other cases the simple weigher cannot even say this much, for instance, consider my lentil loaf example. In this case my desire for lentil loaf can only be further expressed in terms of what tastes good and what I "feel more like eating"—in terms of my desire for it. Unlike weak evaluations the vocabulary of strong evaluations need not end with a performance criterion or a reiteration of one's desire. Because strong evaluations entail a vocabulary of qualitative worth they can give us insight into why certain motivations are worth having and why certain choices are worth making, which in turn gives us insight into individual visions of the good.

Although Taylor is right to point out this difference between weak and strong evaluations, it is important not to overstate his point. The mere use of strong evaluations does not of itself make the user more articulate in our normal use of the term. While we often *have* strong views regarding the ethical import of various desires and choices, these positions are compatible with a limited ability to articulate them. For example, we may

---

[252] Ibid, p. 26.

judge Marc Quinn's sculpture 'Alison Lapper Pregnant' in Trafalgar Square to be indecent. Thus the sculpture's presence may be taken to reflect not only a poor decision by the public works committee, but also Quinn's dubious character. At the same time, however, we may be unable to articulate the vision of the good which motivates this view about Quinn and the sculpture. On the other hand, an individual who only describes her motivations and choices in terms of weak evaluations—if we can imagine such a person—may be able to discuss clearly, within the confines of weak evaluation, exactly why she does what she does. Taylor's point in emphasizing the depth of articulation available to strong evaluations is to comment on the *possibilities* for articulation that strong evaluations open up, not the actual abilities we possess.[253]

In fact articulating the strong evaluations we have, or put differently, describing their qualitative worth is very difficult. For many of us the reasons why we take a motivation to be honorable or a choice ignoble is obscure and hard to say—and that is if we can even say this much about them. But strong evaluations need not, and very often do not end with whatever is espoused by a particular individual—however inarticulate. We can become clearer and more articulate about the ethical import of our motivations and choices as they are further challenged or further developed by others. For instance, if we turn back to my earlier example we can easily imagine a whole host of competing interpretations of what it means for me to decline my job offer. My parents might say that it is an expression of my naivety; my partner might say I'm afraid of success; I myself might wonder if I'm not taking the whole thing far too seriously.

---

[253] This point was brought to my attention by Jeremy Wisnewski.

The different articulations of my choice to decline the job do more than just challenge how I understand my predicament, for recall that the description of the qualitative worth of my motivations and choices is a result of their place within a larger ethical framework in terms of which I understand my life. In offering to augment or challenge my understanding of what it means to decline this job these alternative articulations also challenge my vision of the good and ultimately my own self-understanding.[254] For each articulation not only describes my situation differently—as one of inauthenticity vs. integrity, naivety vs. maturity, fear of success vs. confidence, over-thinking vs. action—but in describing the situation differently each articulation also paints a different portrait of the person I am in declining this job.

Moreover, each of these articulations aim to challenge and develop the way I understand my predicament by imagining my decision in the service of different purposes.[255] For instance, my parents might see my decision as serving no purpose other than that of maintaining my poverty; my partner might see my decision as providing an unconscious defense mechanism; at times I may wonder if my decision does not just serve the purpose of complicating my life. As I discussed in Chapter 3 our understanding of questions and assertions in part depends on their purpose and here I suggest that how we understand our motivations and choices also depends on what we take their purpose or goal to be. Part of the difficulty of understanding the questions in quality of life measures is that the purpose of these questions is often vague and contentious and only in asking the questions in a particular situation did we learn to understand them better. Likewise part of

---

[254] Ibid.

[255] Taylor, C. (1985) 'What is Human Agency' in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, pp. 26-7.

the difficulty in understanding our motivations and choices is that *their* purposes are vague and contentious and only in applying our motivations and choices to concrete situations that we learn to understand *them* better. Nonetheless, just as with quality of life measures, even in a concrete situation individuals will understand things differently.

Coming to a conclusion about the best way to characterize both my desire for this job and my decision to refuse it is not unlike determining how best to understand the questions in quality of life measures. The best way to understand the questions in a measure depends in part on how we understand the construct and how we understand the construct depends in part on how we understand the questions. Likewise the best way to characterize my desire for the job and my decision to reject it depends in part on how we understand what makes for a good life and how we understand what makes for a good life depends in part on how we characterize my desire and my decision. Both our understanding of the questions in a quality of life measure and our understanding of the ethical import of our motivations and choices exhibits the structure of a circle.

Thus, as with my suggestion in regard to quality of life measures we might situate my desire to take the job and my decision to refuse it within a dialogic framework. Recall that in a dialogic framework texts or text-analogues act as our interlocutors regarding how best to understand the subject matter. In this example we might take my decision to decline the job as a partner in a dialogue regarding what makes for a good life. This framework will also include other texts or text-analogues that speak to the subject matter at hand. These other texts or analogues and the individuals who interpret them provide insight into what makes for a good life and thus a starting point for our understanding of my decision. In attempting to understand my decision we will question it and in answering these questions some of our assumptions about what makes for a good life may be altered or replaced. In the end we will come to understand better not only my desire and my decision, but also what makes for a good life.

Unlike a quality of life measure which exists independent from me and my interpretation of it, the self-interpretations that I accept as a result of this dialogic process are partially constitutive of my experiences. For instance, if I come to accept my desire for this job as, say an expression of independence and maturity, then no longer will I feel ashamed to accept the position. Nor will I understand myself in the exact terms that I did before. I will understand myself in terms of an ethical framework of maturity and responsibility rather than authenticity and this understanding will then give weight to other mature and responsible choices I might make.[256]

Strong evaluations support a conception of the self whereby people can enter into a dialogue with others to give voice to their understanding of their motivations and choices, and also have their understanding of these motivations and choices—as well as their understanding of what makes for a good life—deepened or transformed. Part of Taylor's point in emphasizing this dialogic aspect of persons is not unlike my point about quality of life measures: we cannot determine in advance of the dialogue how we might best understand our motivations and choices or how to understand our questions. For Taylor the self *is* dialogic—we only come to know ourselves and our position with respect to the good in a dialogue with ourselves and others. Put differently, we cannot avoid deliberation on

---

[256] To be sure the frameworks in which we understand our motivations and choices do not change readily or absolutely and single decisions are rarely framework altering. Our motivations and choices are understood within frameworks and only over time do these frameworks evolve.

these matters unless we want to limit the kinds of people we might become or with regard to quality of life measures what we might come to know.

Nonetheless, the model of the self that weak evaluations promote suggests that this element of deliberation and in fact the whole possibility of interpretation as an essential element of evaluation and self-understanding is unnecessary. As we saw earlier weak evaluations make this suggestion by first disengaging our motivations and choices from the purpose with regard to which they have their meaning: wanting to run off last night's dinner does not affect how I understand the meaning of my motivation to run 6 miles although it will affect the length of run that I choose; being in the mood for lentil loaf does not affect how I understand the meaning of my desire for it though it does explain my choice to eat it.

Secondly, weak evaluations disregard the importance of interpretation to evaluation and self-understanding by disengaging our purposes from deliberative and interpretive dialogue: what it means to want to run off one's dinner or to be in the mood to eat lentil loaf is inconsequential to understanding my choice. Once disengaged in these ways the meaning of one's motivations and choices with respect to one's goals reduces to, at best, the degree to which the motivations and choices meet these goals.[257] Put differently, if we want to understand a motivation or choice resulting from a weak evaluation we need not enter into a conversation with the agent to deliberate over the motivation or choice's expressive purpose, rather we merely look to the purpose, conceived of as static and assess the extent to which the motivation or choice fulfills it.

---

[257] At worst, as we saw with the lentil loaf example, motivations and choices just become a reiteration of one's initial desire.

In contrast to the model of a person that strong evaluations support in which an understanding of our motivations and choices requires insight into our understanding of what is meaningful in light of our ethical stance, the simple weigher's motivations and choices are described in objective and absolute terms; weak evaluations attempt to describe our motivations and choices in terms independent of the variable meaning they can have for us, for example, as subjects coming from different experiences, histories, cultures, and having different expertise and training. Weak evaluations instead strive to describe our situations objectively and absolutely by expressing them in terms that have but one meaning. Ordinarily, doing this requires using terms which are independent of what Taylor calls 'anthropocentric properties'.[258] Here Taylor takes advantage of the second dictionary definition of "anthropocentric"—interpreting or regarding the world in terms of human values and experiences[259]—to emphasize that weak evaluations in their endeavor to objectify our motivations and choices attempt to eschew terms that rely on variable human experience and values for their meaning.

In many ways this attempt to limit the descriptions of a situation to those that can be given in objective terms mirrors Salmon and Kitcher's, as well as many quality of life researcher's desire to limit the purpose or relevance relations of the questions we ask to just those which are properly scientific. As we have seen Salmon and Kitcher and many quality of life researchers attempt to limit the scope of legitimate relevance relations or purposes in the hopes of limiting the possible ways we understand questions and answers. In limiting

---

[258] Taylor, C. (1985) 'The Concept of a Person', in *Human Agency and Language Philosophical Papers I*, Cambridge: Cambridge University Press, p. 108.

[259] Mish, F. C. *et al.* (eds) (1999) (10th ed) Merriam-Webster's Collegiate Dictionary, Springfield: Merriam-Webster Inc., p. 49.

the description of a particular situation to those that involve non-anthropocentric properties weak evaluations too strive to limit the questions we can ask and the answers we can give regarding the characterization of our motivations and choices.

To be sure, we often make weak evaluations and the motivations and choices involved are, more often than not, clearly understood. But this ease of understanding should not give the false impression that these evaluations find absolute descriptions with which to describe our desires or outcomes. Rather this ease of understanding is best understood to exist within a framework in which the meaning of certain purposes like running off one's dinner or being in the mood to eat a certain meal is already established and unproblematic. Only when we act outside of our familiar frameworks of reference do we fully recognize the implicit understandings that make weak evaluations—weak. For instance, the desire to run off one's dinner in some parts of the world *is* taken as a comment on one's character; it may indicate arrogance, strangeness and perhaps an unnatural obsession. Here the once weak evaluation turns into a strong evaluation given a different context or framework.

Taylor writes that one who *only* weakly evaluated or indeed was characterized as only weakly evaluating would be a rather shallow character—we might say a sociopathic character.[260] It would be someone who was not engaging with others in the expression of motivations and choices which offer unique insight into our lives as subjects *qua* subjects—insight into what is distinctively human.[261] For instance, an individual who was nothing

---

[260] Taylor, C. (1985) 'What is Human Agency', in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, p. 26.

[261] Taylor, C. (1985) 'Self-Interpreting Animals', in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, p. 59-60.

but a simple weigher would be utterly insensitive, qua simple weigher, to the genuine horror that my partner's mother expresses at the thought that someone in the village might recognize me as I run past.

In the next section I return to my study of quality of life measures and take a look at Andrew Edgar's arguments which conclude that quality of life measures characterize respondents as weak evaluators. Following Taylor Edgar takes this assumption to be a criticism of these measures as well as a limit on their ability to provide respondents with reasonable self-determination in decisions about their health care.

II

Treating Respondents as Simple Weighers

A. Weighting Health States

Andrew Edgar's point of departure for his criticisms of quality of life measures differs from mine in that he focuses primarily on the processes by which health researchers weight health states and I have focused on the validation process of these measures. But these two approaches are related in at least two ways. First many of the most popular weighting techniques involve asking people questions in order to elicit the relative importance of certain health states compared to other health states. Because they use *questions* to determine the appropriate weightings these techniques potentially run into the same sorts of problems that I have discussed throughout these chapters.

Secondly, the weighting of questions and the validation of quality of life measures are interrelated since the weightings of different health states translate into part of the meaning that respondents' answers have for the researchers that analyze them. For example, recall from Chapter 1 that in the NHP the question, 'I'm tired all the time' carries a weight of 39.2, if the question receives a yes answer. Put differently, researchers understand a yes answer to this question as negatively affecting quality of life to the extent that the weight indicates.

Although Edgar's interest in quality of life measures is not unrelated to my own, as I have already indicated, he takes a different tack: he argues that the weightings themselves are dubious. He suggests that the weighting methods employed in standardized quality of life measures treat individuals as weak evaluators and in treating individuals as such the weightings may be biased. This bias is a problem for Edgar because, as he sees it, weighting questions is one of the ways that standardized quality of life measures express their antipaternalistic roots and thus give respondents a voice in their measurement outcomes.[262] For Edgar effective weighting exercises allow respondents to challenge certain assumptions regarding the importance of biological prerequisites for a good quality life.

Thus Edgar, unlike proponents of individualized quality of life, takes standardized measures to be the appropriate kind of measure to provide individuals with reasonable self-determination. But unlike most proponents of standardized measures Edgar is critical of the current methodologies used to weight the items in these measures. In other words, Edgar challenges the notion that individual priorities truly are incorporated into the

---

[262] Edgar, A. (1997) 'A Discourse Ethics Approach to Quality of Life Measurement', in *Annals New York Academy of Sciences*, 809: 34-5. The other way quality of life measures express their antipaternalism, according to Edgar, is through involving lay people in the process of content validity, but this is not his focus.

weightings of questions in the measures. In what follows I discuss some of the different methodologies typically employed to weight the questions in quality of life measures.

The methodology that researchers use to determine the weightings of questions often differs depending on the preferences and conceptual orientation of the researchers working on a measure. As we saw in Chapter 1 when discussing the SEIQoL questions may be individually weighted such that each question in the measure is uniquely tailored to an individual's sense of what is most important to her quality of life. On the other hand, the original Nottingham Health Profile (NHP), which I discussed in Chapter 3 and which Edgar also highlights, uses Thurstone-type rating scales to weight questions. This method requires members of the general public to consider a series of similar statements and determine which one of a pair expresses the more severe condition. For instance, individuals may be asked whether 'Things are getting me down' or 'Worrying keeps me awake at night' is the more punishing with respect to quality of life. A weight for each statement is then derived from the proportion of individuals who weight a certain statement as more severe than another.[263]

There are also other methods for weighting questions. For instance, as Edgar notes the original QALY matrix identified 6 marker health states which members of the public were asked to rank in order of severity. This weighting process includes not only a cardinal indication of severity among the health states, but also an ordinal indication of how much more severe one health state is with respect to another.[264] Finally, many—and some might say most—quality of life measures have *de facto* weightings. In this case researchers

---

[263] Edgar, A. (1995) 'Weighting Health States and Strong Evaluation', in *Bioethics*, 9: 244-6.
[264] Ibid, p. 245.

assume that each question has the same relative importance to quality of life as the other questions do and thus all questions are given equal weighting. The SF-36 is an example of a quality of life measure which uses *de facto* weightings.[265]

B. The NHP Weightings as Weak Evaluations

Despite the variety of techniques for obtaining weightings Edgar claims that '...the dominant approaches to health state measurement remain at the level of weak evaluation' and thus presuppose the simple weigher as the model of human agency.[266] To begin his argument Edgar turns his attention to the weightings of the questions in the NHP and draws similarities between Thurstone-type rating scales and Taylor's treatment of weak evaluations. Recall from the previous section that because 'there is nothing to choose' between the qualitative worth in weak evaluations the reasons considered in contingent circumstances settle which of our competing desires is most preferred. If indeed 'there is nothing to choose' between the qualitative worth in weak evaluations and the weightings in the NHP are taken to be weak evaluations, then we would expect researchers to explain systematic differences among the rankings of health states in virtue of different contingent considerations.

In fact this is how those who developed the NHP understood the effects of cultural variation on the revision of the measure for countries outside the United Kingdom. For instance, Edgar notes that the revision of the NHP for use in France had small but important

---

[265] See http; www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html The SF-36, however, does use a weighting system when deriving a 'summary' score of the measure. This 'summary' score involves just 2 of the SF-36's 8 dimensions, namely physical and mental dimensions.
[266] Ibid, p. 244.

differences in the weightings of questions like 'I'm in pain when I am sitting'. In the United Kingdom this description was ranked as the fifth most severe statement about pain out of eight possibilities; in France it was ranked as the least severe out of eight.[267] To explain the differences between the United Kingdom and French weightings researchers like Bucquet *et al.* turned to language barriers, environmental factors like the number of stairs encountered in daily life and habitat such as family structure and employment conditions.[268]

While Edgar recognizes that these factors may be important to the differences in the NHP weightings he is nonetheless skeptical that they are the only important factors.[269] Although Edgar is not particularly clear on this point his skepticism seems to stem from the fact that, as we saw in the previous section, weak evaluations divorce our motivations and choices from the purposes from which they have their meaning. As I discussed earlier, when weighting the NHP individuals are asked which statement in a series of paired statements is more undesirable with respect to quality of life. Consider the following pair: 'I'm in pain when I am sitting' or 'I have pain at night'. As individuals give their replies and the statements are subsequently weighted the implicit assumption is that respondents understand these statements in the same way—that is they are understood with respect to the same purpose. If they are understood in the same way, then it is possible to aggregate the preferences for certain health states over others, draw conclusions about which states are generally taken to be most and least desirable and make comparisons between different groups of people.

---

[267] Ibid, p. 256.
[268] Ibid.
[269] Ibid, pp. 246-7.

Given that respondents understand these statements in the same way we can go about explaining differences in the rankings of health states—differences in weak evaluations—between different populations or people in terms of contingent circumstances. We can explain these differences as Bucquet *et al.* did in terms of environmental factors and habitat, but we could also explain these differences in terms of cultural and personal values. For instance, we might hypothesize that the French are unconcerned with sitting pain and so it is not as important to them as it is to those in the United Kingdom.

But what we generally do not do is explain the differences in rankings in terms of variable understandings of qualitative worth that the statements themselves might express to different people or groups; put differently we do not question what the statements might mean to different people or groups. We do not hypothesize that the French and the English understand sitting pain differently, as expressing, for example, legitimate pain in one case and mere squeamishness on the other. As Taylor argues, however, an individual's environment is mediated by their vision of the good and if we want to fully comprehend why a certain experience of pain is relatively unimportant, then we may have to understand what that experience of pain means to those who participate in these weighting exercises. To emphasize this point in what follows I briefly discuss a case where understanding the different meanings of a particular health state is crucial to understanding its different ranking among different individuals.

In his article 'A Discourse Ethics Approach to Quality of Life Measurement' Edgar points to Segun Gbadegesin's example of a Yoruba woman who refused a mastectomy—despite the fact that it was the only option to save her life—because she took the intervention to result in a deformity which would make life worthless.[270] In this example it is not the case that the Yoruba woman simply does not value a mastectomy and western women do and it is not merely differences in contingent circumstances between Yoruba women and westernized women that explain the difference in ranking. Rather the Yoruba woman has a different understanding of what it is to have dignity; what counts as humiliation; what garners respect; what accounts for beauty and sexual attraction as well as the good life at large. Thus she understands what it means to have a mastectomy differently than perhaps western women do.

In this example we can only truly understand the Yorba woman's decision when we come to see how a mastectomy might be worse than death; when we understand what this procedure would mean given her understanding of the good life. In this example we see how strong evaluations sometimes do figure in the explanations of health state rankings and, moreover, how strong evaluations allow us to consider different dimensions of what makes up a good quality life.

C. The Consequences of Weak Evaluations: Restricted Debate Regarding a Good Quality Life

Edgar continues to build on his discussion of the NHP by examining some of the consequences of taking weightings to be the result of weak evaluations. Although Edgar does not provide the details for an example of these consequences we can begin to make his point. Take, for example, one of the questions in the NHP: 'I find it hard to dress myself'. Within the logic of the Thurstone-type rating scale a health state's ranking is taken to

---

[270] Edgar, A. (1997) 'A Discourse Ethics Approach to Quality of Life Measurement', in Annals New York Academy of Sciences, 809: 35-6.

express the preferences—and perhaps the prejudices—of the general public.[271] Although proponents of individualized quality of life argue that such an aggregation of preferences is incompatible with individual self-determination, as long as respondents understand the question in the same way it is methodologically possible to aggregate respondents' answers and adjust the question's weight accordingly.

But as Edgar suggests respondents do not understand these questions in the same way. Rather he argues that the discrepancies over the appropriate weighting of a health state—discrepancies which are both common and well-documented between healthy people and those living with disability or illness[272]—are not merely due to differences in contingent circumstances, but also exist due to different assumptions about what counts as a dignified life; what makes for a humiliating experience and so on. Put differently, the discrepancies between, say, a doctor and a patient over the significance 'dressing oneself' has to quality of life has its roots in different understandings of what it means to be able or unable to dress oneself.

Moreover, Edgar suggests that different rankings of health states between individuals of varying degrees of health may also indicate how the very experience of illness changes a subject's conception of the good life and this tallies with what we know about an individual's adaptive capacities.[273] To be sure, as I discussed in Chapter 2 with my example of Female Genital Mutilation (FGM) just because individuals can and do adapt does not mean that they should have to adapt. Nonetheless when viewed as the result of

---

[271] Ibid, p. 37.

[272] Janse, A.J. *et al.*, (2004) Quality of Life: Patients and Doctor's Don't Always Agree: A Meta-Analysis', in Journal of Clinical Epidemiology, 57: 653-61.

[273] Edgar, A. (1995) 'Weighting Health States and Strong Evaluation', *Bioethics*, 9: 248.

strong evaluations our current uses of weightings do not merely promote certain people's preferences at the expense of others, rather they overlook the different ways that we might understand the significance of the health states being weighted. Thus, the current use of weightings stymie certain efforts to consider what visions of the good life and what kinds of adaptations are legitimate and which ones ought to be guarded against.

Edgar's point here is that the discrepancies between doctors and patients or caregivers and their charges may not just indicate *contingent* differences in environmental or physical conditions; it is not merely the addition of rheumatoid arthritis that might cause one to rank 'I find it hard to dress myself' as insignificant with respect to quality of life. The point is that in acquiring rheumatoid arthritis one *becomes* someone who has rheumatoid arthritis. Following Taylor, this new aspect to one's identity is not a mere addition to it, but rather encourages a certain amount of reflection regarding how I might incorporate it into my understanding of myself; how I understand the good life now that this has happened to me.

The situation in which one acquires rheumatoid arthritis is similar to my job example in the previous section. Suppose that I decide to take this job despite my reservations. If this is the case, then I have to understand my decision either as a betrayal of my ideals or I must come to understand these ideals differently in relation to my decision. Likewise the individual with rheumatoid arthritis might continue to see this disease as a "betrayal" of her dignity, as something that necessitates a life without dignity. But such an individual might also reinterpret their understanding of ideals like dignity in light of their new limitations just as I might reinterpret my understanding of integrity and authenticity in light of my decision about the job. This reinterpretation indicates a transformation about what matters, what is significant, what provides meaning in one's life; the discrepancies between those living with different health states perhaps point to a difference in the goals, expectations, standards or concerns that figure in the lives of people with different capacities.[274] Insofar as our quality of life measures do not account for these differences in goals, expectations, standards or concerns they limit our ability to critically evaluate quality of life.

D. The Consequences of Weak Evaluations: Restricted Self-Determination

Current methods for deriving weightings not only restrict discussions regarding what is important in a good quality life, but they are also an obstacle to realizing the self-determination which partially motivates the use of these measures. For instance, consider again the case in which the general population consisting of mostly healthy individuals is enlisted to determine the relative weightings of various health states. Imagine, as in the example above, that 'I find it hard to dress myself' is ranked as a health state expressing low quality of life and weighted accordingly. This ranking becomes a problem if those whose quality of life the measure is meant to assess do not experience difficulty dressing as undignified or adversely affecting their quality of life. Individuals who answer yes to the question will have their quality of life score diminished despite their individual or even collective experience. If quality of life measurement is meant to give respondents a voice in their health care and as Edgar suggests weighting is one of the mechanisms that allows this to happen, then this discrepancy is a serious problem.

---

[274] Edgar, A. (2003) 'The Ontological Status of Quality of Life' at http://www.iprs.it/ITAweb/progetti%20conclusi/documenti/ontolstatquallife.htm, 13/01/2003.

This difficulty with self-determination is similar to the criticism made by proponents of individualized quality of life: individual understandings and rankings of health states may be neglected in favor of majority preferences. But as I argued in Chapter 2 this problem cannot be solved by simply acquiescing to whatever each respondent does mean by their answer: we can be wrong about ourselves; we can and often do have an imperfect understanding of what is valuable and of most importance. Edgar makes this same point against individualism when he emphasizes that no expression of a health preference is self-evidently more worthy than another[275] and also when he says that, 'The antipaternalism inherent in quality of life measures cannot be so construed as to suppress the physician's voice altogether'.[276]

For Edgar the differences between the responses of different individuals to the evaluation of different health states like a mastectomy or finding it hard to dress oneself suggest the different places that health states may occupy in different visions of the good life—they are differences in strong evaluations. But because, as we saw in the previous section, strong evaluations allow for a level of articulation that weak evaluations do not, these differences in evaluating health states do not represent an impasse. Rather, strong evaluations provide an opportunity to articulate the place our preferences occupy in our ethical frameworks and expose the discrepancies in strong evaluations that exist between doctors and lay people as well as among the population at large.[277] In doing this we enter into an ethical dialogue with one another in which we attempt to characterize our different

---

[275] Edgar, A. (1995) 'Weighting Health States and Strong Evaluation', in *Bioethics*, 9: 250.

[276] Edgar, A. (1997) 'A Discourse Ethics Approach to Quality of Life Measurement', in *Annals New York Academy of Sciences*, 809: 36.

[277] Ibid.

preferences. As Edgar notes some characterizations may be indefensible; others may illuminate areas of significance which we had not taken seriously in the past and which may transform some of our preferences as they widen our scope of understanding.[278]

E. Problems with an Emphasis on Weights

Edgar's emphasis on the importance of understanding the meaning of answers to questions and thus the importance a dialogic solution to the evaluation of different health states is similar to my own conclusions in the previous chapters. But by focusing on the weighting process as both the mechanism where antipaternalism is expressed and where individuals are treated as simple weighers Edgar's point has less force than it might. While I agree with Edgar that the weighting process is one place where quality of life measures treat individuals as simple weighers I disagree with his implicit suggestion that properly adjusted it might be an adequate mechanism for antipaternalistic interests.

First, even when weightings *are* accurate and thus *do* reflect individual experience they can still adversely influence the outcomes data from quality of life measures and as a result undermine respondent self-determination. Take, for instance, the Quality of Well-Being scale (QWB) which is used in a 1995 study to determine the quality of life of cochlear implant recipients. In the QWB respondents are given a series of statements and asked to mark those that apply to themselves. In this study researchers found that the most important factor in implant recipients' increase in quality of life was that over time they exchanged more heavily weighted (more undesirable) descriptions of themselves for less

---

[278] Ibid, p. 37.

heavily weighted (less undesirable) ones.[279] One of the substitutions they cite is the following heavily weighted description:

> Trouble learning, remembering, or thinking clearly

which, over time, is traded-up for one of the following less heavily weighted descriptions:

> 1. Pain or discomfort in one or both eyes
> 2. Overweight or underweight for age and height; skin defects such as scars, pimples, warts or bruises
> 3. Pain in ear, tooth, jaw, throat, lops, tongue; missing or crooked permanent false teeth; stuffy, runny nose; any trouble hearing—including wearing a hearing aid

In this example *even if* on Edgar's criteria trouble learning really ought to be weighted more heavily than, for instance, pain in the eye and *even if* there is a consensus regarding how we should understand the significance of trouble learning it is still not clear that this substitution is good evidence that cochlear implants *improve* quality of life. In other words, it is not clear what these substitutions in the QWB really mean.

Recall my discussion in Chapter 2 that the inability to hear can represent a different, but nonetheless legitimate quality of life. Even so being deaf can, practically speaking decrease one's quality of life when social and environmental circumstances are such that the inability to hear limits forms of communication or community. Under these conditions deaf activists argue that while quality of life may not be good for deaf individuals, this assessment has nothing to do with hearing per se. Instead it reflects inequality of opportunity and a biased management of social resources. Returning to the QWB while it may be true that individuals have trouble learning without implants deaf advocates will

---

[279] Harris, J.P., *et al.* (1995) 'An Outcomes Study of Cochlear implants in Deaf Patients', in *Archives of Otolaryngology Head Neck Surgery*, 121: 401.

argue that this fact is unrelated to their inability to hear since the deaf life, at least in principle, represents a life of unique quality.

Moreover, if we accept that a deaf life represents a good quality life and that resources are biased toward helping people overcome deafness instead of living with it, then an improved ability to learn due to implants may be taken to reflect a loss of diversity for citizens at large and a loss of community for the deaf themselves. On this reading, cochlear implants could issue in a loss of a kind of quality of life—even though they reduce trouble learning. Put differently, just because we agree that trouble learning is very important with respect to quality of life does not mean that *any* intervention which reduces trouble learning automatically counts as an improvement in quality of life.

In arguing for the possibility of this conclusion I want to suggest that answers to questions in the weighting exercises *and* answers to questions in the quality of life measures ought to be treated as strong evaluations. Thus, if we really want to understand the significance of respondents' answers in the measure above, then we need to understand, as Edgar recommends, why 'trouble learning' is so important, but we also need to understand the significance of having or not having 'trouble learning'. Thus while Edgar is right to emphasize the importance of strong evaluations to weighting exercises he does not go far enough, answers to questions in our quality of life measures should also be understood on the model of strong evaluations.

The second problem with an emphasis on weights is that the majority of quality of life measures do not weight their questions. This is because a variety of studies have shown that weighting questions does not affect the overall scores on quality of life measures. For instance, a now famous study found that the scores of the weighted version of the NHP, which Edgar discusses, and the scores from the version which only used *de facto* weights correlated with one another almost perfectly.[280] Similarly an unpublished study of the SEIQoL found that weighted and unweighted versions of the measure also correlated with one another to a high degree.[281] The lack of effect that weighting seems to have on the scores of documented comparisons suggests that weighting questions may be irrelevant to antipaternalistic concerns.

Edgar's argument assumes that weighting questions is important because it offers respondents one of the best opportunities to influence quality of life outcomes. I have suggested, however, that a focus on weightings is insufficient. Rather than focus on the weighting of questions I suggest that we refocus on the analysis of respondent answers to the questions in the quality of life measures—however they happen to be weighted. But in doing so we ought to take from Edgar the insight that respondents—whether they are weighting questions or answering them—are illegitimately treated as simple weighers.

Although Edgar and I focus on different aspects of quality of life measurement we both worry that an expressed preference or answer may be misunderstood. Moreover, we both attribute this potential misunderstanding, at least in part, to the inability of our current methods—weighting methods or validation methods—to appreciate or arbitrate among individual differences in purpose, goals, expectations, standards or concerns that in turn affect the meaning of questions and answers. Thus Edgar argues that our weighting methods mistakenly treat respondent answers as weak evaluations and in light of this claim

---

[280] No correlation from this study was less than 0.98. See Jenkinson, C. (1991) 'Why are we Weighting? A Critical Examination of the Use of Item Weights in a Health Status Measure', in *Social Science Medicine*, 32: 1413-16.

[281] Personal communication with an author of the study, 12 May 2006.

we might understand my own work in the previous chapters as suggesting that construct validity erroneously treats answers to questions as weak evaluations. In what follows I want to deepen the suspicion that construct validity does treat answers as weak evaluations by exposing its philosophical foundation.

III

Making Quality of Life Scientific

Construct validity—the process by which we test a measure and its theory by generating hypotheses regarding the measures outcomes—was not always accepted by psychologists as an appropriate psychometric tool. According to Benjamin Superfine, it was not until after 1955 that construct validity started to gain currency and not until after 1989 with Samuel Messick's work that it took its place as the dominant form of validation that we see today.[282] The introduction of construct validity to the already established methods embodied in criterion and content validity was seen as quite radical. This is in part because construct validity implies that there are grades of validity according to the extent to which different hypotheses pertaining to the construct are supported with evidence. It thus replaces the unambiguous decisions of validity rendered through the use of criterion and content validity. Moreover, in addition to grades of validity construct validation implies that the validity of a particular construct is never complete.[283] Because we are dealing with unobservable constructs, multiple instances of evidence are needed in

---

[282] Superfine, B. M. (2004) 'At the Intersection of Law and Psychometrics: Explaining the Validity Clause of No Child Left Behind', *Journal of Law and Education*, 33: 479-80.

[283] Rust, J. & Golombok, S. (1989) *Modern Psychometrics The Science of Psychological Assessment*, London: Routledge Press, p. 81.

order to give good arguments that a measure does what we think it does, but in line with Popper no hypothesis or theory—no construct—can ever be verified.[284]

The introduction of construct validity meant that the validation process was now not only more ambiguous, but also more complicated. As I discussed in Chapter 1 construct validity requires a theory from which we can draw hypotheses about the construct in question. Therefore the importance of theory-building came to the forefront of validity testing. But because psychometricians have to rely on theories, which can never be verified, they can no longer merely assume the existence of a construct as they had done when only using content and criterion validity. Now they have to define the construct and gather evidence to support it. Evidence in favor of a construct makes its existence more likely, but never verifies it and unfavorable evidence is ambiguous since it is not always clear what is to blame—the hypothesis/theory or the measure.

Although it was a radical departure from the relative simplicity of criterion and content validity, construct validity was and still is seen as an advance because it does not blindly assume the existence of a construct, because it requires good evidence before we provisionally accept a construct and because it aids in theory building in as much as it relies on theory. Put differently, the introduction of construct validity into psychometric practice made that practice more scientific. More pointedly, as Rust and Golombok say, construct validity is, 'a reflection of a particular view of the scientific process, and is integrated within the positivist and hypothetico-deductive view of science'.[285]

---

[284] Superfine, B. M. (2004) 'At the Intersection of Law and Psychometrics: Explaining the Validity Clause of No Child Left Behind', *Journal of Law and Education*, 33: 480.

[285] Rust, J. & Golombok, S. (1992) *Modern Psychometrics The Science of Psychological Assessment*, London: Routledge Press, p. 81.

Thus construct validity is taken to embody the logic of the H-D method and scientific inference. The specific logic of the H-D method is that of an argument. The premises consist of a hypothesis from which observable consequences can be drawn—for instance we might hypothesize that cochlear implants improve quality of life—and the initial conditions of the test or measure—we might include the fact that all those in the study have been deaf for at least 10 years. The conclusion drawn from the premises is an observation prediction, for example we might predict that cochlear implant recipients will score higher on quality of life measures than those without implants. When we get the results of the measure or the test we look to see if the prediction is true.[286] If the prediction is true, then we have reason to think our hypothesis and thus our theory is better supported; if the prediction is false then either our hypothesis and theory is flawed or the measure used to elicit the observable consequences is imperfect.

The H-D method, and thus construct validity, illustrates a type of epistemology, which Charles Taylor claims has been dominant in much of science since the 17th century and particularly dominant in experimental psychology today.[287] The first principle of this epistemology is that all data must be intersubjectively univocal; *ceteris paribus* it is in principle possible for anyone to perform an experiment and achieve the same results as anyone else performing it.[288] This principle refers to the fact that scientific experiments and scientific inferences are meant to be replicable; whatever differences there are between experiments they are not accounted for in terms of the anthropocentric qualities of the

---

[286] Ibid, pp. 44-5.

[287] Taylor, C. (1985) 'Peaceful Coexistence in Psychology', *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, p. 117.

[288] Ibid.

individual researchers or the individual participants in the study. Instead, they are accounted for in terms of the contingent features of the experiment.

For our purposes this means that *ceteris paribus* the choice of a researcher to analyze the outcomes data will not affect the outcome. But it also means that the choice of *respondents* should not affect the outcome. If we do find a difference between different groupings of respondents, for instance in the revision of the NHP in which French people ranked 'I'm in pain when I am sitting' as the least severe option opposed to those in the United Kingdome who ranked it 5th out of 8, then we attribute this difference to contingent circumstances between the two measures or between the groups measured. If you recall, in this particular example the differences were attributed to language barriers, environmental factors and habitat.

The reluctance to rely on anthropocentric qualities to explain differences in quality of life assessment can be traced to one of the defining features of standardized quality of life measures, namely their understanding of 'subjectivity'. Recall from Chapter 1 that in the context of standardized measures 'subjectivity' is understood as what I call 'embodiment'. 'Embodiment' refers to the fact that we are all in similar bodies and as a result have certain similar unobservable experiences. If subjectivity is due to our embodiment, then quality of life as a measure of subjective experience can be standardized, that is replicated. Thus quality of life measurement is purified from problems of understanding meaning because in virtue of being in similar bodies the kinds of experiences we have remain similar.[289]

---

[289] To be sure individualized quality of life measures understand subjectivity as idiosyncratic and on the face of it may seem to embrace anthropocentric qualities and thus may be thought to serve as the appropriate

One of the consequences of intersubjective univocality and hence construct validity is that scientific data—the questions and answers of quality of life measures as well as the hypotheses about the outcomes—are taken as weak evaluations.[290] They must be taken as weak evaluations, as requiring no further articulation or explanation in order to understand their meaning, because otherwise we allow for potentially unarbitrable disputes.[291] As a result the experiments or the measures would not be replicable: the results and the experiment or measure itself would be understood differently depending on who is involved in the study.

This impediment to replication is indeed a problem for many areas of scientific research. But quality of life assessment relies on questions and their answers as the basic unit for gathering observable information about unobservable constructs and as I argued in the previous chapter we cannot limit our understanding of questions and answers to a single meaning. Valid quality of life measures cannot be treated as univocal pieces of research. How we understand the meaning of the data *does* depend on the researchers and respondents involved in the measure.

Respondents are not merely simple weighers, but when we use construct validity as our primary validation procedure we treat them as though they were and in so doing deprive them of the self-determination that quality of life measures were thought to provide. It is here in the validation of quality of life assessment and not in the weighting exercise that the notion of the simple weigher not only does its most damage, but is also at

---

alternative to traditional measures. But as I have already argued in Chapter 2 it is unclear that idiosyncrasy best expresses our anthropocentricity.

[290] Ibid, p.118.

[291] Ibid.

its deepest in terms of conceptual orientation. Construct validity was thought to make psychological testing more scientific, but when it comes to quality of life measurement it serves better to conceal our understanding of different constructs than to help us discover new facets of them.

Taylor's notion of weak evaluations and simple weighers as well has his insights into the H-D method help to highlight and unify much of my work throughout these chapters. Nonetheless, while I have argued for the indeterminate nature of our questions and answers on the grounds of what it is to ask and answer questions, Taylor argues that the descriptions of our motivations and choices cannot be absolute because this conception is inconsistent with our best understanding of what it is to be a person. To be sure, these two arguments are different, yet we might see them as two sides to the same coin. For if Taylor is right that strong evaluations represent our self-interpretations and I am right that valid interpretations require a dialogic solution, then to the extent that our identity consists of our self-interpretations we *are* both the source of these questions and their answers. If persons have the depth of character which Taylor attributes to them, then *we* are a kind of text-analogue and thus subject to the process of asking and answering questions in order to best understand ourselves—our desires, our motivations, our choices, our preferences.

It follows from what has been said so far that investigations into the quality of life that a set of answers expresses require some insight into a respondent's vision of the good life. But we can only really achieve this insight insofar as we—quality of life researchers and respondents—are open to reevaluating what people might mean by their questions and answers by asking further questions. Building on that with which I ended in Chapter 4, investigations into quality of life are not just dialogic—they are also *ethical* dialogues; quality of life is an essentially ethical or we might say broadly speaking moral question.[292]

---

[292] See Taylor, C. (1989) *Sources of the Self the Making of the Modern Identity*, Cambridge: Cambridge University Press, pp. 14-15, for a brief defense of expanding our normal use of "moral" i.e. our respect for and our obligations to others, to include our understanding of what makes for a full life as well as a range of notions concerned with dignity.

Conclusion

In this thesis I have linked complications in understanding the meaning of questions and answers in quality of life measures with issues of validation. My claim has been that these complications render the validation standards of classical test theory inadequate. To be sure, our measures can be better or worse—more or less free from bias and prejudice—but they cannot be independent instruments whose outcomes, like the outcomes of a sphigmometer or stethoscope, are meant to be clear to anyone educated in how to read them. Throughout the pages of these chapters I have discussed my reasons for this claim and there is no reason to repeat them here. In what follows, however, I would like to point to two further topics that weave throughout this thesis and may not have received sufficient attention.

It is often said that quality of life lacks a gold standard. This statement is typically taken to mean that quality of life is unobservable and that we thus do not have evidence of the construct independent of the measures that assess it. We have no way of directly validating the measures we use, but we also do not know what a measure's outcomes signify. The lack of a gold standard suggests both that we can only grasp quality of life through our measures and that understanding these measures is itself a problem. The typical solution to this difficulty appeals to construct validity: construct validity provides not only an account of a measure's accuracy in lieu of a gold standard, but in doing so it also tells us what the questions and answers might mean.

When we say that quality of life lacks a gold standard we might be understood as saying that we do not know what quality of life *is*. Normally when we do not know what something is we ask questions about it and indeed quality of life measures ask respondents questions about their quality of life. But at the heart of genuine inquiry, unlike the questions in quality of life measures, is what Gadamer calls the '...questionability of what is questioned.'[293] By this he means that in questioning a subject matter the answers we seek are not settled and thus our questions, too, are open to a certain amount of reinterpretation; when we ask a question we open up inquiry into a subject matter which we imperfectly understand and thus the questions we ask are also imperfectly understood. We come to a better understanding of the meaning of our questions and answers as we come to understand the subject matter.

But unlike genuine inquiry the use and validation of quality of life measures assumes—indeed requires—that the respondents understand the questions and answers in our measures in the same way as researchers imagined they would. Put differently, neither the questions nor answers in our measures are open to reinterpretation and as Gadamer writes, a question without this openness is '...basically no more than an apparent question'.[294] Apparent questions do not reveal the indeterminacy or the "questionability" of a subject matter and instead suggest that the subject matter is already understood. Quality of life, however, is not supposed to be understood independent our measures for this is just the difficulty that the lack of a gold standard underwrites.

My point here is to emphasize the discrepancy between certain claims made about quality of life and the reality of quality of life measurement. We might say that the lack of a gold standard directs us to a problem, a problem which is perhaps best understood in

---

[293] Gadamer, H. (2003) (2nd ed) *Truth and Method*, trans Joel Weinsheimer & Donald G. Marshall, Continuum Press: New York, p. 363.
[294] Ibid.

terms of a circle—we cannot understand quality of life unless we understand the measures that asses it, but we cannot understand our measures unless we understand quality of life. Importantly, construct validity does not solve this dilemma, for construct validity implies that we already know how to understand our questions and answers. Rather than resolving the difficulty of how to understand quality of life, this response requires that we already understand it.

Thus, my claim in this thesis is two-fold. Not only do problems with understanding the definitive meaning of questions and answers in quality of life measures render the validation standards of classical test theory inadequate. In addition quality of life research often fails to grasp the significance of the fact that quality of life lacks a gold standard. This lack asks that we face what I we might call the hermeneutic problem. This problem cannot be solved with a methodological slight of hand whereby we grant the lack of a gold standard at one moment only to treat quality of life research as though it had a gold standard at another.

Quality of life measures face another difficulty which weaves through this thesis, namely the difficulty of creating measures that allow for a reasonable amount of respondent self-determination. This difficulty is intimately connected to the lack of a gold standard because when we treat the questions and answers in our measures as though we know what they mean we open up the possibility of misunderstanding them. If at least one of the important motives for measuring quality of life is to give respondents a voice in their health care, then the standardization of the meaning of questions and answers is a problem. We do not give people self-determination when we determine a priori what our questions and their answers already mean. To do so is to ask only apparent questions and to do this is to be less than genuine in our inquiry into quality of life. Moreover, it is to treat people as something less than self-determining agents. Thus limiting how we understand the meaning of the questions and answers in our measures may not only be an epistemological loss, but also an ethical infringement on individuals' capacity to express what they understand to be a good quality life.

Quality of life research attempts to standardize the meaning of the questions and answers in our measures in order to make quality of life a scientific enterprise. But what this thesis has tried to show is that questioning is more of an art than a science and that this art is part of the art of conversation in general. As Gadamer writes there is no '...method of learning to ask questions...'[295] We simply learn to ask good questions by taking part in inquiry and in attempting to understand the things around us. Moreover, we cannot understand these things—we cannot participate in inquiry—unless we ask questions. If we want to measure quality of life, then we must also understand it. But this undertaking does not require additional efforts to make quality of life measures independent and self-contained. Rather it requires that we incorporate these measures into a kind of dialogue that helps us to expand and critically to evaluate our conceptions of the good life.

In his discussion of the limitations of our methods in helping us to ask good questions Gadamer refers us to Socrates and his claim that the important thing in asking good questions is the knowledge that one does not know.[296] For Gadamer it is the Socratic dialogue which creates the conditions for the possibility of good questions and thus the possibility of knowledge. Gadamer's reference to Socrates indicates just how long the

---

[295] Ibid, p. 365.
[296] Ibid.

issue of questioning has been hovering around the edges of philosophy. In this thesis I have taken up the case study of quality of life measures—a modern phenomenon, which one might say is the result of modern technologies. But I have used this case study to highlight an old and often neglected topic, namely our understanding of questions. Questions are often overlooked in philosophy in favor of answers—explanations, reasons, causes. But if Gadamer is correct when he writes that the priority of the question over the answer is the basis of the concept of knowledge[297], then perhaps we would do well to pay them more attention.

---

[297] Ibid.

Bibliography

Abelson, J., *et al.* (2003) 'Deliberations about Deliberative Methods: Issues in the Design and Evaluation of Public Participation Processes', in *Social Science Medicine*, 57: 239-51.

Agency for Healthcare Research and Quality (2001) 'Healthcare Costs: Why Do They Increase? What can We Do?', United States Department for Health and Human Services, at http://www.ahrq.gov/news/ulp/costs/ulpcosts1.htm

Ahmedzai, S. (1993) 'Quality of Life Measurement in Palliative Care: Philosophy, Science or Pontification?', in *Progress in Palliative Care*, 1: 6-10.

Allison, P.J., *et al.* (1997) 'Quality of Life: A Dynamic Construct', in *Social Science Medicine*, 45: 221-9.

Alexander Graham Bell Association for the Deaf and Hard of Hearing at www.agbell.org

Amundson, R. (2005) 'Disability, Ideology, and Quality of Life: A Bias in Biomedical Ethics' in *Quality of Life and Human Difference Genetic Testing, Health Care and Disability*, David Wasserman et al., New York: Cambridge University Press, pp. 101-124.

Aristotle (1999) (2nd ed) *Nicomachean Ethics*, trans by Terence Irwin, Indianapolis: Hackett Publishing Company.

Austin, J. L. (1980) (2nd ed.) *How to do Things with Words*, J. O. Urmson & M. Sbisa eds, Oxford: Oxford University Press.

Beauchamp, T.L. & Childress, J. F. (1983) (2nd ed) Principles of Biomedical Ethics, New York: Oxford University Press.

Bellan, L. (2005) 'Why are patients with no visual symptoms on cataract waiting lists?', *Canadian Journal of Ophthalmology*, 40: 433-38.

Birger, M., *et al.* (2004) 'Cochlear Implants and Health Status: A Comparison with Other Hearing-Impaired Patients', in Annual of Rhinology and Otolaryngology, 113: 914-21.

-- (2005) 'Cochlear Implants and Quality of Life: A Prospective Study', in *Ear & Hearing*, 26: 186-94.

Bowling, A. (1991) *Measuring Health A Review of Quality of Life Measurement Scales*, Milton Keynes: Open University Press.

Bramston, P., *et al.* (2005) 'Conceptual Principles of Quality of Life: an Empirical Exploration', in *Journal of Intellectual Disability Research*, 49: 728:-33.

Brock, D. (1993) 'Quality of Life Measures in Health Care and Medical Ethics', in *The Quality of Life*, Martha C. Nussbaum & Amartya Sen eds, Clarendon Press: Oxford, pp. 95-132.

-- (2005) 'Preventing Genetically Transmitted Disabilities while Respecting Persons with Disabilities', in *Quality of Life and Human Difference Genetic Testing, Health Care, and Disability*, David Wasserman et al. eds, Cambridge: Cambridge University Press, pp.67-100.

Brooks, R., *et al.* (1996) EuroQol: the Current State of Play, in *Health Policy*, 37: 53-72.

Browne, J., *et al.* (1997) 'Conceptual Approaches to the Assessment of Quality of Life', in *Psychology and Health*, 12: 737-51.

Browne, J., *et al.* (1997) 'Development of a Direct Weighting Procedure for Quality of Life Domains', in *Quality of Life Research*, 6: 301-10.

Browne, J. (1997) *Individual Quality of life in Older People, Conceptual and Methodological Challenges*, PhD Thesis in Psychology, Dublin University.

Cartwright, N. (2004) 'Causation: One Word Many Things', *Philosophy of Science*, 71: 805–19; reprinted as *Causality: Metaphysics and Methods* (2003), Technical Report, CTR, CPNSS, LSE.

Cella, D.F. (1992) 'Quality of Life: The Concept', in *Journal of Palliative Care*, 8: 8-13.

Chee, G.H., *et al.* (2004) 'Benefits of Cochlear Implantation in Early-Deafened Adults: The Toronto Experience', in *The Journal of Otolaryngology*, 33: 26-31.

Cheng, A.K. *et al.* (2000) 'Cost-Utility Analysis of the Cochlear Implant in Children', in *The Journal of the American Medical Association*, 284: 850-56.

Clark H. H. & Schober, M.F. (1992) 'Asking questions and influencing answers', in J. M. Tanur (ed.) *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, New York: Russell Sage Foundation, pp. 15-47.

Cohen, C. (1982) 'On the Quality of Life: Some Philosophical Reflections', in *Circulation* 66 (suppl III): III29-33.

Cohen, J. (1998) 'The Deaf Identity Double-Bind: Culture Versus Disability', in *The Advocate's Forum*, 5.

Cross, C.B (1991) "Explanation and the Theory of Questions", *Erkenntnis*, 34: 237-60.

Cummins, R.A. (2005) 'Moving from the Quality of Life Concept to a Theory', in *Journal of Intellectual Disability Research*, 49: 699-706.

Dean, K. (2004) 'The Role of Methods in Maintaining Orthodox Beliefs in Health Research', in *Social Science Medicine*, 58: 675-85.

Donovan, J.L., *et al.* (1993) 'Assessing the need for health status measures', *Journal of Epidemiology and Community Health*, 47: 158-62.

Edgar, A. (1995) 'Weighting Health States and Strong Evaluation', *Bioethics*, 9: 240-51.

-- (1997) 'A Discourse Ethics Approach to Quality of Life Measurement', Annals New York Academy of Sciences 809: 30-9.

Edgar, A. *et al.*, (1998) *The Ethical QALY Ethical Issues in Healthcare Resource Allocations* Haslemere: Euromed Communications

Edgar, A. (2003) 'The Ontological Status of Quality of Life' at http://www.iprs.it/ITAweb/progetti%20conclusi/documenti/ontolstatquallife.htm.

Edlund, M. & Tancredi, L.R. (1985) 'Quality of Life: An Ideological Critique', in *Perspectives in Biology and Medicine*, 28: 591-607.

Fausto-Sterling, A. (2000) *Sexing the Body*, New York: Basic Books.

Ferrans, C.E., *et al.* (2005) 'Conceptual Model of Helath-Related Quality of Life', in *Journal of Nursing Scholarship*, 37: 336-42.

Ferrara, P. J. (1994) 'Power to the People Positive Alternatives to the Oregon Health Plan', at http://www.cascadepolicy.org/health94.htm.

Gadamer, H. (2003) (2nd ed) *Truth and Method*, trans. J. Weinsheimer & D. G. Marshall, New York: Continuum Press.

Gallaudet Research Institute at http://gri.gallaudet.edu/Literacy

Guion, R.M. (1977) 'Content Validity: Three Years of Talk—What's the Action?', in *Public Personnel Management*, 6: 407-414.

Hacking, I. (1988) 'The Self-Vindication of the Laboratory Sciences', in *Science as Practice and Culture*, Chicago: University of Chicago Press, pp. 29-64.

Hajiran, H. (2006) 'Toward a Quality of Life Theory: Net Domestic Product of Happiness', in *Social Indicators Research*, 74: 31-43.

Hallberg, L.R. & Ringdahl, A. (2004) 'Living with Cochlear Implants: Experiences of 17 Adult Patients in Sweden', in *International Journal of Audiology*, 43: 115-121.

Harding, S. (1998) *Is Science Multi-Cultural?*, Bloomington: Indiana University Press.

Harris, J.P. (1995) 'An Outcomes Study of Cochlear Implants in Deaf Patients', in *Archives of Otolaryngology Head Neck Surgery*, 121: 398-404.

Hawthorne, G., *et al.* (2004) 'Evaluating the Health-Related Quality of Life Effects of Cochlear Implants: A Prospective Study of an Adult Cochlear Implant Program', in *International Journal of Audiology*, 43: 183-92.

Hayry, H. (1991) 'Measuring the Quality of Life: Why, How and What?', in *Theoretical Medicine,* 12: 97-116.

Heidegger, M. (1996) *Being and Time*, trans Joan Stambaugh, Albany: State University of New York Press.

Herman, B (2006) 'Ends Justifying Means', University of California, Riverside 16th Annual Conference *Normativity & Universality from a Kantian Perspective*, 24 February.

Hunt, S.M. & McEwen, J. (1980) 'The Development of a Subjective Health Indicator', in *Sociology of Health and Illness*, 2: 231-246.

Hunt, S.M. (1997) 'The Problem of Quality of Life' in *Quality of Life Research*, 6: 205-12.

Hyland, M.E. (1992) 'A Reformulation of Quality of Life for Medical Science', in *Quality of Life Research*, 1: 267-72.

Janse, A.J., *et al.* (2004) 'Quality of Life: Patients and Doctor's don't Always Agree: a Meta-Analysis', in *Journal of Clinical Epidemiology*, 57: 653-61.

Jenkinson, C. (1991) 'Why are we Weighting? A Critical Examination of the Use of Item Weights in a Health Status Measure', in *Social Science Medicine*, 32: 1413-16.

Jenkinson, C. (1994) *Measuring Health and Medical Outcomes*, London; University College London Press.

Joyce, C.B.R., *et al.* (1999) *Quality of Life in Individual Quality of Life Approaches to Conceptualisation and Assessment*, Australia: Harwood Academic Publishers.

Kaplan, D.M., *et al.* (2003) 'Early-Deafened Adult Cochlear Implant Users: Assessment of Outcomes', in *The Journal of Otolaryngology*, 32: 245-49.

Kuhn, T. (1996) (3rd ed) The Structure of Scientific Revolutions, Chicago: Chicago University Press.

Lane, H. (1993) *The Mask of Benevolence*, New York: Vintage Books.

Lloyd, E. A. & Anderson, C. G. (1993) 'Empiricism, Objectivity, and Explanation', in *Midwest Studies in Philosophy*, XVIII: 121-31.

Longino, H.E. (1990) Scien*ce as Social Knowledge Values and Objectivity in Scientific Inquiry*, Princeton: Princeton University Press.

Lowe, F. & Rapin, H. (1994) 'The Paradoxes of Quality of Life and its Phenomenological Approach', in *Journal of Palliative Care*, 10: 37-41.

Maillet, C.J., *et al.* (1995) 'Change in the Quality of Life of Adult Cochlear Implant Patients', in *Annuals of Otolaryngology Rhinology Laryngology Suppl.*, 165: 31-48.

Mallinson, S. (2002) 'Listening to respondents: a Qualitative Assessment of the Short-Form 36 Health Status Questionnaire', in *Social Science & Medicine*, 54: 11-21.

McClellan, T.D. 'Deaf Mom Fights to Keep Kids form Ear Implants', in *The Grand Rapids Press*, at http://www.bridges4kids.org/articles/9-02/GRPress9-6-02.html.

McDowell, I. & Newell, C. (1996) *Measuring Health: A Guide to Rating Scales and Questionnaires*, New York: Oxford University Press.

Michalos C. A. (2002) 'Social Indicators Research and Health-Related Quality of Life Research', in *Social Indicators Research*, 67: 27-72.

Mish, F. C., *et al.* (eds) (1999) (10th ed) Merriam-Webster's Collegiate Dictionary, Springfeild: Merriam-Webster Inc.

Montgomery, C. (2002) 'The Cochlear Implant Trial', in *The Ragged Edge Magazine*, at http://www.raggededgemagazine.com/extra/deaftrail.

National Institute for the Deaf (NAD) Board of Directors (2000) 'Cochlear Implants NAD Position Statement', at www.nad.org/site.

Nussbaum, M.C. (1999) Sex and Social Justice, New York: Oxford Press.

O'Boyle *et al.* (1992) 'Individual Quality of Life in Patients Undergoing Hip Replacement', in Lancet, 339: 1088-1091.

O'Boyle A, *et al.* (1994) 'Quality of Life: Assessing the Individual', in *Advances in Medical Sociology*, 5: 159-180.

Office of New Drugs and the Office of Medical Policy in the Center for Drug Evaluation and Research in cooperation with the Center for Biologics Evaluation and Research and the Center for Devices and Radiological Health at the Food and Drug Administration (2005) *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims--Draft.*

O'Neill, O. (2004) 'Informed Consent and Public Health', in *Philosophical Transcripts of the Royal Society London*, 359: 1133-6.

Rawls, J. (1999) (revised ed.) *A Theory of Justice*, Oxford: Oxford University Press.

Rayens, M.K. & Hahn, E.J. (2000) 'Building Consensus Using the Policy Delphi Method', in *Policy, Politics, & Nursing Practice*, 1: 308-15.

Rosenberg, R. (1995) 'Health-Related Quality of Life Between Naturalism and Hermeneutics', in *Social Science Medicine*, 41: 1411-14.

Rust, J. & Golombok, S. (1989) *Modern Psychometrics The Science of Psychological Assessment*, London: Routledge Press.

Salmon, M. H., *et al.* (1999) *Introduction to the Philosophy of Science*, Indianapolis: Hackett Publishing Company.

Salmon, W. & Kitcher P. (1987) 'Van Fraassen on Explanation', in *The Journal of Philosophy*, 87: 315-30.

Sechrest, *et al.*, (1996) 'Calibration of Measures for Psychotherapy Outcome Studies', in *American Psychologist*, 51: 1065-1071.

Sen, A. (1999) 'Capability and Well-Being' in Martha C. Nussbaum & Amartya Sen eds. *The Quality of Life*, Oxford: Clarendon Press, pp. 29-53.

-- (1999) *Development as Freedom*, Oxford: Oxford University Press.

SF-36 Instructions for Use at www.rand.org/health/surveys_tools/mos/mos_core_36item_scoring.html

Sprangers, M.A.G. & Schwartz, C.E. (1999) 'Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model', in *Social Science Medicine*, 48: 1507-15.

Streiner, D.L. & Norman, G.R. (2003) Health Measurement Scales a Practical Guide to their Development and Use, Oxford: Oxford University Press.

Superfine, B. M. (2004) 'At the Intersection of Law and Psychometrics: Explaining the Validity Clause of No Child Left Behind', *Journal of Law and Education*, 33: 479-80.

Taillefer, M., *et al.* (2002) 'Health-Related Quality of Life Models: Systematic Review of the Literature', in *Social Indicators Research*, 67: 293-323.

Tanner A., *et al.* (1993) 'The Phenomenology of the Knowing Patient', in *Journal of Nursing Scholarship*, 25: 273-280.

Taylor, C. (1985) 'Peaceful Coexistence in Psychology', *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, p. 117-138.

-- (1985) 'Self-Interpreting Animals' in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, pp.45-76.

-- (1985) 'The Concept of a Person' in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, pp. 97-114

-- (1985) 'What is Human Agency' in *Human Agency and Language Philosophical Papers 1*, Cambridge: Cambridge University Press, pp. 15-44.

-- (1989) *Sources of the Self the Making of the Modern Identity*, Cambridge: Cambridge University Press.

Tennant, A. (2004) 'Application of Rasch Analysis in the Development and Application of Quality of Life Instruments', *Value in Health*, 7 (suppl I): S24.

University of Washington School of Medicine (1998) 'Ethics in Medicine: Resource Allocation' at http://depts.washington.edu/bioethx/topics/resall.html

Van Fraassen, B. (1980) *The Scientific Image*, Oxford: Oxford University Press.

Vermeire, K., *et al.* (2005) 'Quality-of-Life Benefit from Cochlear Implantation in the Elderly', in *Otology & Neurotology*, 26: 188-95.

Warnke, G. (1987) *Gadamer Hermeneutics, Tradition and Reason*, Stanford: Stanford University Press.

-- (1999) *Legitimate Differences Interpretation in the Abortion Controversy and Other Public Debates*, Berkeley: University of California Press.

Washington University School of Medicine in St. Louis, Department of Otolaryngology, 'Cochlear Implant—What is it?' retrieved on 29 June, 2004 at http://wuphysicians.wustl.edu/dept.

Wright, L. (2001) *Critical Thinking an Introduction to Analytical Reading and Reasoning*,

Oxford: Oxford University Press