UNIVERSITY OF CALIFORNIA SAN DIEGO

The Capabilities Approach to Well-being: Characterizing Capabilities and Measuring Them

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Philosophy (Science Studies)

by

Travis Dean Chamberlain

Committee in charge:

Professor Nancy Cartwright, Chair
Professor Craig Callender
Professor Monte Johnson
Professor Martha Lampland
Professor David Wiens

2021

The dissertation of Travis Dean Chamberlain is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

For Ida, Skai, and family,
more valuable according to any measure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my appreciation to the UCSD Philosophy and Science Studies Departments for providing resources, networking, and all manner of training. I'm especially thankful for the diverse disciplinary insights given by the UCSD Science Studies Program including the focus on ethnographic interpretations of scientific practice.

I would like to express my sincere gratitude to my committee chair, Nancy Cartwright, who inspired me to embark on this journey when I was studying in London, and has truly been a wayfinder. It has been my privilege to work under your direction. My sincere thanks also to the members of my committee, to Monte Johnson for clarifying some of the deeper connections in the history of philosophy and for insisting that I look into capabilities in the first place, to Craig Callender for thoughts on writing style and audience, to Martha Lampland for insights regarding the sociology of economic practice - especially the politics surrounding the creation of indices -, and to David Wiens for input regarding representations and the importance of technical rigor. I would also like to thank Georgios Anagnostopoulos for thoughts on Aristotle's politics and connections to Nussbaum and prohairesis. Thank you also to Richard Arneson for taking extra time to discuss capabilitarian approaches to well-being, especially as related to sufficientarian thresholds and cost-benefit analysis.

I'm also grateful to a number of academics who are more thoroughly embedded in the practice of theorizing about capabilities and developing measures of capabilities. I would like to thank Sebastian Östlund for numerous consultations and comments on drafts. I would also like to thank José Manuel Roche for practical insights related to the development of indices and participatory methods within the capabilities approach. Thank you also to Mozaffar Qizilbash for critical input, to Nicolai Suppa and Jakob Dirksen of OPHI for further conversations and research suggestions related to the MPI, and to Jaya Krishnakumar for responses regarding the application of various econometric techniques.

I also wish to thank my colleagues at UCSD for ongoing dialogues and feedback. Thank you especially to Ryan Stringer, Gil Hersch, Danny Weltman, Jonathan Knutzen, Blythe Greene, Kyle Sereda, and Aaron Finley for collaboration in various stages of this project, and to numerous philosophy and science studies colleagues for their collaborative engagement. I would also like to thank Michael and Mary McKay of the Rady School of Management, because the kind of leadership they teach is resonant of the applied feminism and participatory methods that are central to this thesis' practical interpretation of the capabilities approach.

Most importantly, thank you Ida for your patience, prudence, and wisdom. This could only be done with a resilient team, and I'm unbelievably lucky to have you as a co-captain. Of course, my thanks to Skai for her patience and effervescent creativity, to Grandpa Sky for the (rather insistent) nudge, and to Grandma Theo and Grandpa Dean for your endless support. Thank you to Mom and Ron for the encouragement, and to Dad for so many years of zany dialogue.

VITA

2006	Bachelor of Arts in Economics and in Philosophy, University of California Irvine
2009	Masters of Science in Philosophy and Public Policy, London School of Economics
2014 - 2018	Teaching Assistant, University of California San Diego
2019 - 2021	Lecturer in 'MGT 18: Managing Diverse Teams,' Rady School of Management, University of California San Diego
2019	Lecturer in 'Philosophy, Politics and Economics,' University of San Diego
2021	Doctor of Philosophy in Philosophy (Science Studies), University of California San Diego

ABSTRACT OF THE DISSERTATION

The Capabilities Approach to Well-being:
Characterizing Capabilities and Measuring Them

by

Travis Dean Chamberlain

Doctor of Philosophy in Philosophy (Science Studies)

University of California San Diego, 2021

Professor Nancy Cartwright, Chair

This thesis concerns the capabilities approach, which is widely adopted and cited as a way to conceptualize human well-being, also called welfare. The concept of 'capabilities' has been developed in different but overlapping ways by Amartya Sen and Martha Nussbaum. The views of both authors are rooted in Aristotle's theory of living well, which I outline and then use in my exegesis of their concepts.

If the concept of capabilities is to do serious work in evaluating individual and population-level 'capabilitarian well-being,' a measure must be developed that fits with the concept. Accordingly, I investigate how well different ingredients of capabilities measures - including their representations and relevant operational procedures - cohere with core features of Sen and Nussbaum's conceptual characterization. Both authors characterize capabilities as value-laden, and both authors explain that different kinds of capabilities cannot in general be traded off against one another to create total orderings of capabilitarian well-being. Ergo, measures that represent capabilities as not value-laden are inconsistent with respect to the value-ladenness of Sen and Nussbaum's characterization. Similarly, measures that provide total orderings by trading off different kinds of capabilities indicators are inconsistent with respect to Sen and Nussbaum's general commitment to avoid such trade-offs. Outputs for such measures are what I call 'off-target,' meaning there is good reason to believe such measures do not measure capabilities as characterized by Sen and Nussbaum.

Yet even if off-target, such measures may remain useful as rough metrics depending on context or purpose. For example, even total orderings of capabilities may be justified with respect to the purpose of furthering an international conversation, or as rough measures used to identify poverty as different kinds of capability deprivation. I conclude by analyzing two contemporary capabilities measures – Jaya Krishnakumar's improved Human Development Index and Sabina Alkire and James Foster's Deprivation Depth Measure - as exemplars of measurement systems that are inconsistent with respect to Sen's and Nussbaum's characterization of capabilitarian well-being yet remain useful as approximate indicators with respect to a limited set of purposes.

Introduction

"Whether a measure of well-being is valid should not be mainly a technical question."

- Anna Alexandrova (2017: xlv)

The capabilities approach to conceptualizing and measuring people's well-being continues to have wide-ranged impact on welfare-oriented public policy. Capabilities have been called "a new theoretical paradigm… that has had increasing impact on international agencies discussing welfare, from the World Bank to the United Nations Development Programme" and also the "Sarkozy Commission Report on the measurement of economic performance and social progress."[1] As Ingrid Robeyns (2017) explains, the capabilities approach is used by academics and politicians to inform "how to make policies, views about what social justice requires, or bottom-up views about development" and is particularly valuable because it "prompts us to ask alternative questions, and to focus on different dimensions when we make observations or when we gather the relevant data for making evaluations or judgments."[2] The approach has also shown promise for measuring and addressing human poverty, for example through the Multidimensional Poverty Index (MPI) created by the Oxford Poverty and Human Development Initiative (OPHI).[3] In sum, the capabilities approach represents a progressive shift in the tradition of welfare economics; it broadens the informational space used to conceptualize and measure lives worth living.

According to the capabilities approach, well-being is constituted by people's capabilities - their opportunities to achieve doings and beings, also called 'functionings,' that they have reason to value. Functionings refer to different kinds of abilities and activities such as being part of a social community, learning to reason well, maintaining good bodily and psychological health, maintaining personal relationships, etc. Yet beyond a mere cataloguing of opportunities for functionings, the characterization of capabilities can incorporate further variance of features depending on the purpose to which the relevant

[1] See Nussbaum *Creating Capabilities* (2011: x-xi).
[2] See Robeyns *Wellbeing, Freedom and Social Justice* (2017: 7).
[3] See Alkire in Comim, Qizilbash and Alkire (eds.) *The Capability Approach: Concepts, Measures and Applications* (2008: 27).

measure is lent. For example, practitioners might ask questions such as whether the characterization sufficiently captures the collective sense of well-being for a particular community, whether what it captures is useful for comparisons of well-being *across* populations, whether it accurately captures important senses of oppression and injustice, or whether it captures vulnerabilities such as higher risk of unemployment, exposure to COVID-19, or further forms of structural oppression. Though this variety of characterizations enhances the innovative nature of the capabilities approach, practitioners continue to debate which sets of features are salient with respect to which measurement purposes.

Though there is some variation to the concept, there continue to be two dominant characterizations of the capabilities within the approach: those of founding authors Amartya Sen and Martha Nussbaum. These two characterizations have significant overlap, yet, as we shall see, lend themselves more readily to different purposes and surrounding theories such they are often juxtaposed.[4] In this thesis I give an analysis of 'capabilitarian well-being,' a specialized term I use to refer to the overlap between Nussbaum and Sen's characterizations and related commitments for measuring capabilities at both the individual and population level. In the former case, there are measures of capabilitarian well-being that can look to what kinds and levels of functionings are available to some individual, while in the latter case there are measures that are used to compare opportunities for different kinds of functionings across groups of individuals. I offer reconstructed characterizations of both authors' concepts to explain how their views overlap regarding key features of capabilitarian well-being and related commitments for measuring it. I use these features and commitments to analyze various measures of capabilitarian well-being, and indicate a few areas of caution, particularly for measures that can always provide total orderings of different kinds - often called 'dimensions' - of capabilities.

I start the analysis by explicating the conceptual roots of the capabilities approach: the ethical and political writings of Aristotle, who was one of the first Western philosophers of social science. I unpack Aristotle's claim that people who exhibit what he calls 'practical wisdom' - meaning they have had the

[4] See, for example, Robeyns (2017) and Claassen (2018).

right kinds of experiences that enable them to make prudent judgments - are best able to recognize and evaluate the kinds of doings and beings that constitute what Aristotle calls a 'flourishing' life of *eudaimonia*. I trace these Aristotelian ideas through both Sen and Nussbaum's characterizations of capabilitarian well-being. Even though both authors distance their characterizations from particular elements of Aristotle's philosophical view, they both acknowledge that the Aristotelian view substantially influences their characterizations of their concept.

Moving from the Aristotelian roots, I first give an exposition of Amartya Sen's work on capabilities, which continues to be heavily cited by practitioners working to measure well-being. For example, Sen's concept features in the creation of the United Nations Development Programme (UNDP) Human Development Index (HDI) and Sarkozy Commission Report cited above. Sen characterizes capabilities as *the extent of an individual's freedom, also called 'opportunities,' to achieve doings or beings they have reason to value*. I explicate Sen's pluralistic emphasis on *different kind*s of opportunities that are, in turn, recognized according to different kinds of values, and show that this Sennian pluralism sanctions certain procedures for evaluating different kinds of capabilities. As part of this analysis, I also highlight the mathematical and logical structures that Sen offers as formal representations of capabilities as he characterizes them, and explain how practitioners who cite Sen's work often take liberties with those structures such that they may no longer measure Sen's concept.

From Sen's concept, I distil two key commitments for measuring capabilitarian well-being, both of which overlap with Nussbaum's concept yet tend to be incautiously downplayed by practitioners working within the approach. First, as mentioned above, Sen explains that capabilities are 'value-laden' such that the "approach is concerned *primarily* with the identification of value-objects, and sees the evaluative space in terms of functionings and capabilities to function" (Sen 1992: 43). Because the approach recognizes capabilities as functionings that people have reason to *value*, practitioners must implement the right kind of measurement practices that allow stakeholders who may be affected by

associated policies to express, reasonably articulate, and give influence to this value-ladenness of capabilities.

Second, Sen describes a key feature of capabilitarian well-being using generally synonymous terms including 'heterogeneous,' 'irreducibly diverse,' or even 'non-commensurable.' Each of these terms indicates that capabilities are constituted by different kinds of doings and beings that, as mentioned above, are valued in diverse kinds of ways. Because of this heterogeneity, different kinds of capabilities are heavily resistant to being rendered into total orderings, even though different kinds of capabilities can often be compared, say in order to make a hard choice between achieving one or another valuable functioning (Sen 2009: 240).[5] Sen allows, as a practical matter, that because limited comparisons of different kinds of capabilities are possible, a measure may, in rare cases, yield total orderings of capabilitarian well-being may (ibid: 241).[6] Yet measures that can *always* provide total orderings of capabilitarian well-being, especially those that do so by trading off different kinds of capabilities such that an increase of any one can be substituted for a decrease of another, are inconsistent with Sen's characterization and representations of capabilities as heterogeneous. I argue, following Sen (1992, 1999, 2009), that measures that keep different kinds of capabilities disaggregated or that only provide partial rankings (or both) are more likely to yield results that track Sen's and, as we shall see, Nussbaum's concept of capabilitarian well-being.[7]

Nussbaum characterizes capabilities such that they are *the situationally-sensitive list of people's basic abilities to freely achieve opportunities for valuable doings or beings*. Though different in a few substantive ways, Nussbaum's characterization of capabilitarian well-being largely mirrors that of Sen's, especially the key commitments introduced above. Capabilities are value-laden on Nussbaum's view just

[5] See, for example, Comim et al (2008: 75, 184, 294) or Robeyns (2017: 207). See also Sen in *Development as Freedom* in which Sen opposes the idea of "completeness of the evaluative ranking" to the term 'partial ranking' (Sen 1999: 82).
[6] Total Orderings are also often called 'complete rankings.' Rare cases refer to those of what are called 'set dominance.' These terms are clarified in chapter two.
[7] See Sen *Inequality Reexamined*, in which Sen argues that "if an underlying idea has an essential ambiguity, a *precise* formulation of that idea must try to *capture* that ambiguity rather than lose it" (1992: 48 – 49, original emphasis). Sen makes similar arguments in *Development as Freedom* (1999: 75 – 82) and *The Idea of Justice* in which Sen argues that "capability is entirely consistent with reliance on partial rankings and limited agreements … The main task is to get things right on the comparative judgments that can be reached through personal opinion and public reasoning, rather than to feel compelled to opine on every possible comparison that could be considered" (2009: 241 – 243). I return to these points, among others relevant for both Sen and Nussbaum's characterizations, in subsequent chapters.

as they are for Sen, though Nussbaum takes a more philosophically-grounded stance on what it means for people to recognize and measure the value-ladenness of different kinds of capabilities. Nussbaum explains that capabilities are "evaluative and ethical right from the start", such that the approach "asks, among the many things that human beings might develop the capacity to do, which ones are the really valuable ones, which are the ones that *a minimally just society* will endeavor to nurture and support" (Nussbaum 2011: 28, *emphasis added*). Following Aristotle's views on *eudaimonia*, Nussbaum characterizes the value-ladenness of capabilities not only according to a broader notion of how well individual people's lives are going, but also according to a narrower political emphasis on the provision of a just social minimum at the level of a society.[8]

Secondly, Nussbaum mirrors Sen in claiming that capabilities are 'irreducibly heterogeneous' such that they are each "distinctive, and all need to be secured and protected in distinctive ways" (ibid: 35). Nussbaum's position on this 'qualitative distinctiveness' of capabilities is grounded on Aristotle's philosophical view of a good life as composed of different kinds of goods - also called virtues -, each of which can be recognized according to the excellent achievement of distinct sets of functionings. According to this Aristotelian view, the excellence of one or another kind good is *particular* to some specific set of experiences, even if the relevant virtue can be spoken of generally. For example, friendship is a general kind of virtue, yet affiliation can be recognized across a wide variety of particular experiences. Nussbaum's concept of capabilitarian well-being follows this Aristotelian focus on the particular; I show that though Nussbaum characterizes capabilities as a list of central features that allegedly applies for all human beings, the list remains sensitive to more particular combinations of context and culture. In the same way that what it means to achieve friendship varies across cultures, so

[8] See chapters one and four regarding connections between Nussbaum's characterization and the Aristotelian philosophical background.

Nussbaum's list is sensitive to variances of distinct kinds of values that influence how particular people or communities might define opportunities for valuable doings or beings.[9]

A concomitant of this focus on the particular is that, similar to Sen, Nussbaum argues against measurement processes that evaluate distinct kinds of capabilities in terms of one another such that a 'tragic lack' in one can be compensated by substitution with some excess of another. For example, an excess of affiliation can certainly be a balm for some tragic lack of physical health like a chronic ailment or terminal disease, but the tragedy remains in that there is no real cure. Even if given an excess of one or multiple kinds of capabilities as substitutes, a person with a tragic lack like a chronic ailment or a lack of education would continue to be influenced by that specific kind of deprivation. Nussbaum recognizes that though different kinds of capabilities can, as a practical matter, be compared, there are significant moral reasons not to engage in such comparisons, especially comparisons that suggest they are readily intersubstitutable. For example, in her article on "Costs of Tragedy," Nussbaum claims that cost-benefit analysis methods used to assign comparative weights to different kinds of capabilities for the purpose of trading them off against one another do not "encourage us to ask which options involve denying some citizens a central entitlement, one they have a right to demand, and which options do not" (Nussbaum 2000c: 1032). Total orderings of capabilitarian well-being tend to render these kinds of important questions invisible, such that the total ordering fails to capture important features of the target concept.

As mentioned above, measures that are faithful to Nussbaum and Sen's concept of capabilitarian well-being can generally offer only partial orderings. This general claim is supported by a wide assortment of arguments and demonstrations - as I show in this thesis -, yet the point is intuitively accessible. Consider that, at the individual level, if someone has the opportunity to achieve a fulfilling vocational goal at the expense of significant levels of social connectedness and physical health, while someone else can achieve a middling level of all three but not significant levels of any, there may be no

[9] Nussbaum calls this 'multiple realizability' – see chapter three. I contrast 'multiple realizability' with claims that Nussbaum's list is 'perfectionistic' where the term is taken to indicate Nussbaum's list imposes a universalist interpretation of what constitutes a good life onto different cultures and concomitant sets of values.

fact of the matter regarding which has more capabilitarian well-being.[10] Similar problems apply at the population level: it can be challenging to measure which populations have more capabilitarian well-being when populations not only can attain different levels of achievement, but may also have different sets of doings and beings available to them. In such cases, there may be no fact of the matter in terms of total orderings of Sen and Nussbaum's capabilitarian well-being.

And yet, as mentioned above, practitioners working within the approach often aim to develop measures that can always yield total orderings of capabilitarian well-being across individuals or across populations. They generally manage to do so, with a few interesting exceptions, by assigning weights to and trading off levels of different indicators, each taken to represent different kinds of capabilities.[11] I argue that many such practices produce a measure that does not correspond to core features of capabilities as characterized by Sen and Nussbaum, and so the measure may be '*off-target*,' meaning the procedures will not provide assignments that track the capabilitarian well-being of those individuals or populations. By contrast, some measures of capabilitarian well-being - for example the Alkire and Foster Capability Deprivation Headcount Measure - represent different kinds of indicators separately, in tabular or matrix form, such that they avoid tradeoffs among indicators and are more likely to remain on-target with respect to key commitments of Sen and Nussbaum's characterization.[12]

I develop my argument according to the '*representational theory of measurement*' (RTM), which originates in the work of Patrick Suppes and Duncan Luce. I give a more detailed presentation of RTM and the relevant mode of analysis in the beginning of chapter four, and for now offer a brief introduction. According to RTM, there are three ingredients of a measure: its characterization, its representation, and the relevant set of operational procedures. The characterization is the description of the features taken to hold for the measured concept, the representation is the formal apparatus that gives mathematical or

[10] There are some cases in which orderings are not partial; the point is that orderings are often partial or incomplete. I consider differences between full and partial orderings in chapter two.
[11] Interesting exceptions include, for example, Dasgupta and Weale's (1992) Borda Count Rule, and also fuzzy set theory (see Martinetti and Roche 2009). Yet most multi-dimensional measures of capabilities that are currently in practice (the bulk of which cite Sen or Nussbaum or both) treat different kinds of capabilities as if they can *always* be weighted and traded off against one another in order to produce a total ordering.
[12] I discuss these kinds of measures in chapter five, including the Alkire and Foster measure in chapter five section three.

diagrammatic expression for the characterized features, and the operational procedures are the set of practices used to generate outputs for the features measured. For example, in the case of a mercury thermometer the target phenomenon, temperature, is characterized according to a concept of 'absolute temperature' in terms of kinetic energy such that certain amounts of heat energy correspond to proportional expansions of mercury in a thermometer tube. Here, the measurement system provides an output as degree markings placed at standardized intervals along the thermometer tube. In the case of the thermometer, degree markings represent a cardinal scale in which magnitudes of intervals between units on the scale can be meaningfully compared to one another, for example that the difference between 122 degrees and 116 degrees is twice the difference between 13 degrees and 10 degrees. The operational procedures for measuring temperature according to a thermometer include those that assure that the thermometer works correctly, for example that it has been applied in a context where there is a temperature differential that can be reliably measured, that the thermometer has been designed and implemented appropriately, etc.

According to RTM, measures are '*incoherent*' when core ingredients of the measure do not cohere with each other. In such cases, the measure may no longer track one or many features characterized for the intended target. As a basic example, consider that if the 100 °F mark has not been set according to a well-justified set of procedures on a mercury thermometer, there is good reason to doubt that the thermometer measures temperature - as represented on a cardinal scale - in a manner that is on-target with the concept. This is a well-discussed example, as Fahrenheit famously intended to mark the temperature of the human body at 100 degrees. However, since the Fahrenheit discrepancy is fairly minor (1.4 degrees) such that one could argue that, as a practical matter, the incoherence is relatively trivial, let's also consider the challenge faced by Royal Society chemists who attempted to set 100 degrees Celsius according to the boiling point of water. The boiling point of water can vary - in some cases quite significantly - according to the type of container, the presence of oxygen, the purity of the

water, the elevation level, etc.[13] The wrong kind of operational procedure, say the boiling of water in the wrong type of vessel or at the wrong elevation, may not be coherent with respect to how the target phenomenon has been represented operationalized, and so the measure may be off-target.

I prefaced this issue of incoherence for measures of capabilitarian well-being just above when I argued that Sen and Nussbaum characterize and represent capabilitarian well-being according to certain key commitments, yet some capabilities measures include representations or operational procedures that are inconsistent with those commitments. Writing from an economics background that includes his significant contributions to Social Choice Theory, Sen uses certain mathematical and logical structures to formally represent features characterized of his concept, yet some practitioners cite Sen but use other representational structures that do not cohere with Sen's representation and characterization in their measures. Similar arguments for incoherence apply, mutatis mutandis, regarding liberties that some practitioners take with respect to Nussbaum's characterization, representation, and recommended operational procedures.

Using this kind of coherentist analysis from RTM, I consider both the Improved Human Development Index (HDI) developed by Jaya Krishnakumar and the measure of Capability Deprivation Depth developed by Sabina Alkire and James Foster.[14] Among other incoherences, both measures trade off different kinds of indicators against one another such that they can always provide total orderings, and so both measures are in danger of being off-target with respect to Sen and Nussbaum's characterization of capabilitarian well-being.[15] I offer both of these analyses as support for a more general argument against the class of capabilitarian well-being measures that include procedures for trading off different kinds of capabilities against one another in order to always produce total orderings. I argue that by including such

[13] See Hasok Chang in *Inventing Temperature*, which also would have us consider the challenge of expanding the Celsius scale into unknown realms of temperature that could not be easily measured by a thermometer. See also Bridgman (1927). I return to this discussion in chapter four section two.

[14] See chapter five.

[15] For a list of similar kinds of indices, see Roché and Chiappero-Martinetti's "Operationalization of the Capability Approach, from Theory to Practice: A Review of Techniques and Empirical Applications" in which the authors list over twenty such measures that follow the capabilities approach (2009: Table 2). Or, similarly, see Decancq and Lugo (2010) whose "Weights in Multidimensional Indices of Well-Being: An Overview" provides a similar list (2010: Table 3) that includes, among those mentioned above, the 'Proportional Deprivation Index,' the 'Life Conditions Index,' the 'Index of Economic Well-being' etc.

procedures, practitioners act as if representations and procedures that tend to be used in decision-making processes - in this case weighted tradeoffs among different kinds of options - also work well-enough to provide on-target measures of capabilitarian well-being as characterized by Sen and Nussbaum. However, methods for decision-making and methods for measuring capabilitarian well-being have different purposes and different domains of use.[16] This point connects to another important nuance of analysis according to RTM, which is that we must consider the purpose of a measure when we question the coherence of its ingredients.

As Anna Alexandrova argues in A Philosophy for the Science of Well-Being (2017), well-being measures - of which measures of capabilitarian well-being are a species - admit of a wide variety of purposes and contexts. I allow that some combinations of context and purpose might be sufficient to justify the use of incoherent measures of capabilitarian well-being, even measures that can always provide total orderings. I recognize, following practical discussion from Elizabeth Anderson, that such measures might be useful as rough 'performance standards' used to rank something like capabilitarian 'well-being' of different populations - much in the same way that Olympic figure skating competitions set performance standards for ranking different competitors in terms of figure skating ability. Many capabilities indices do seem to be implemented according to this kind of standard-setting purpose. The problem is that they are then lent to purposes in which the output yielded by the index is treated as something other than a rough ranking.

Two points must be made here regarding the practical analysis that I give along with the coherentist analysis of measures of Nussbaum and Sen's capabilitarian well-being. The first is that a measure of what I'm calling 'capabilitarian well-being' may in practice or in fact be a measure of some other characterization of 'capabilities,' even when practitioners developing the measure cite the work of Sen and Nussbaum. However, in such cases, practitioners would need to have provided deep conceptual work to clarify their characterization of the concept 'capabilities' such that the alternate characterization

[16] I explain these points particularly in terms of 'deliberation' and 'phronesis,' both of which are discussed throughout the thesis. See chapter two for an introduction to Aristotle's notion of phronesis.

justifiably differs from that given Nussbaum and Sen, and such that it affords with the relevant representations and operational procedures. Most authors working within the approach simply refer to some concept like 'multidimensional well-being' or 'capabilities' and provide little or no characterization work other than references to foundational authors like Nussbaum and Sen. If practitioners developing some measure reference the characterization work of the founding authors and otherwise provide no significant characterization work of their own, then it is appropriate to assume the measure is intended to follow Sen and Nussbaum's concept of capabilitarian well-being. I assume for the purpose of this thesis that many contemporary measures of 'multi-dimensional well-being,' 'capabilities,' or even 'human development,' including at least the exemplars considered in chapter five, attempt to measure a concept that generally matches what Sen and Nussbaum have characterized and what I refer to as 'capabilitarian well-being,' at least in terms of the commitments presented in this introduction and defended in the following chapters.

Secondly, practical analysis of a measure should allow that though a measure is incoherent such that it may be in danger of being off-target with respect to the intended characterization, the measure still might provide an indication that remains roughly useful with respect to some purposes. I argue that the indexical 'performance standards' mentioned above carry a similar function; though they are clearly incoherent with respect to core commitments for measuring capabilitarian well-being, such measures may yield results that approximate capabilitarian well-being closely enough for some purposes, for example as a means of furthering an international political conversation regarding how generally to improve people's lives.[17] Such standards may not be useful for more particular purposes, say for an in-depth, cross-national assessment of capabilitarian well-being among three or more nations. I argue that regardless of the purpose, it is important that the incoherence of the measure be made transparent for practitioners and

[17] See, for clarification, Fukuda Parr (2003: 305): "[Mahbub ul Haq] was convinced that a simple combined measure of human development was essential for convincing the public, academics, and policy-makers that they should evaluate development by advances in human well-being and not only by advances in the economy. Although Sen initially opposed this idea, he went on to help Haq develop the Human Development Index (HDI), a composite index of achievements in human development. Sen was concerned by the difficulties of capturing the full complexity of human capabilities in a single index. But he was persuaded by Haq's insistence that only a single number could shift the attention of policy-makers from material output to human well-being as a real measure of progress (United Nations Development Programme 1999)."

stakeholders. Moreover, if there are significant harms associated with the possibility that the measure may be off-target, then further investigation of the incoherence is warranted.

I clarify these points regarding purpose and warrant later in the thesis, but to introduce the argument, let's consider the Multidimensional Poverty Index (MPI), which provides total orderings of 'multidimensional poverty' of different populations. So long as the concept 'multidimensional poverty' is taken to be a version of Nussbaum and Sen's capabilitarian well-being as defined in this thesis, the MPI is not true to that concept inasmuch as it includes procedures for assigning weights to and trading off different kinds of indicators such that total orderings are always provided. Despite this limitation, the MPI is certainly useful for identifying which people in a population are the 'haves' and which are the 'have-nots,' especially when we find, following Wolff and DeShalit (2013) that deprivations of capability cluster such that those who experience multidimensional poverty tend to experience it across many different dimensions of capabilities, and, similarly, that lack of capability deprivation also tends to cluster across dimensions. In such cases, even if the measure is incoherent such that it may be off-target, it is likely to provide a picture that is sufficiently useful for policy-makers and other practitioners working to identify and design policies to alleviate different kinds of poverty. Still, even in this case the incoherence may warrant further investigation depending on how significant are the possible harms associated with the measure mis-identifying cases of capability deprivation, etc.

Later in the thesis, I offer a few further practical guidelines for the development and implementation of contemporary measures of Nussbaum and Sen's capabilitarian well-being. A central guideline is that commitments to value-ladenness must be transparently recognized not only by practitioners but also by groups of deliberators if and when they engage in the evaluative processes used to measure capabilities. I recognize, following Ingrid Robeyns (2017) and Morten Byskov (2018), that there might be a limited theoretical justification for treating capabilities as 'value-free' or 'value-neutral,' though characterizing capabilities as such is no longer faithful to Sen and Nussbaum's concept of capabilitarian well-being. I argue that practitioners should treat capabilitarian well-being as value-laden

at all stages of the measurement process, especially in the context of Participatory Learning and Action (PLA) practices that are often used to determine which policies create more capabilities for developing communities.[18] Regarding the commitment to avoiding tradeoffs, I allow that there is some justification for treating different kinds of capabilities as roughly comparable when they are in excess above a context-sensitive social minimum. Yet I argue, following Nussbaum's characterization, that tradeoffs of different kinds of indicators below the relevant social minima are unlikely to carry significant meaning given that deprivations of such indicators are, in most cases, not inter-substitutable. I draw these considerations into a summary of practical recommendations for measuring capabilitarian well-being in at the end of chapter five.

The thesis progresses as follows: In chapter one I outline portions of Aristotle's political and ethical philosophy that are relevant to how practitioners should conceptualize and attempt to measure capabilitarian well-being. In his Nicomachean Ethics, Aristotle claims that well-being can be understood according to many different kinds of goods, some of which are Aristotelian 'virtues' (like generosity or courage). These different kinds of goods can best be recognized by people who have had the right kinds of experiences. Aristotle also posits an experience-based, practical wisdom – in Greek called *'phronêsis'* that prudent people use when they recognize such goods in particular situations, for example when they evaluate different life choices. When making prudent decisions, agents of *phronêsis* recognize distinct kinds of goods - including different kinds of virtues - in order to make their decisions. I return to Aristotle's concept of *phronêsis* in chapter three.

In the Nicomachean Ethics, Aristotle also claims that different kinds of goods are not actually measured by their economic exchange values because different kinds of goods are non-commensurable - in Greek '*asummetria*' - such that they are resistant to being rendered in terms of one another. One implication of *asummetria* is that though different kinds of economic goods can be traded off against one another in terms of a common currency, the representation of how such goods differ in value according to

[18] By 'measurement process' I include, inter alia, processes that select which kinds of doings and beings are relevant for measuring capabilitarian well-being and also which indicators best work as proxies in the given context. See chapter four for further discussion.

that common currency no longer tracks the entire set of values that are relevant to those different kinds of economic goods.[19] From the discussion of *asummetria*, *phronêsis*, and other elements of Aristotle's philosophy, I defend two Aristotelian commitments to measuring well-being that preface Sen and Nussbaum's characterization of capabilitarian well-being. Those two commitments are (i) that different kinds of goods are value-laden, and (ii) that inasmuch as we seek to capture the different kinds of values that themselves constitute those goods, we should avoid trading off different kinds of goods in terms of one another, as if they are intersubstitutable.

In the second chapter I show that Sen's characterization of capabilitarian well-being includes commitments both to the value-ladenness and the heterogeneity of capabilities. I offer an exposition of Sen's concept, emphasizing his commitment to pluralistic deliberation as a process of evaluation that should feature in any attempt to measure capabilitarian well-being. I also analyze Sen's representation of capabilities as sets of co-possible sets of functionings that a person values (positively or negatively) and that they have the freedom or opportunity to attain, given their resources and life circumstances. Advocates of Sen's approach often refer to these as 'the set of lives worth living' available to the person.[20] I explain that Sen's representation of capabilitarian well-being as sets of co-possible sets does not cohere with the 'Vector View' representation that contemporary authors often seem to attribute to Sen. A 'Vector View' representation represents capabilities according to a set of indicators, a structure that mirrors the kind of representation often used for cost-benefit analysis evaluation in decision-making processes. Representing the set of indicators as a 'vector' elides important ways in which sets of co-possible sets of doings and beings can be evaluated holistically against one another.

[19] Consider Cass Sunstein's (1994) argument that "different kinds of valuation cannot without significant loss be reduced to a single 'superconcept'" (Sunstein 1994: 784), for example a common 'currency' like dollars or, in the context of measuring well-being, units of utility or pleasure, etc. Sunstein argues that the act of trading different kinds of goods, themselves featuring different kinds of values or valuation, along a single metric can 'do violence' to our considered judgements between or among those goods. For example, there are market limitations, including laws, that prevent the economic valuation of children or kidneys, even when the market can, of course, provide a metric by which to compare children, kidneys, and economic staples like gasoline or cell phones, etc. (ibid: 796). Sunstein's example helps us to see the import of Aristotle's argument, and also bolster's Nussbaum's more direct claim that we should avoid, whenever possible, tragic tradeoffs among different kinds of capabilities, which are themselves qualitatively distinct. See chapter three for further discussion.

[20] See, for example, Lawrence Hamilton's use of the term in *Amartya Sen* (2019: 150), or similar use by Noam Peleg in "Reconceptualising the Child's Right to Development: Children and the Capability Approach" (2013: 539), or by Runhardt and Cartwright "Measurement" (2014) in Cartwright and Montuschi, eds. *Philosophy of Social Science: An Introduction*, etc.

Using a vector of indicators also lends itself more readily to trading off different kinds of vector elements to produce total orderings. I explain how Sen (1980b) in "Plural Utility" gives a limited discussion of the Vector View as an alternative to single grade utilitarian metrics of well-being. Sen recognizes that the Vector View of utilitarian well-being can be aggregated by trading off different kinds of indicators according to what Sen calls a 'secondary view of utility' in which the common currency for trading off indicators is some interpersonal standard of intensity of pleasure or desire (Sen 1980b). However, Sen questions the interpersonal comparability of such a standard, and he argues in favor of the disaggregated picture. I show that, similarly, Sen argues that evaluative orderings of capabilities as sets of co-possible sets will generally be incomplete, or partial.

I argue that practitioners who cite Sen's work yet proceed to trade off different kinds of indicators to provide total orderings do not actually follow Sen's characterization and representation of capabilities.[21] If anything, they seem to misapply the Vector View - itself a part of Sen's critique of utilitarianism - in a capabilitarian context. It is possible, however, that practitioners who use such systems do so because the vector measure - even an incoherent, single grade total ordering constructed from it - is still roughly useful with respect to the relevant purpose, and may serve that purpose better than other vector or single-grade measures that don't employ any of the capabilitarian ideas about what welfare consists in. I claim that such practitioners may misinterpret Sen's cautious endorsement of complete choice orderings in practical decision-making contexts, believing, perhaps, that the same kind of total ordering justifiably applies in the context of trying to accurately measure capabilities. With such practitioners in mind, I consider a class of measures that appear to use one or another feature of capabilities - for example, the level of 'substantive freedom' or 'navigational agency' - as a common currency for the purpose of providing a total ordering of items as a rough, if incoherent, estimate. I argue that only limited,

[21] Consider, for example, the work done in support of the Human Development Index and the OECD Better Life Index, both of which cite Sen and, so it would seem, are developed on the basis of at least Sen's if not Sen's and Nussbaum's characterization work. Consider also, in the context of this thesis, Alkire and Foster's Multidimensional Poverty Index, which includes numerous references to and citations of Sen's work. See also Krishnakumar (2015), which is considered along with the MPI in chapter five. See also, as a partial sampling, Decancq (2013) and Decancq and Lugo (2010). See also Nguefack-Tsague (2011), Peruzzi (2014), Quizilbash (1987), Pattanaik and Xu (1990), Kuklys (2005), Comim (2001), Chowdhury and Squire (2006) as listed in the bibliography of this thesis.

particularly urgent contexts provide sufficient warrant for the practical application of members of the class.

In chapter three I argue that Nussbaum's objective list characterization of capabilitarian well-being also includes the commitments to value-ladenness and the 'qualitative distinctiveness' of different kinds of capabilities. I present Nussbaum's objective list characterization and representation, explaining that the list admits of more open-endedness than is generally recognized by those who call the list 'perfectionistic.' In fact, the number of different 'central features' within the list is not set, and central features can be understood differently in different contexts – what Nussbaum calls 'multiple realizability.' I argue the plasticity of the list follows from Nussbaum's commitment to Aristotelian *phronesis*, or practical wisdom. Nussbaum argues, following Aristotle, that the agent of *phronêsis* exhibits a well-honed ability to recognize different kinds of deep expressions of value in context – an ability called '*prohairesis.'* The ability to recognize different kinds of values supports, in the context of measuring well-being, the commitment to measuring different kinds of capabilities separately. Following Nussbaum's defense of *phronêsis* as *prohairesis* - what Nussbaum also calls 'desiderative deliberation' - I argue that central features within the objective list representation are not characterized such that they can be traded off against one another to provide total orderings.[22]

However, as in chapter two, I recognize one can argue that measures that are not entirely true to capabilitarian wellbeing as characterized by Nussbaum may still provide an estimate that is sufficiently useful for some purposes and contexts. I respond, following Nussbaum, that the wide application of the reductive practice of trading off different kinds of capabilities among one another can lead practitioners to misinterpret the tragic nature of difficult policy decisions. Especially in contexts in which deliberating individuals or groups must choose among different kinds of capabilities, practices that reductively represent different kinds of capabilities in terms of one another can render invisible important expressions that could otherwise significantly influence public deliberation. I bolster this argument with a survey of

[22] See Nussbaum 1986a: 308.

'pragmatic inquiry' as defended by Hilary Putnam, which I take to be indicative of the kind of measurement process defended by both Sen and Nussbaum in that such practices recognize objectively-observable 'facts' about values. In that context, I also consider a set of pragmatic arguments from Elizabeth Anderson that justify practices that trade off different kinds of well-being values to create total orderings, but only relative to the purpose of setting what she calls a 'performance standard.' Though it is incoherent with respect to Nussbaum and Sen's concept of capabilitarian well-being, a total ordering performance standard may be useful for a limited combination of contexts and purposes.

In chapter four I provide a set of arguments that indicate how certain representations or operational procedures do not cohere with the characterization of capabilitarian well-being according to Sen and Nussbaum. I also outline the representational theory of measurement (RTM) that provides the basis for these arguments. In recognition of the commitments to value-ladenness and the qualitative distinctiveness of different kinds of capabilities, I give extended examples of how incoherence with respect to either commitment can manifest. The first incoherence, *Value-Omission Incoherence*, occurs when capabilities measures treat - at any point in the measurement process - capabilities as value-free or value-neutral rather than value-laden. I argue, following Bernard Williams, that such treatment is incoherent with respect to the characterization of capabilitarian well-being because even though the doings and beings themselves can be abstractly represented as value-free, their selection is necessarily a value-laden exercise. I demonstrate the implications of this kind of incoherence by discussing Participatory Learning and Action practices in which groups of local community members deliberate over relative contributions to well-being of different policy initiatives. Without the right kinds of practices that foster transparent recognition of value-ladenness, deliberators might simply parrot the wishes of community organizers or other non-local advocacy groups. If policies are to best serve the interests of stakeholders, the values expressed by relevant stakeholders must be recognized and given influence within deliberation from the start.

The second incoherence, *Numerical Scale Incoherence*, occurs when capabilities measures that follow Sen or Nussbaum's characterization (or both) treat different kinds indicators or capabilitarian well-being values as if they can be traded off against one another in order to provide total orderings. I show, following the recommendation of Decancq and Lugo (2010), that a deeper look into such measures inevitably exposes tradeoffs among different intervals of indicators, that, when made explicit, are at least unintuitive and may be unacceptable to relevant stakeholders.

In the final chapter I analyze two contemporary capabilities measures, the Alkire and Foster Capability Deprivation Measure and the Krishnakumar 'improved' Human Development Index. I consider these two measures as exemplars for a number of contemporary capabilities measures that similarly use tradeoffs among indicators to provide total orderings. In the case of Krishnakumar's Human Development Index, which follows more directly from Sen's work, I critique the use of a statistical practice, factor analysis, to set weights used to trade off different 'dimensions' of indicators. In the case of the Alkire and Foster measure, I argue that there are good justifications for certain forms of combining indicators into a single measure, for example through the Alkire and Foster Method headcounts of the number of capability deprivations faced by individuals within a population. However, the Alkire and Foster measure can be used to provide measures of capability 'deprivation depth' derived by trading off intervals among different kinds of indicators. I argue this practice is incoherent with respect to the characterization of capabilitarian well-being.

I conclude the dissertation with a brief summary of what I think has been accomplished.

Chapter One:

The Aristotelian Background to the Characterization of Capabilities:
Commitments to Value-Ladenness
and against Total Orderings of Different Kinds of Well-beings

1.1 Introduction

In this chapter I outline the Aristotelian background to Amartya Sen and Martha Nussbaum's foundational characterizations of capabilitarian well-being. Aristotle's philosophical view is relevant for measuring well-being because it offers a practical, well-defended framework for recognizing and attaining a 'flourishing' life of *eudaimonia*- one that is healthy, worthwhile, and conducive to happiness. By 'happiness' I refer to different kinds of deeply meaningful ways one can appreciate one's experience, in a manner that "supervenes on activity 'like the bloom on the cheek of youth'" (Nussbaum 2008: 585). That there are different kinds of happiness is a recurring theme among theories for measuring well-being. For example, in chapter two I show that Sen (1980b) follows philosophers such as John Stuart Mill, Karl Marx, and, yes, Aristotle, in claiming that people experience different kinds of pleasures. The happiness we may experience when eating a great meal is not the same kind that we may experience when making new friends, gaining wisdom, finishing important personal projects, and so on. Aristotle similarly claims that well-being, as a flourishing existence, is constituted by distinct kinds of goods – what Aristotle calls 'virtues' and 'virtuous activities.' As we shall see, the claim that virtues are of distinct kinds has important implications for how we can attempt measure a life of *eudaimonia*.[1]

According to Aristotle, well-being is measured as 'value-laden,' given that it is done so according to how well people can recognize different kinds of virtues and virtuous activities. Given the more practical orientation of this thesis, I use philosopher Ruth Chang's 'ordinary' notion of values, where values "include any evaluative abstracta, including obligations, rights, duties, utility, excellences and so on, and are not limited to evaluative criteria, like pleasure, that can be aggregated by a cardinal unit of

[1] According to the representational theory of measurement introduced in the previous section, by 'measure' I refer to a system of characterization, representation, and operational procedures – what I also call a 'measurement system' - that justifiably captures the features characterized of the target phenomenon. See chapter four section two.

measure" (Chang 2015: 204).[2] Following Chang's ordinary notion, I specifically focus on how values are evaluative expressions that influence the development of abstract guidelines for individual or group action. In this and the next two chapters I further clarify some of the repercussions of Sen and Nussbaum's claims that values can be expressed as deep emotions or coherent sets of reasons, or as complex combinations of both. We shall see that Sen provides a wide definition of the kind of 'reasonable' rather than formally 'rational' expressions that people may give when deliberating about well-being, and Nussbaum defends prudent peoples' abilities to recognize different kinds of 'values' that are relevant when such people deliberate regarding the well-being of different life choices, policy options, etc.[3]

I add here, following Chang (above) and also Anna Alexandrova (2017), that values tend to be action-guiding, both for individuals and for groups.[4] For example, an individual or population might evaluate some state of being according to the kinds of well-being that state is taken to promote. The evaluation, whether positive or negative, may then influence the justifications given for policies taken to further or limit those kinds of well-being. I grant that values can be given more specific definitions and roles when they are understood according to one or another well-defended theory of ethics. However, in the practical world of policy implementation, the set of expressions that can be understood as providing abstract guidelines for action includes not only ethical and legal theories but also a wide range of cultural and religious perspectives. Values are diverse, especially when expressed relative to practical purposes,

[2] I provide the following as clarification of 'cardinal' vs. 'ordinal': On ordinal scales, magnitudes of intervals between units do not give meaningful measurement information for the relevant property. Say, for example, that populations A, B, and C are measured along an ordinal scale of utility such that A has 10 utils, B has 20 utils, and C has 200 utils. If the scale is ordinal, we know that C > B > A as a ranking, but the difference in intervals does not actually tell us how much more utility there is for C versus B than for B versus A. Even though C and B differ by 190 utils and B and A differ by 10 utils, we still don't know whether people experience more pleasure, desire, or have more preference for C versus B than for B versus A; other than for creating ordering relationships of greater or less than, the intervals between different numbers on the scale are meaningless in terms of the measurement information they convey.
For cardinal scales, intervals between units do carry meaning so that, returning to the example above, we know that the difference in experienced pleasure, desire, or preference for B versus A is nineteen times less than for C versus B. Cardinal scales can furthermore be either ratio or interval scales. The difference between the two types of cardinal scales can be demonstrated by scales of temperature: Celsius and Fahrenheit scales lack a zero value (the zero is arbitrarily set) and so only interval comparisons are possible. Absolute Kelvin, by contrast, has a true zero temperature, allowing for ratio comparisons of different units along that scale. On the Kelvin scale, 50 degrees Kelvin have five times the magnitude of temperature as 10 degrees Kelvin, while on the Celsius or Fahrenheit scale no such claim can be made because there is not a true zero to either scale.
[3] From Nussbaum's (1986a) *Fragility of Goodness*. See chapter three section three.
[4] See Alexandrova (2017) A Philosophy for the Science of Well-Being pp. xv – xxx. This point foregrounds the discussion of 'practical inquiry' in chapter three.

and so I use Chang and Alexandrova's characterization because it does not deny the practical influence of some values in exchange for those justified by one or another particular ethical or legal theory. I return to this theme of diversity in the next two chapters, as it is strongly presupposed by the characterizations of both Sen and Nussbaum. For now I note, as an introduction to the concept, that Aristotelian philosophy includes the commitment to recognizing important doings and beings in terms of different kinds of Aristotelian virtues, which are themselves a form of value according to Chang's ordinary sense of the term.

A second important point from Aristotelian philosophy is that the well-being values of different kinds of economic objects are not, in general, fully captured when such objects are represented as if they can all be traded off among one another according to a common currency of exchange. Different kinds of economic objects exhibit qualities that resist being rendered in terms of one another – a property Aristotle calls '*asummetria,*' often translated as 'non-commensurability.' This rendering issue is common enough in the scientific world as it relates to different kinds of physical qualities. For example, it generally makes no sense to say something like "x units of pH are equivalent to y units of temperature" or to defend some similar equivalence across different kinds of physical qualities, even though one can imagine contexts in which it might make sense to compare the pH of one object to the temperature of another, or, similarly, to treat different kinds of physical qualities as if they can be traded off among one another, particularly with respect to some specific purpose. Consider, for example, how the low pH of some substance might prove much more useful relative to some specific purpose than the ease with which another substance combusts to give off heat. Similar forms of comparison are common in the world of economics, even when the relevant qualities are significantly different such that they resist being rendered in terms of one another. Consider how easily one might accept, as an economic exchange, that x paintings are equivalent to y gallons of gasoline, even when it remains unclear how the artistic value of the former can be rendered in terms of the qualities of the latter.

In Aristotle's Economic Thought (1995), Scott Meikle discusses Aristotle's analysis of exchange value and highlights how *asummetria* of different kinds of values delimits actual 'measurement' of those values. By 'exchange value' I refer to how different goods are traded against one another in fixed ratios through the use of currency, and by 'measurement' or 'measure' I refer to how well those values are captured and reflected by the relevant metric. Aristotle explains that exchange value doesn't take into account the fact that different economic goods feature different kinds of well-being values. For example, laws of supply and demand that determine the cost of shoes relative to cell phones will generally not reflect how shoes are valued relative to health or mobility while cell phones are valued relative to how we communicate and maintain personal relationships. According to Meikle's Aristotle, though exchange value works sufficiently for capturing a narrow sense of economic value, it is at best a limited stand-in for relevant well-being values and is not a true measure of those values inasmuch as it fails to capture them.

If Aristotle is right, the recognition that different kinds of goods often exhibit *asummetria* of different qualities, especially their well-being values, can be taken to support the general claim that measures of well-being that always allow for a total ordering of items will often fail to capture all relevant values. Sen and Nussbaum both take it to do that. Consider that different kinds of goods can easily be put into total orderings in terms of their exchange value. Yet when such goods exhibit *asummetria* of the different kinds of qualities constitutive of or relevant to each, total orderings will generally fail to capture those further qualities, and so will not stand as faithful measures of those qualities. There are, however, entire literatures devoted to analyzing the wide variety of methods by which different states of well-being can be put into total orderings, and so I am not attempting a defense of this claim for measuring well-being here. The claim I defend is that measures of *capabilitarian well-being* that fail to treat well-being as value-laden or that always provide total orderings (or both) do not fit Sen and Nussbaum's Aristotelian commitments to the value-ladenness of the concept and to allowing that different kinds of well-being values are resistant to being captured in terms of total orderings.

We shall see in subsequent chapters that Sen follows Aristotle when he argues that capabilities are 'heterogeneous' - they are constituted by different kinds of doings and beings that are evaluated according to different kinds of values.[5] We shall also see how Nussbaum draws from Aristotle's ethical and political philosophy to defend the practical import both of treating different kinds of capabilities as distinct and avoiding the representation of different kinds of capabilities as if they can be traded off among one another, especially to provide the possibility of total orderings in all cases. Both authors' characterizations of capabilities are not, however, purely Aristotelian. Sen distances his own characterization from Aristotle's philosophical view of well-being, which he claims provides a narrowly "objectivist framework based on a particular reading of human nature" (Sen 1993: 47). Similarly, in her later writing Nussbaum characterizes well-being through the broader notion of 'human dignity' rather than the more restrictive Aristotelian *eudaimonia*.[6] Yet neither Sen's nor Nussbaum's distancing substantively weakens either author's reliance on commitments to value-ladenness and against measurement practices that always provide total orderings of capabilitarian well-being.

In this chapter I also introduce how those who make decisions according to Aristotelian practical wisdom – in Greek '*phronêsis'* – must treat different kinds of well-being goods – what I continue to call 'values' according to Chang's 'ordinary' sense of the term - as distinct, meaning such values are to be recognized separately. There is a quandary here regarding what it means, in practice, to treat different kinds of well-being values as distinct. Those who act according to *phronêsis* must, like all practical decision-makers, make choices among options that exhibit *asummetria*. According to one fairly popular method for choosing among such options, decision-makers can treat distinct values *as if* they can be traded off among one another, for example according to a common exchange value. Using such a method, practical decision-makers can create total orderings of options - even when such options exhibit *asummetria* of different kinds of relevant well-being values.

[5] See chapter two sections two and three.
[6] For an excellent summary of the difference between Nussbaum's earlier emphasis on the *eudaimonic* and her later emphasis on dignity, see Chad Kleist (2016) *Developing Capabilities: A Feminist Discourse Ethics Approach* pp.10-34.

Let's consider, as an illustrative example, a case in which a city council decides whether to build a school or a retail outlet in the same physical space. After conferring with representatives from the community, the council evaluates options by recognizing different kinds of values that can, in turn, inform abstract guidelines used for justifying either decision. Members of the council might, for example, recognize values related to educational development that can be abstractly represented as positive effects on the mental, social, or physical health of individuals in the community. They might also recognize values related to commercial development that can be abstractly represented as positive effects on the level of freedom and health due to increased wages and infrastructure. Though these values – what the Aristotelian would call different kinds of virtues or virtuous activities – feature '*asummetria'* in that they are of different kinds, at some point the council makes a decision among options. However, if the council forces total orderings, especially by using a procedure that allows all different kinds of capabilitarian well-being values to be traded off among one another, the question remains: does the ordering decided on really reflect an ordering of the capabilitarian well-being of the options, or is the measure 'off-target' such that there is good reason to suspect it is not measuring the intended concept?

There are two points of concern here. First, the philosophical point according to the representational theory of measurement: does the measure that always allows for a total ordering cohere with the characterization of Sen or Nussbaum? Second, the practical point: if the measure does not, toward what purposes and concomitant sets of operational practices can the representation be considered as even roughly-useful rather than dangerously off-target? In many cases in which researchers or decision-makers (or both) measure capabilitarian well-being as characterized by Sen or Nussbaum, both the philosophical and the practical concerns seem to be disregarded.[7] Given the recognition of incoherence with respect to the target concept, the relevant measurement system warrants further investigation as to whether its outputs are sufficient to the task for which it has been implemented and, in many cases, designed.

[7] The paradigmatic case for this thesis is the Human Development Index (and similar indices such as the Better Life Index), but also contemporary measurement systems such as the Alkire/Foster Capability Deprivation Matrix and the Krishnakumar improved Human Development Index, both of which were introduced in the introduction to this thesis and are further analyzed in chapter five.

Let's posit a formal clarification of the practical issue at hand: if a decision-making individual or group must choose among options, say options A, B, C, or D, one way to model the decision is by first setting up some way to assign units to the different kinds of qualities - call them $x_1, x_2, \ldots x_n$ in a n-quality case – shared among the options, second to give weights to the qualities, third to multiply the weights by the units of each quality and add up the resulting numbers for each option. This gives a final number that can be used to give a complete rankings of A – D, much in the same way that judges award scores to the performances of different athletes in gymnastics or figure-skating competitions.[8] As Sen (2009a) notes, similar measurement procedures that always provide total orderings of items are extensively used as representations, perhaps even measurement systems, in contexts of practical decision-making. Yet so long as the measure is intended to cohere with Sen and Nussbaum's characterization of capabilitarian well-being, such operational procedures do not coherently *measure* the capabilitarian well-being of A, B, C, and D in cases in which those options exhibit *asummetria* of any of qualities x_1 - x_n, similar to how pH and viscosity are of different kinds.

Still, there remains a practical challenge to the claim that such measures are not coherent with respect to Sen and Nussbaum's characterization and may be off-target. Many practitioners operating within the capabilities approach simply give a shoulder-shrugging response when they discuss (if they even do) whether measures that always provide total orderings of capabilitarian well-being - especially those that do so by treating different kinds of capabilities as if they can all be traded off among one another - are actually measuring Sen and Nussbaum's concept. This is not to say the problem is unrecognized. Indeed, many authors working within the capabilities approach question the credibility of practices that weight different kinds of capabilities indicators and then combine them into total orderings – often according to a procedure similar to that for A-D above. In her (2018) article on "The Weight of Weighting – an Empirical Study Based on the OECD Better Life Index (BLI)," Mónika Galambosné uses randomized permutations of the weighting scheme for the BLI to show that 15 of 38 countries in the BLI

[8] See chapter three section five.

can shift by at least 10 ranking positions depending on how weights are arranged. Similarly, in an earlier critique of the Human Development Index (HDI), Allen Kelley argues that "weighting requires detailed justification, about which the HDI is substantially silent" (Kelley 1992: 320). And in a recent analysis of the Multidimensional Poverty Index (MPI), Sabina Alkire et al. recognize that "more substantial changes in the … weighting schemes do have relevant implications, e.g. for subnational rankings" (Alkire et al. 2019: 21). Such authors explain that small changes to weights cause unintuitive effects on rankings, suggesting the procedure may not be capturing the target concept, or that, in the least, clearer justifications should be given for the implementation of that kind of procedure.

Yet despite widespread recognition of the concern with weights used to provide total orderings, practitioners working within the approach typically do not pay sufficient attention to the deeper conceptual issue: the lack of coherence with respect to Sen and Nussbaum's original characterization. Chowdhury and Squire (2005), for example, survey various methods for setting weights among indicators, and argue in support of some methods and against others without discussing how *all* methods they consider always provide total orderings of capabilitarian well-being, typically by treating different kinds of indicators as if they can all be traded off among one another.[9] Others authors, such as Alkire et al. (2011, 2015a, 2015b), and Flavio Comim et al. (2008: 163-164), do connect the recognition that these factors are genuinely different such that there is no general way to fix their importance relative to one another with the problem of how to set weights used to provide total orderings. Yet such authors then tend to rely on generally-accepted statistical remedies such as sensitivity analysis or various 'robustness checks' on weights, side-stepping the underlying philosophical issue.[10] Call this kind of shoulder-

[9] Similarly, see Nguefack-Tsague et al. (2010) as an example of an argument for a particular method of weighting the HDI via Principal Component Analysis, for which weights are justified because they explain a high proportion of the 'variance of the data' - in other words how much change in rankings each well-being indicator can be 'responsible' for depending on permutations of the relevant weighting coefficient. While this kind of justification is operationally valid, it is not coherent if Nguefack-Tsague intend to measure capabilitarian well-being as characterized by Sen and Nussbaum. See chapter five section two, in which discuss a similar incoherence for the factor analysis used by Jaya Krishnakumar (2007) to set the weights for her improved human development index.

[10] Following the precedent set by the UNDP's weighting scheme used for the Human Development Index, Alkire and Foster (2009) argue that weights among different kinds of capabilities indicators might as well be equal in the default, so long as the variables are commensurate. Further arguments - not discussed in their paper - would be needed to justify alternate weighting arrangements. Alkire and Foster (2009) cite Atkinson (2003) when they claim the following: "By defining a poverty measurement methodology based on deprivation counts and simple averages, we have implicitly assigned an equal weight of w_j=1 to each dimension j. This is appropriate when the dimensions have been chosen to be of relatively equal importance. As Atkinson *et al* observe, equal weighting has an intuitive appeal: 'the interpretation of the set of indicators is

shrugging position an argument according to the principle of the Justification for Total Orderings on the Basis of Practical Comparison (JTOBPC). The proponent of JTOBPC follows the deliberating council from the example above when she argues it is often at least useful - and arguably reliable - to treat numerical representations of practical comparisons *as if* they are processes that can reliably be used to always provide total orderings of capabilitarian well-being.

Let's take a more detailed look at how this JTOBPC kind of argument plays out. In "Understandings and Misunderstandings of Multidimensional Poverty Measurement," Alkire and Foster (2011) note that procedures that treat different kinds of capabilities indicators as if they can be traded off among one another in order to always provide total orderings of items could sometimes be inappropriate if those indicators are 'noncommensurate.' In such 'noncommensurate' cases, those capabilities should be analyzed by a method that does not treat all indicators as if they can be traded off against one another, and presumably one that does not always provide total orderings (Alkire and Foster 2011: 13). Alkire and Foster seem to indicate that some kinds of capabilities indicators are of such inherently different kinds that there is no fact of the matter regarding how to rank them. However, in the same (2011) paper, Alkire and Foster argue that "values or weights afforded … are straightforward in interpretation and lead to measures whose numerical values *convey meaning*" (ibid: 17, emphasis added). They then argue that, because they convey sufficient (if sufficiently vague) 'meaning,' equal weights are sufficient for purposes of always creating total orderings of capabilitarian well-being (ibid: 18).[11]

---

greatly eased where the individual components have degrees of importance that, while not necessarily exactly equal, are not grossly different' (2002, p. 25; see also Atkinson 2003 p. 58)."
Yet Atkinson (2003) also gestures to Sen (1987a) who notes that a "range of weights be considered" (Atkinson 2003: 59), and Atkinson only discusses equal weight-setting in the context of a survey of different functional forms of the social welfare function that can be used to avoid 'overly demanding' dominance conditions for multidimensional welfare measures. It would seem that Atkinson is taken out of context by Alkire and Foster 2009. Moreover, Atkinson (2003) also offers the familiar kind of shoulder-shrugging response to the problem of setting weights for different kinds of indicators, then proceeds to a more involved discussion of headcount measures (rather than poverty depth measures), which appear to have less restrictive dominance conditions (Atkinson 2003: 63). I return to a discussion of headcount measures in chapter five section three.

[11] As cited by Alkire and Foster (2011), Decanq and Lugo (2013) provide an apt summary of different methods by which weights can be selected. Like Alkire and Foster, Decancq and Lugo note that the setting of weights is a problem that regards the normativity of different values, and so it has normative implications. After even-handedly analyzing different weight-setting methods, the authors conclude that the "form of the transformation of the original variables into commensurable units and the parameter of substitution between dimensions also play an important role. However, these components are more often than not ignored in the literature." (Decanq and Lugo 2013: 24). Yet Decancq and Lugo, like Alkire and Foster (2011), simply shrug away the problem, other than to recommend sensitivity analysis of alternative weighting schemes (ibid: 24). See also chapter four section four.

Note with respect to Alkire and Foster's method that if meaning-carrying weights were to be created only for kinds of capabilities that are not 'non-commensurate,' the measure would avoid the issue. Yet Alkire and Foster argue for and make use of indicators that measure living standards, health, and education. Such indicators must, if they are to follow the characterization of Sen or Nussbaum, either be evaluated according to different kinds of well-being values or stand for different kinds of capabilities (or both), and so they will be 'non-commensurate,' unless Alkire and Foster (2011) mean something different by that term.[12] Alkire and Foster do not directly address this concern. Instead they argue that "because the dimensions, indicators, weights and cutoffs of our methodology are flexible, they can be filled in in many ways. Robustness checks should normally be implemented to ensure that key points of analysis are robust to a range of plausible parameters" (ibid: 20). I question this claim that robustness checks – or some similar statistical practice - provide a sufficient remedy for recognizing and then mitigating ways in which the measure may be off-target in terms of the weights provided.[13] I follow Decancq and Lugo (2013: 30) in questioning whether the trade-offs implied by the Alkire and Foster weighting procedure are in keeping with the deeper intuitions of those doing the evaluation, especially when such trade-offs do not fit with Sen and Nussbaum's commitment to respect the qualitative integrity of different capabilities, and Nussbaum's commitment to avoid such trade-offs in general. I provide a deeper analysis of the Alkire and Foster measure of deprivation depth in chapter five.

In summary, practitioners such as Alkire and Foster (2011) or Chowdhury and Squire (2006) act in accord with the JTOBPC when they develop or defend measures that always provide total orderings of capabilitarian well-being without addressing the commitments of the characterizations given by Sen and Nussbaum and informed by Aristotle. Similarly, practitioners such as technocrats and politicians who use such measures of capabilitarian well-being to justify different policy orientations do not provide arguments for why they treat the measure as if it is generally on-target when this may not actually be the

[12] Their (2011) paper, unfortunately, contains no definition for the term 'non-commensurate.' It is possible they mean 'non-commensurate' in the sense that different kinds of capabilities are not 'roughly on par,' but then Alkire and Foster would seem to be suggesting that living standards, health, and education are valued such that they are 'roughly on par.' The implication is at least unintuitive, as I demonstrate in chapter four.
[13] Of cause for concern is the assumption that checking into 'a range of parameters' provides a sufficient basis for determining the reliability of the measure. I return to this point in chapter five.

case. I claim that their unspoken position can be determined from their actions. It is possible that these practitioners are committed to treating 'capabilitarian well-being' as a single grade, in which case they no longer measure the multi-dimensional concept offered by Sen and Nussbaum. However, if they are committed to Sen and Nussbaum's characterization, possibly they are attempting to ignore the incoherence because the measure might be roughly useful with respect to the intended purpose. In such cases, the incoherence at least warrants further investigation in order to determine the significance of the danger that the measure may be off-target. I return to a passing discussion of this more practical issue in subsequent chapters, including consideration of how the purpose to which the measure is lent provides grounds for considering the harms that may result when the measure is off-target.

1.2 Aristotelian Philosophy, Well-being Values, and Practical Wisdom (*Phronêsis)*

A core concept discussed throughout Aristotle's Nicomachean Ethics – and one that informs both Sen and Nussbaum's characterizations of capabilities - is that of practical wisdom (henceforth *phronêsis*). *Phronêsis* is an intellectual virtue by which an individual – or, as we shall see in Aristotle's Politics below, a group of individuals – recognizes the goods relevant to a given situation in order to realize and promote excellent – or '*aretaic'* ethical action. Individuals who act according to *phronêsis* combine the experienced recognition of different goods – particularly virtues - with action-guiding comparison of options. Importantly, *phronêsis* is not a pure form of knowledge like that of mathematics. It is instead an intuitive receptivity to the particular qualities that are relevant in a given situation, a receptivity that Aristotle claims humans at least partially share with animals *(EN* VI.7, 1141a30). Unlike that of animals, however, human *phronêsis* incorporates further cognitive processes like the use of language and abstract reason. Aristotle describes *phronêsis* as the following:

> Practical Wisdom … is concerned with things human and things about which it is possible to deliberate; for we say this is above all the work of the [person] of practical wisdom, to deliberate well, but no one deliberates about things invariable, or about things which have not an end which is a good that can be brought about by action… Nor is practical wisdom concerned with universals only- it must also recognize the particulars; for it is practical, and practice is concerned with particulars.
> *(EN* VI.7, 1141b8-18. tr. Ross).

Phronêsis concerns complex, particular situations about which we can deliberate. We shall see in chapter three that Nussbaum argues that recognition of different kinds of values – especially the ways in which those values are expressed - is key for processes of deliberation.[14] What Aristotle calls 'goods of the soul' such as different virtues or different kinds of virtuous actions are members of the same set of phenomena that Nussbaum – like Ruth Chang (2015) - would call 'values.'[15] Nussbaum claims to follow Aristotle in

[14] See, for example, Nussbaum (1986a) *The Fragility of Goodness* Nussbaum pp. 284-304, discussed at length in the next chapter. See also Nussbaum (2000a) *Women and Human Development: The Capabilities Approach* pp. 34-110. The topic is a point of return for the remainder of the thesis. In chapter three I also relate how Elizabeth Anderson's pragmatic theory of expressive value characterizes the process of deliberation in a way that independently validates Nussbaum and Sen's characterizations of deliberation. See chapter three section five.

[15] See, for example, Nussbaum in *The Fragility of Goodness* in which Nussbaum claims "Aristotle has … denied the commensurability of values" (1986a: 309) or – among many other similar claims – Nussbaum's later point that "Commensurability loses us the distinct nature of each of the values we cherish" (ibid: 310). See also section four of this chapter for further discussion regarding 'goods of the soul' and Aristotelian virtues.

claiming that ethical agents deliberate by recognizing the particular values that are constitutive of concrete, practical problems - problems like policy assessments and interventions – instead of relying only on abstract, idealized values. Similarly, Sen (2009a) defends a 'non-ideal' process of deliberation, one that compares different real situations against one another, rather than attempting to maximize one or another abstract concept like total utilitarian pleasure or even equality.[16]

To get a better sense of how deliberation involving *phronêsis* might work, let's consider an example from Sen that I also discuss in the next chapter: Annapurna, a social worker, must choose whether to give the job to Dinu, the poorest, to Bishanno, the most depressed, or to Rogini, who has the least physical health.[17] If Annapurna – or by extension a deliberative panel assessing policies in terms of their impact on capabilitarian well-being - is to make a prudent decision, she must recognize and even *empathize* with the different kinds of experiences of Dinu, Bishanno and Rogini as part of her evaluation process. The values related to economic efficiency (Dinu), mental well-being (Bishanno) or physical health (Rogini) must be *separately* recognized along with other concerns related to equality, well-being, agency, or surrounding social norms and values.[18] Moreover, as is discussed in chapter three, Annapurna's selection of one worker does not nullify the tragedy of not being able to help the other two, as if an increase is one is substitutable for a lack of the other. One kind of well-being value is heavily resistant to being traded off with the others because different kinds of well-being values exhibit *asummetria*, introduced in the previous section.

Aristotle claims "each state of character has its own ideas of the noble and the pleasant, and perhaps the good [person] differs from others most by seeing the *truth in each class* of things, being as it were the *norm and measure* of them" *(EN* III.4, 1113a32, tr. Ross, emphasis added). Here the 'good

[16] See Sen *The Idea of Justice* (2009: 15-18).
[17] See chapter two section two.
[18] See chapter two section two regarding, for example, the important difference between agency freedom and well-being freedom. As discussion in chapter three of Nussbaum's *tragic* trade-offs will make clear, the reduction of one kind of value to another allows for a flawed recognition of the particular situation.

person' refers to the agent who best uses practical reason.[19] Annapurna's experience-based recognition of the particulars relevant to each situation functions as a practical basis for her rather difficult decision; in this sense, her recognition of relevant particulars is the *measure* of each situation. And again, the particularity of the relevant values is a feature of that recognition: the *aretaic* agent of *phronêsis* not only recognizes but even *measures* different truths for different classes of phenomena, for example the subtle and distinct kinds of values that are relevant for Annapurna's evaluation. Translator W.D. Ross' use of the term 'norm' is also illustrative here; the agent of *phronêsis* is aware of the relevant social norms as types of values, for example norms concerning how to address lack of mental health or physical health, etc. In this sense it matters whether Annapurna has had the kind of experiences that enable her to recognize and apply the relevant social norms, or to credibly interpret and evaluate the life-positions of Rogini, Dinu and Bishanno.

But is experience really necessary as a guide to prudent action? It is, after all, plausible to claim that if one has the technical knowledge, say of the relevant abstract principles and how to at least coherently combine them, then one can contribute well-reasoned – or at least influential – claims within processes of deliberation. However, Aristotle defends the role of experience as necessary for making prudent decisions when he claims "while young [people] become geometricians and mathematicians and wise in matters like these, it is thought that a young [person] of practical wisdom cannot be found. The cause is that such knowledge is concerned not with universals but with particulars, which become familiar from experience" *(EN* VI.8, 1142a12-15, tr. Ross). We should intuitively be able to understand that there is a difference between technical knowledge of general theory, say the kind of knowledge given in a book, and that given by experience.

Aristotle makes a similar point at the beginning of the Nicomachean Ethics, when he notes that political philosophy is the most authoritative 'master art' that one can study, but that it can only be properly practiced by an experienced person *(EN* I.2, 1094b, tr. Ross). "Hence," says Aristotle, "a young

[19] Aristotle makes a similar claim that the good man is the measure "of each class of things" *(EN* IX.4, 1166a10, tr. Ross) and again claims that "virtue and the good man as such are the measure of each thing" *(EN* X.5, 1176a18, tr. Ross).

[person] is not a proper hearer of lectures on political science; for [she] is inexperienced in the actions that occur in life, but its discussions start from these and are about these… it makes no difference whether [she] is young in years or youthful in character; the defect does not depend on time, but on [her] living" *(EN* I.3, 1095a1-10, tr. Ross). The distinction here is important: inexperienced people may provide abstract intellectual representations of the relevant situation, but Aristotle makes the intuitively accessible claim that some values can only be recognized through one's experience of the particular.[20] It matters, for example, whether Annapurna or others on the deliberative panel have actually experienced poverty or disabilities related to mental or physical health. Previous experiences shape how the relevant expressions of value are recognized and impact further decisions made on the basis of the measurement. As is further discussed in the next chapter, even the articulation of the reasons used to make the relevant evaluation are influenced by one's experienced recognition of the particular.

How might an experienced recognition of the particular work as it relates to recognition of different kinds of values relevant for justifying policy decisions? Let's look to another relevant bit of Aristotelian philosophy. In the third chapter of the Politics, Aristotle makes - in the context of discussing the ideal of justice - an interesting claim about kinds of political qualities that should not be rendered in terms of one another for the purpose of determining how to distribute political offices. Aristotle considers the argument that there are different kinds of qualities that might matter more for rewarding a political office than would others. Purported qualities include one's level of wealth, nobility, and virtue, and Aristotle argues that any distribution given according to comparisons – and implied tradeoffs - among these qualities of these would not be a just basis of distribution.

> But, surely, if this is true, the complexion or height of a man, or any other advantage, will be a reason for his obtaining a greater share of political rights. The error here lies upon the surface… upon this principle, any good may be compared with any other. For if a given height may be measured against wealth and against freedom, height in general may

[20] Parent-hood is an apt example. I would claim, along with many others, that it is simply not possible to relate what it means to become a parent, even to the most apt and attentive audience. 'Parent-hood' must be experienced to be understood; the experience of becoming and inhabiting the role of 'parent' is a simultaneous re-tooling of one's affect, cognitive functions, and in many cases one's identity. As Ernst Cassirer would note, there are experiences that outstrip even our most accurate propositional attempts to characterize them; even the most perspicuous analytic descriptions point merely to a shell that lacks affective pith, the kind of thing that Sen or Nussbaum, following Aristotle, might call a 'value.' For more on this point, see the discussion of expressions within public deliberation in the next chapter.

> be so measured. Thus if A excels in height more than B in virtue, even if virtue in general excels height still more, all goods will be commensurable… But since no such comparison can be made, it is evident that there is good reason why in politics men do not ground their claim to every sort of inequality any more than in the arts. (*Pol.* III.12, 1282b28-1283a13, tr. Ross).[21]

The passage indicates that different kinds of intuitively important qualities cannot be compared relative to a measure of political merit, especially when the measure is used to justify one or another distribution of different kinds of resources. If wealth and virtue could be compared, then, says Aristotle, surely height and virtue, or other odd combinations of qualities could be compared, as if they could be traded off in terms of one another, in order to award resources. Note that systems that measure total capabilitarian well-being of individuals or populations function within a similar kind of context. As summarized by Ingrid Robeyns (2006), they can help to assess policies that distribute resources such as health, education, or other kinds of capabilities. Though Aristotle's example specifically regards the meritocratic distribution of government offices, the general principle we can distil is that there are practical reasons for why different kinds of values that feature for different items should not be compared – and by implication traded off - with one another for the purposes of political distribution of resources.[22]

Phronêsis concerns not only the deliberations of individuals but also that of groups. In the Politics, Aristotle further discusses the role of *phronêsis* in a social context in which groups of people deliberate in order to arrive at action-guiding conclusions. In that context, the well-governed state need not be constituted entirely by virtuous and wise citizens (*Pol.* III.4, 1277a). In the democratic case in particular, public deliberation is a communal act in which different kinds of virtues expressed or recognized by different citizens influence the holistic appreciation of the relevant situation. Aristotle defends this kind of evaluative process when he compares the wisdom of the multitude to that of a small governing body exhibiting practical wisdom:

[21] A point of clarification must be made here. Though this English passage uses the word 'commensurable', the above argument emphasizes that different kinds of political qualities are *non-comparable.* What Ross translates as 'commensurable' is not the Greek term *summetria*. Consider that W.E. Bolland translates the relevant phrase as the following: "So that if A excels B in size more than B excels A in virtue, and if, on the whole, size excels virtue more than virtue excels size, all things could be brought into *relation and comparison*" (*Pol.* III.12, 1283a7, tr. Bolland, emphasis added). Bolland's notion of 'relation and comparison' seems more apt here than Ross' translation.

[22] I return to this point regarding tragic trade-offs in chapter three, and I return to this Aristotelian point about distribution in chapter four.

> For the many, who are not as individuals excellent men, nevertheless can, when they have come together, be better than the few best people, not individually but collectively, just as feasts to which many contribute are better than feasts provided at one person's expense. For being many, each of them can have some part of virtue and practical wisdom, and when they come together the multitude is just like a single human being, with many feet, hands, and senses, and so too for their character traits and wisdom. (*Pol.* III.11, 1281b1-6, tr. Ross).

Let's first note the obvious role given to experience, and the related need for a deliberative body to include people who can recognize and express the relevant kinds of values. When reading the quote, one might wonder how the multitude can be a better judge than a few excellent people. Aristotle recognizes that in some cases, especially when most members of the multitude lack practical wisdom, it is actually imprudent for the multitude to make a decision that could have been made by a smaller group of agents who do exhibit sufficient practical wisdom. Yet, so long as practical wisdom is shared above a certain threshold such that the relevant values can be recognized and reasonably considered through public discussion, groups of people can make better prudent decisions in practical situations, and often can make better-informed situations than could any single individual or small group.[23] The implication for this thesis, then, is that groups of people can and do deliberate and act according to *phronêsis*. Furthermore, if group deliberation according to *phronêsis* is to be in accord with the political philosophy that informs Sen and Nussbaum's characterization, it should be constructed such that deliberation focuses on, protects, and gives influence to the distinctiveness of particular values that are best recognized by experience.[24]

Groups of citizens think and act according to the constitutions of their political state. In a discussion of how a statesman should classify different kinds of governments, Aristotle offers the illuminating aside that "one must not overlook the varieties of the constitutions, how many they are and how many ways they can be combined. And it is with this same practical wisdom that one should try to see both which laws are best and which are appropriate for each of the constitutions" (*Pol.* IV.1, 1289a10-14, tr. Ross). Here again we see an emphasis on the recognition of the particular; though constitutional

[23] A good contemporary analytic discussion and proof of this concept comes from Scott Page, who argues in "Making the Difference: Applying the Logic of Diversity" that it is precisely the greater level of experience of a larger, more diverse group, that gives the group a stronger predictive ability than a less diverse group of a similar size. See Page (2007: 6-20).
[24] I return to this point in chapter four section three as it relates to what I call *Value-Omission Incoherence*.

law may be written to be understood as universal principle, the contextual assessment of whether or not some part of it is appropriate is actually a practical judgment of the fit of the constitution to the history and values particular to that context. As we shall see in chapter three, the relevance of *phronêsis* to the analysis of constitution and law also foreshadows the constitution-building purpose linked to Nussbaum's characterization of capabilities.[25]

[25] I return to this point especially in chapter three section two. This point is furthermore relevant because, as discussed in chapters three and four, the vector representational form – what I call the 'Vector View' in later chapters - that structures most indices or similar total orderings of capabilitarian well-being is more clearly linked to Nussbaum's characterization and representation of capabilities than it is to Sen's settled representation of sets of sets of co-possible human activities, or functionings.

1.3 Problematic Total Orderings of Capabilitarian Well-being: Aristotle's Analysis of Exchange Value and 'as if' Measurement of Different Kinds of Well-being Goods

As Scott Meikle relates in his book Aristotle's Economic Thought, Aristotle not only created a theory regarding the roles of different kinds of political and moral values in ethical decision-making, he also considered the related question of how economic value can be assigned to objects that have different constitutive qualities. Aristotle gives this analysis in a few short chapters of his Nicomachean Ethics Book V and in the Politics Book I chapters 8-10. Though the writing is brief, it covers a weighty topic that demarcates the neo-classical economic theories of Jevons, Gossen, Walras and Menger from the classical economic theories of Adam Smith, Ricardo, Mill and Marx. That topic is s*ummetria* – often translated as 'commensuration' - of economic goods.[26] Meikle explains that the neo-classical economists believed that different kinds of goods can be measured, among other outputs, and totally ordered in terms of their common utility value. In contrast, economists in the value-supporting camp deny that a single metric can universally capture the different kinds of ways that humans value such goods. We can from this list of names better see the scope of the Aristotelian tradition: Aristotle's analysis provides a conceptual nexus for an ongoing debate regarding the roles of different kinds of values in economic measurements, including whether and how different kinds of values can be put into total orderings.[27]

Aristotle's analysis of *summetria* as it relates to economic goods proceeds through a set of intuitive examples to demonstrate two important claims: (i) exchange value as a measure does not fully capture the different values – including well-being values – of different economic goods, and (ii) exchange value as a measure does not actually *create* the value or set of values that is treated as a common currency used for the universal exchange of those goods. Regarding (i), let's consider Aristotle's well-discussed examples of shoes, houses, and beds. It may be possible to trade these objects in market-established ratios such that - as Aristotle maintains - one house is traded for five beds, or that

[26] See Meikle (1995). Further analysis of the ontological and metaphysical distinctions underlying Aristotle's use of *summetria* is provided in the final section of this chapter.
[27] For an excellent discussion of this topic, see Sen's On Ethics and Economics (1987b)

50 pairs of shoes trade for one bed (assumed here for mathematical ease). Though any real market would actually feature contextually-sensitive variations of these ratios depending on factors like bargaining positions, trade barriers, or amount of information available to negotiating agents, let's assume that market-established tradeoffs are generally held consistent such that one house goes for 250 shoes, etc.[28] Aristotle explains that currency can stand in for the ratios established by the barter economy, for example that one house is ten minae, since one bed trades for two minae, and a shoe is also traded as $1/25^{th}$ of a minae, etc. It could be argued, then, that because there is some common exchange value that holds among different economic objects, the exchange value universally captures the ways in which those objects are valued, and also allows for the possibility of their total ordering in all cases.

However, different kinds of objects exhibit *asummetria*: they are constituted by particular qualities that cannot actually be reduced to a common exchange value. A bed and a house arguably share some properties in terms of the potential leisure value they create or the materials they are made of, but further constitutive qualities of a house - including the doings or beings that it supports - differ significantly from those of a bed, and even more clearly those of a shoe (or an orange, a litre of gasoline, and so on). For example, we cannot say that properties of beds such as having mattresses or providing a place to rest are fully captured when rendered according to properties of shoes such as having soles or providing for better mobility, even if beds and shoes can, as a practical matter, be traded off in terms of their exchange value. Though beds, houses and shoes can be represented *as if* they are 'commensurate' according to their exchange value, their relative prices are shaped by an underlying market system that gives value according to market principles such as scarcity, access to information, tariffs, elasticity, etc. Note that it is possible to counter that some of the qualities constitutive of different goods may actually be reflected through processes of exchange – for example qualities like size or mass as captured by shipping cost or labor cost, or chemical composition as captured by how well price captures scarcity of certain resources. But even then it remains unclear how market principles provide the means for measuring

[28] For an excellent ethnographic perspective on the reality of such practices, see Market Threads: How Cotton Farmers and Traders Create a Global Economy by Koray Çalişkan. In fact, exchange values established by markets are usually starting points of negotiation, and trading proportions fluctuate significantly depending on myriad factors related to context and practice.

further kinds of related human doings and beings and, similarly, their related well-being values in terms of one another. For example, in 'Incommensurability and Valuation in Law,' Cass Sunstein (1993) explains that we have strong moral and legal reasons not to value and trade *all* potential economic objects in terms of one another. Humanity has a problematic history of treating living things – including human beings – according to their economic value.

Regarding (ii) above, Meikle's Aristotle holds that a measure "does not create the property in things which it measures" (Meikle 1995: 23). If minae is the common currency taken to be the what-it-is that allows beds, houses and shoes to be rendered into measures that yield total orderings, one might assume it could only be because there is an underlying 'economic value' intrinsic to that common currency that represents the total set of values featured by and relevant for each of those different objects. Supporting this assumption, again, is the tendentious claim that the value created by the market really just is the value of the object, but here that claim can be used to reinforce the further view that the currency-qua-market-instrument is actually *measuring* the values of either object. Meikle argues that to make this kind of an argument would be to put the cart before the horse; the value of the object is there *before* the exchange value 'measures' it, in the way that length is there before the ruler measures it (ibid: 23). We cannot, therefore, claim that the values that feature in beds or houses are represented in terms of one another by the process of 'measuring' their exchange value. It is, again, the practical combination of factors and principles of the marketplace that is captured by exchange value. That underlying system of factors and principles is not sufficient for the purpose of capturing different kinds of qualities – especially well-being values - constitutive of different kinds of economic objects.

Given the preceding arguments, we now have a solid case against the claim that Aristotle allows that exchange value accurately represents different kinds of constitutive qualities - including different kinds of well-being values - of economic goods. Yet Aristotle looks to another plausible underlying basis that might justifiably capture all constitutive qualities. If money is not what accurately represents different kinds of constitutive qualities against one another, then perhaps it is simply need – in Greek

chreia - that does so. Here the use of *chreia* as an underlying, universal currency mirrors how contemporary economists use utilitarian pleasure or desire. If underlying need – often called 'demand' - rather than money captures the total set of constitutive qualities found in different economic goods, then different states of well-being could on a similar basis be represented in terms of the underlying 'demand' – in terms of utilitarian pleasure or desire - for each. Meikle, however, explains that Aristotle "never [directly] links *chreia* with commensurability (*summetria*)" (ibid: 25) and, most importantly, Meikle translates Aristotle as claiming that when different goods are represented against one another according to *chreia*, they are done so *only sufficiently*, merely for practical purposes (ibid: 25; see also EN V.5, 1133^{b} 19-20, tr. Ross).[29] Though *chreia* can be used as a basis for representing the well-being value of different objects such that they can be traded off against one another, it does so artificially. It is not a basis that can be used to *measure* the different, essential qualities of different economic objects.

We shall see in the next chapter that Sen takes a similar position against the proponents of JTOBPC who ignore the commitment to avoid measurement processes that always create total orderings of capabilitarian well-being, especially when such processes treat different kinds of capabilities as if they can all be traded off in terms of one another. Different kinds of capabilitarian well-being values are not *measurable* according to some common currency, even if such processes of exchange often do occur as a practical matter. The discussion in this section helps us to see how Aristotle argues that it is possible, in an economic context, to misinterpret a system of exchange – in this case that of trading houses and beds according to their common exchange value in *minae* - for comprehensive, unit for unit measurement as *summetria.* The analysis also demonstrates that according to Aristotle different economic goods have different constitutive qualities, especially in terms of their well-being values. We shall see that just as on Aristotle's view different economic objects cannot, in general, be put into total orderings in terms of their

[29] As a means of validating Meikle's translation, consider the same lines translated by David Ross: "Now in truth it is impossible that things differing so much should become commensurate, but with reference to need they *may become so sufficiently.*" (EN V.5 1133^{b} 19-20, tr. Ross, *emphasis added*). For an illuminating cross-reference, see also the same lines translated by Terrence Irwin: "And so, though things so different cannot become commensurate in reality, they can become *commensurate enough in relation to our needs*." (EN V.5 1133^{b} 19-20, tr. Irwin, *emphasis added*). Both translations recognize (i) that commensuration of different kinds of economic goods is not possible, and (ii) that they *may* be treated as if commensurable, or commensurate *enough*, relative to our practical needs as agents interacting in a market context.

well-being values, especially according to how those objects are traded off in the marketplace, so too the qualities that constitute different kinds capabilities cannot be coherently measured by a system that always provide total orderings of Sen and Nussbaum's capabilitarian well-being.

1.4 Commitments to Value-Ladenness and against Total Orderings: A Pragmatic Prelude

In his article "Plural Utility," Sen (1980) discusses claims from Marx and Mill, who follow Aristotle's assertion that different kinds of well-being are both value-laden and are resistant to being rendered in terms of one another. Mill famously alleges that there are different *kinds* of pleasure, the relevant values of which are known by experience, while Marx claims that total individual utility – as either pleasure or desire - amounts to a 'metaphysical abstraction' that distorts the different ways phenomena are valued, much in the same way that universal exchange value fails to capture different kinds of well-being values.[30] Nussbaum and Sen take this line of thinking into their approach to measuring capabilitarian well-being. In their paper "Internal Criticism and Indian Rationalist Traditions," they argue in favor of an Aristotelian process of deliberation – what they call a 'reflective assessment of values' – that allows a group to find "a view according to which they can live together in community – a shared and sharable view of value" (Sen and Nussbaum 1987: 24).

This evaluative process includes a feature Sen and Nussbaum call 'non-commensurability,' which they define as a commitment to "treating each of the values involved as a qualitatively distinct item, not reducible to any other item, not conceivable as simply a certain quantity of something else." I take this sense of 'qualitative distinctiveness' to be a coherent paraphrase of Aristotle's *asummetria*, at least insofar as the term indicates that different kinds of capabilitarian well-being values are resistant to being rendered in terms of one another, especially according to measures that always provide total orderings of items. In this section I further outline Aristotle's method for measuring well-being and I introduce the practical importance of remaining committed to the 'qualitative distinctiveness' of different kinds of capabilities in the context of measuring capabilitarian well-being.[31]

[30] See previous section.

[31] There are, of course, other principles that can be taken as key to the Aristotelian process. In the same (1987) paper, for example, Nussbaum and Sen both recognize the importance of the "essentiality of the particular" and the "role of emotions and the imagination." Both of these features are discussed in this and the next two chapters, most notably through this chapter's discussion of *phronêsis* but also through the next chapter's analysis of Sen's pluralistic focus to deliberation, and Nussbaum's defense of *prohairesis*, or emotional awareness, in chapter three. However, for the purpose of critiquing contemporary capabilities measures, this thesis focuses on the commitments to value-ladenness and qualitative distinctiveness as central to the Aristotelian process.

Before I outline how Aristotle's philosophy can be used to justify the commitments to value-ladenness and qualitative distinctiveness in the work of Sen and Nussbaum, I address a question commonly voiced in development work. Putting aside the recognition that Aristotle is the first known scholar in the Western tradition to discuss *asummetria* of well-being values, why does the capabilities approach to measuring well-being have in its foundations all this discussion of one ancient Greek philosopher? Surely the risk of invoking the writings one or another foundational intellectual is that values particular to that philosopher or to their culture (or both) might exclude other relevant perspectives? Sen seems to address this concern of cultural particularity when he claims that according to Nussbaum's (1988) translation of Aristotle there is – at a certain level of generality – one perfectionistic list of capabilities that defines a *eudaimonic* life (Sen 1993: 45).[32] Sen argues against the implication that the one list is appropriate to all cultures. To be fair to Sen's intuitions, let's note that Aristotle's philosophy does uphold a uniquely Greek or Western notion of certain virtues.[33] Consider, for example, Aristotle's defense of '*megalopsuchia,'* a Greek notion of pride that supports one's having a quickness to anger.[34] If it really defended a particular set of cultural norms without giving due consideration to other ways of life, Aristotle's philosophical view would seem to be a weak basis for a universally-appropriate method of measuring well-being.

Claims regarding the narrowness or even the indefensibility of some moral or political elements of Aristotle's philosophy should not be understated - especially, for example, his repeated insistence that some classes of people lack the natural abilities to become full citizens. Yet the Aristotelian background can still be defended as appropriate for measuring capabilitarian well-being in diverse, contemporary cultural contexts. In the next chapter I explain that Sen's characterization of capabilities supports a culturally-sensitive process of public deliberation, one that recognizes expression of - and reasons given

[32] For further discussion of perfectionism as it allegedly relates to the characterization of capabilities, see chapter three.
[33] Further support for Sen's position comes from acknowledging that Aristotle notoriously claims that women and slaves lack sufficient reasoning abilities to be treated as equal citizens of a political community. Nussbaum recognizes this limitation of Aristotle's writing, and can be considered a 'Neo-Aristotelian' inasmuch as she denies some of the less egalitarian claims made by Aristotle as historically contingent, arguing that deeper principles within Aristotelian political and moral philosophy do imply that woman and slaves have the same inner potential and deserve to be treated with the same dignity as citizens. See Nussbaum (1987) Wider Working Paper 31 *Nature, Function and Capability: Aristotle on Political Distribution*
[34] See, for example, Nussbaum (1987b) Wider Paper 32: *Non-Relative Virtues* pp. 11-13.

in support for - different kinds of values. Nussbaum, as detailed in chapter three, emphasizes a similar sensitivity to different cultural contexts – what she calls 'multiple realizability.' On Nussbaum's view the set of norms and emotions associated with a given kind of value, for example one's self-confidence or pride, can be recognized through different cultural lenses, for example as either Greek *megalopsuchia* or Christian *humility* (among many others). In both cases the root *kind* of value - what Aristotle calls a 'virtue' – grounds the relevant set of affective responses, yet the same root virtue remains multiply realizable in different contexts.[35] For example, if one were to receive an insult, the Greek cultural tradition may support immediately fighting back while the Christian tradition may support turning the other cheek, and so on. On Nussbaum's view, a given culture can settle on different kinds of accepted responses - and thus interpretations - for evaluating each kind of virtue.

So how does Aristotle's account of different kinds of well-being values – which themselves admit of significant context-sensitive diversity - actually figure in Sen's account? In his essay on "Capability and Well-Being," Sen describes Aristotle's justification as "an account of the human good [that] is explicitly linked with the necessity to 'first ascertain the function of [humans]' and … then proceeds to explore 'life in the sense of activity'" (Sen 1993: 46). Note that the exploration of 'life in the sense of [each] activity' is necessarily value-laden, while 'ascertaining of function' invokes the commitment that there are distinct kinds of functionings. Sen, like Nussbaum, refers to Aristotle's Nicomachean Ethics (henceforth NE) book I chapter 7, in which Aristotle argues that what is 'good' – or conducive to well-being - depends on what we determine to be the important functions of the relevant phenomenon *(EN* I.7, 1097b). For example, the goodness of the eye is that it can see; an eye that could not see would not be a 'good' or 'well' eye. Similarly, Aristotle notes the following:

> [F]or all things that have a function or activity, the good and the 'well' is thought to reside in the function… if the function of [humans] is an activity of soul which follows or implies reason, and if we say 'a so-and-so' and 'a good so-and-so' have a function which is the same in kind, e.g. a lyre player and a good lyre player, and so without qualification in all cases, eminence in respect of goodness being added to the name of the function…

[35] In the next two chapters I spend some time discussing the related multiple realizability of capabilities. The limited perspective offered in the *megalopsuchia* example is discussed in Nussbaum WIDER Paper 32: *Non-Relative Virtues*, pp.10-11

> human good turns out to be activity of soul exhibiting virtue, and if there are more than one virtue, in accordance with the best and most complete
> *(EN* I.7, 10987b25 – 1098a18, tr. Ross)

Aristotle argues that by chronicling the best – in Greek *aretaic* – human functionings we can generate an account of human well-being. Note that there are distinct kinds of virtues, each of which is exhibited by different kinds of 'activities of the soul.' In fact, according to Aristotle's ontology there are different types of kinds of goods: there are different kinds of external goods such as houses and shoes as discussed in section three, different kinds of goods of the soul such as the different virtues just listed, and different kinds of actions and activities that follow the application of the relevant virtue or set of virtues *(EN* I.8, 1098b).[36] Note also that actions or activities recognized according to different virtues are value-laden, following Ruth Chang's 'ordinary' sense of values from first section of this chapter. When they are recognized as components of a *eudaimonic* life, items that feature or are evaluated according to different, qualitatively distinct goods of the soul, for example activities like getting an education or getting exercise, exhibit *asummetria* such that they recognized separately, following Nussbaum and Sen's use of 'qualitatively distinct' above.

Aristotle furthers the case for qualitative distinctiveness by listing different goods of the soul along with corresponding actions and activities. The list includes the aforementioned appropriate use of human reason, but also includes courage, temperance, being liberal (or even magnificent) when taking and giving, having a certain kind of pride, seeking truth, cultivating friendship, and others *(EN* II.7, 1107b). In each case, Aristotle describes the relevant virtue as a culturally-habituated mean, the attainment of which is pleasurable in itself and the deficiency or excess of which can be described as a vice. For example, how we moderate our pleasantness around others is one kind of virtue, the lack of which is surliness and the excess of which is obsequiousness *(EN* II.8, 1108a). A surly person might be rude to well-meaning friends simply because she is angry for unrelated reasons, while the obsequious person might flatter even those who are unkind to her. Note also that the actual understanding of each

[36] These differences are returned to directly in the analysis of Sen's characterization of the diversity of functionings given in chapter two.

virtue depends on surrounding social and cultural context, as exemplified via the contrast between humility and *megalopsuchia* given above. Yet in whatever context, different virtues are of different kinds, meaning a virtuous agent must be able to distinguish which virtue or set of virtues is relevant for a given situation, and act accordingly. Recall, for example, Annapurna's decision discussed in section two: if Annapurna is to act as a prudent decision-maker, her experience is the basis from which she is able to recognize, separately and as needed, the relevant virtue or set of virtues.

From the foregoing explication we can start to see that Aristotle's picture of well-being is, in practice, quite complex given that different sets of virtues are relevant for different actions or activities.[37] Aristotle summarizes the experienced understanding of this complex picture as *phronêsis*, or practical wisdom, introduced earlier in the chapter. The agent of practical wisdom recognizes that human functionings are intrinsically value-laden given the combination of virtues particular to each, and she also recognizes that different virtues are particular. For example, the agent of *phronêsis* would find it difficult to measure how much some action affects one's pride against how much some other action affects one's feelings of friendship. This latter point is given a more general theoretical framework in EN I.6, in which Aristotle gives an argument against Plato's idea that the good is of a single grade.[38] Aristotle's argument resonates with Mill's above, namely that "the good… is not something common answering to one idea" *(EN* I.6, 1096b20, tr. Ross). Rather, the good – or well-being – is constituted by different, value-laden kinds that are constitutive of or relevant to the particular context.

But is it really practical to treat different kinds of well-being values as qualitatively distinct, especially when this treatment supports the commitment to avoid procedures that always provide total orderings of capabilitarian well-being? I provide a deeper response to this question in chapters three and four, and for now respond with an argument by analogy. So long as practitioners intend to follow the Aristotelian commitment to *phronêsis* as it informs Nussbaum and Sen's characterization of capabilitarian well-being, measurement systems that respect the distinctiveness of different kinds of capabilities are

[37] See, for example, Aristotle's take on the complexity of virtuous activity in *EN* II.9, 1109b.
[38] Plato's argument for a single grade of the good is further discussed below and also in chapter three.

better suited to most purposes than are those that allow for the possibility of total orderings in all cases, especially those that do so by treating different kinds of capabilities as if they can be traded off against one another. As Sen and Nussbaum both allege, in many cases there is simply no fact of the matter whether some items can be ordered with respect to their capabilitarian well-being, and in such cases a faithful measure would have to indicate the same - that the picture provided is 'incomplete.'

Consider for now, as an analogical example, how a two dimensional 'orange peel' projection – the Goode Homolosine projection – provides an aesthetically incomplete representation of the Earth, but also presents a reliable representation of the areas of the different continents.[39] The Goode Homolosine projection allows that, if the globe is to be represented according to a two-dimensional view, there will gaps in the picture provided. As a contrast, consider how the Mercator projection of the Earth produces significantly more off-target measurement outcomes – for example in terms of relative landmass of different continents or spatial mappings of different locations - in exchange for aesthetic completeness.[40]

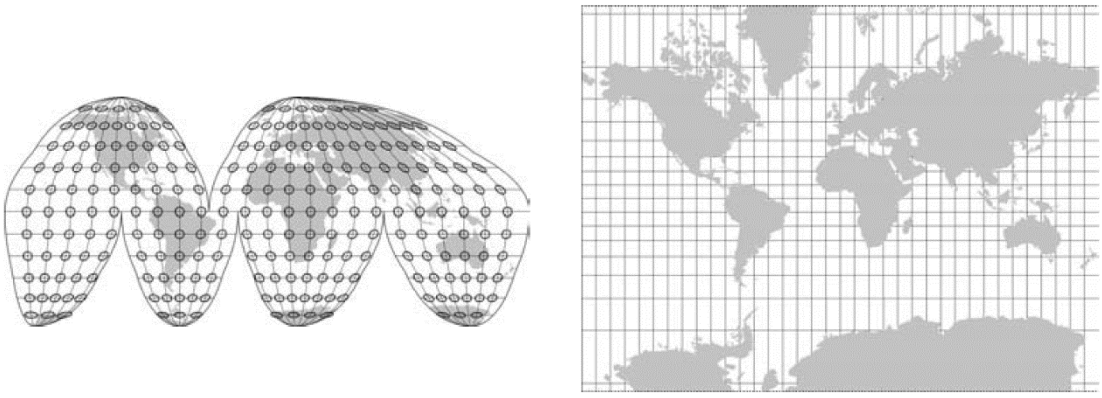

Figure 1: Goode Homolosine Projection (left) and Mercator Projection (right)
from Monmonier *Rhumb Lines and Map Wars* (2004: 131, 33)

[39] Sen and Nussbaum's proposed measures of capabilities follow analogically, though in different ways. Sen (1993) defends what he calls the 'incompleteness' of his capabilities approach. Incompleteness in this sense entails that rankings of different kinds of well-being must be, for the most part, partial if they are to provide an on-target output in terms of Sen and Nussbaum's capabilitarian well-being. As mentioned in the introduction to this thesis, Nussbaum argues against tragic trade-offs, and an implication of her argument is that many total orderings of capabilitarian well-being transgress this commitment. See chapter three section three.

[40] Note, for example, the difference in landmass between Greenland in either projection The difference between different kinds of map projections might matter significantly if, as Mark Monmonier explains in his social history of the subject, one is a world war II artillery officer. The Mercator Projection interestingly dates back to an earlier period of sailing navigation in which it was useful to envision the entire navigable world, even if the picture was warped. That picture became less useful in modern contexts. See Monmonier (2004) pp. 99-109.

Measures that always provide total orderings of capabilitarian well-being are analogous to the Mercator Projection in that they always manage to provide a 'complete' output, one that might be roughly useful for limited purposes such as deciding among policy options. Yet in cases in which there is no fact of the matter in terms of capabilitarian well-being, the 'complete' measure incorporates some set of reasons external to the original concept as characterized. Just as a set of assumptions ground the distorted view of total landmass in the Mercator Projection, a set of reasons external to Sen and Nussbaum's characterization of capabilitarian well-being ground the possibility of providing total orderings for all cases. As I argue in later chapters, further investigation of capabilitarian well-being measures using such procedures is warranted to determine whether, among other practical concerns, the consequence of being off-target might be significant.[41]

There are ways that practitioners might misinterpret Sen and even Nussbaum as tacitly supporting such procedures. Sen (1993) broadens the implications of 'incompleteness' when he claims that "the capability approach can be used with different methods of determining relative weights and different mechanisms for actual evaluation" (Sen 1993: 48).[42] Such methods could be conceived to include those that always provide total orderings (what Sen calls 'complete rankings') of capabilitarian well-being, including those that do so by treating different kinds of capabilities as if they can all be traded off in terms of one another. That broadened conception is useful to those who argue according to the JTOBPC. I allow that such operational procedures, analogically similar to those that create a Mercator Projection, may be indeed called 'roughly useful' measures of capabilitarian well-being relative to certain purposes, for example those for which there are not significant harms associated with being off-target in the way the measures are. The Mercator Projection is, after all, a useful heuristic that provides a general picture of how the continents are situated with respect to one another. Yet that same picture is not even roughly useful for other purposes: it might, for example, lead a navigator seriously off course.

[41] See, in particular, chapter four.
[42] Sen, in fact, makes a number of claims like this that can be misinterpreted, as the next chapter details.

Let's also consider what philosophical justifications might support the JTOBPC-type argument. Two philosophical positions provide substantive justification and are worth analyzing, as they mark a continuum of possible reasons, other than usefulness with respect to limited purposes, that practitioners can give for ignoring the commitment against tradeoffs into total orderings. The first position is that different kinds of well-being values are not qualitatively distinct because what might be called 'well-being' has, deep down, a single measurable quality. This position is defended by Plato's character Socrates in The Protagoras and is further discussed in chapter three. One can hold Plato's position and then claim that magnitudes of single grade 'well-being' units can be measured along a single numerical scale, for example as Benthamic Pleasure, much in the same way that we measure a physical property like degrees Kelvin. Recall that Aristotle challenges this view in EN I.6 when he claims there is no 'common idea' – or single general form – of the good.

The second position is that of the moral relativist who dismisses *asummetria* along with all strong metaphysical assumptions. This position is arguably that of Chowdhury and Squire (2006) whose article on setting weights for the Human Development Index (HDI) surveys multiple methods for setting indicator weights that can be used to provide total orderings in all cases. Chowdhury and Squire accept all surveyed methods and then average the results into a set of 'averaged weights,' one that allegedly validates the equal weighting of indicators used by the HDI. Given that they do not discuss the deeper problems related to Sen and Nussbaum's commitments to both the value-ladenness and the qualitative distinctiveness of different kinds of capabilities, Chowdhury and Squire appear to accept that different kinds of capabilities indicators can be rendered into total orderings or not depending on the kind of quantification assigned contingently by a survey or some similar process. If, for example, some process of deliberation produces a set of specific weights and allows that indicators can be universally traded off according to those weights in order to always provide total orderings of items, then it would seem that Chowdhury and Squire can conclude that the relevant metaphysical commitments – whatever they may be

– favor the claim that the measure succeeds in capturing the relevant set of capabilitarian well-being values.

The Platonic and the extreme moral relativist positions represent two different views along a continuum: one view commits to a metaphysically thick position that supports a single metric, while the other view commits to a metaphysically thin – and arguably circuitous - position that accepts the outcome of deliberation as its own evidence. I have introduced both positions as defeasible in the context of measuring capabilitarian well-being and will continue to argue accordingly. Once we remove both extremes from the continuum, we find ourselves in a pedestrian middle-ground where practical considerations can and do challenge what it might mean to try to use a measurement system that features significant incoherence. At various points in the next three chapters I focus on that practical position as it relates to Sen and Nussbaum's characterizations and representations, exploring contextual differences that preclude or warrant certain capabilities measurement system representations and/or operational procedures.[43]

Even in that pedestrian middle-ground, the extreme moral relativist practitioner that ignores the commitment to avoid procedures that always provide total orderings can still attempt to justify her actions through a further misinterpretation of Sen's work. As we shall see, Sen supports a pluralistic evaluative process, one that allows for many different kinds of moral theories to influence deliberation.[44] The extreme moral relativist could argue that Sen's pluralistic position is actually extremely relativist and so, as indicated in the Chowdhury and Squire discussion above, rankings could be complete or not depending on the whim of those engaged in the deliberative process. However, the sense of value-pluralism defended by Sen is not radically relativist in the way that the position of the extreme moral relativist is radically relativist. Sen's commitment against procedures that always provide total orderings helps to

[43] The Platonic view is further addressed in the next chapter, as it provides partial grounds for treating some core feature of capabilities as a common-currency proxy that can be used to measure different intranational and international populations. See chapter two section five. I call a representative proxy of this type a 'Universal Covering Condition' or UCC and argue, following Anna Alexandrova (2017) that there is no key feature that can suitably work in this manner to represent capabilitarian well-being such that it always provides total orderings of items.
[44] See chapter two section five for further discussion of this topic.

frame a deliberational context in which different kinds of valuable activities or actions cannot be reduced to one another without substantive justification, for example on the basis of particularly urgent reasons given in a public context (Sen 2009a: 241-243). Moreover, Sen (2009a) has argued that deliberative forms of democracy that support the expression of plural forms of value are statistically correlated with the prevention of famines, increased political involvement and political voice, and correspond to sustainable economic development (ibid: 348-354). So while Sen supports pluralism, his own characterization of the incompleteness of capabilities stands against the position of the extreme moral relativist. I return to this discussion in the next chapter.

I give two further responses here to the extreme relativist view that I take practitioners like Chowdhury and Squire (2006) to be holding. The first is that the extreme moral relativist position is a *non-starter* for any observation-based capabilities measurement system committed to descriptive strength. If the nature of capabilitarian well-being is considered to be contingent on the moral position accepted in that context, then the entire process measures little more than a set of dominant social preferences that may themselves be adaptive or oppressive. I return to this point in chapters two and four, for now noting that, for example, if a deliberative body nearly unanimously expressed that the 'proper' vocations of any woman citizen should be limited to caretaking and then set weights in favor of policies promoting gender-based vocations that give poverty-level compensation, the relativist would have no basis for challenging the way in which those weights – and the concomitant total orderings of items - overpower the reasonable counter-expressions given by a small minority.[45] Without some set of methodological and procedural checks that give expression and influence to a wide range of values, deliberation likely tracks away from a pluralist representation of the target phenomena.[46] If there is no foundation for how to discuss, characterize, represent, select, or evaluate capabilities, there is little foundation for claiming the measure is on-target.

[45] The members of such a panel might question why, after all, should there be an equal gender division on the panel.
[46] See, for example, Kevin Olson (2011) and Jack Knight and James Johnson (2011)

Second, though commitments to value-ladenness and to avoid procedures that always provide total orderings certainly can be defended according to various moral positions, they are also defensible through observation and a few pragmatic principles established by Aristotle and defended by Sen and Nussbaum. This pragmatic 'rescue' might help convince the relativist who finds it challenging to follow what might seem to be unsubstantiated metaphysical claims. Let's note that Aristotle's philosophical view relies on a significant amount of observational evidence. For example, the virtuous mean between recklessness and cowardice is developed according to real data, in this case the consideration of real human actions and reactions, and what it means to best 'function' as a human in that real world. It is on this practical basis that Aristotle argues there are different kinds of virtues.

Consider again the different kinds of qualities featured by different physical objects, for example qualities like density and voltage.[47] That we recognize the difference of such qualities is the result of observing the physical world, sometimes called the 'empirical' or 'real' world, though I am not here defending a nuanced interpretation of either of those terms. Similarly, a commitment to best-available observational evidence cannot rule out that virtues – what Aristotle calls the 'goods of the soul' – exhibit *asummetria*. Though we may be able to weakly compare a courageous act with a friendly one, there is an intuitive recognition that, in some cases, there may be no fact of the matter whether the courageous act has more or less value than does the friendly one in terms of capabilitarian well-being. More importantly, it would be a tragic reduction to claim that one such act could be substituted for the other, as if more courage could be a balm for a lack of friendship, and vice versa. Such easily-observable intuitions are central to how Sen and Nussbaum characterize and represent capabilitarian well-being.

[47] See also the next section and chapter two section two for further discussion of different physical qualities.

Chapter Two:

Sen's Characterization of Capabilities
and the Coherence of Related Representations and Operational Procedures

2.1 Introduction

As mentioned in the introduction to the thesis, I use the representational theory of measurement to analyze how a measurement system can exhibit incoherence among key ingredients of characterization, representation, and operational procedures such that there is reason to suspect the system may be 'off-target,' meaning the system is not measuring what it is supposed to be.[1] In this framework the characterization is the set of features conceptualized of the phenomenon, the representations are the logical, mathematical and/or diagrammatic structures used to represent those features, and the operational procedures reflect those features through various practices of measurement, for example the collection and organization of data, use of proxies, implementation of statistical procedures, etc. If an operational procedure or representation (or both) is inconsistent with the characterization, or if an operational procedure does not fit the relevant representation, the system is incoherent.

In this and the next chapter I defend the claim that commitments both to value-ladenness and to avoid procedures that always provide total orderings feature in the characterizations and respective representations of both Amartya Sen and Martha Nussbaum. Systems that allegedly follow either author's characterization yet transgress one or both of these commitments are incoherent. In particular, we should be concerned when the measure still provides an output for situations for which the characterizations indicate there should be no fact of the matter. As discussed in the previous chapter, in such situations the system incorporates external reasons used to justify the rendering of a 'complete' output. Practitioners should at least attempt to make such external reasons transparent as part of an effort to weigh the significance of the possibility that the measure is off-target.

[1] See chapter four section two.

Let's turn to Amartya Sen's characterization and representation. Sen's characterization of *capabilities* can be given the following introductory shorthand:

> **Sen's Characterization (Shorthand):** An individual's capabilities are the extent of that individual's freedom, also called 'opportunities,' to achieve doings or beings they have reason to value.

In the next section I further explicate the characterization, but for now let's note that the characterization emphasizes the 'freedom' that individuals have to actually achieve valuable types of being, for example a meaningful career, a feeling of happiness, or - in a more practical sense - being able to complete a certain level of education, vote, get enough nutrition, etc. Doings and beings that individuals are 'free' to achieve are also those that people can autonomously choose rather than being forced by external pressures into 'choosing' one or another option. In this non-nominal sense, the freedom to achieve various capabilities is what Sen calls 'substantive freedom,' further discussed in subsequent sections.

According to the above shorthand, Sen's characterization directly supports the commitment to value-ladenness in that individuals have 'reason to value' the relevant opportunities. The shorthand does not, however, directly support the commitment that different kinds of capabilities are resistant to always being measured as total orderings. I therefore take some time in the next section to defend the following *reconstructed* characterization as closer to what Sen intends:

> **Sen's Characterization:** Capabilities are the extent of the freedom individuals have to achieve *different kinds of* doings and beings they have reason to value *in diverse ways*.

The italics are included to indicate phrasing that differs from the 'typical' characterization of Sen's capabilities sketched above.[2] This reconstructed characterization incorporates presuppositions that Sen repeatedly defends, for example in his more recent (2009) The Idea of Justice. Sen's capabilities are of 'different kinds' not only because doings and beings are different kinds of phenomena, but because

[2] See the introduction to the next section for further examples of 'typical' characterizations following Sen.

opportunities for these different kinds of doings and beings are *evaluated* according to diverse kinds of values.[3] These differences in kind are what Sen casually labels as 'heterogeneous,' 'qualitatively distinct,' or even 'non-commensurable.' I take Sen's somewhat meandering use of these labels to follow Aristotle's use of '*asummetria'* of different kinds of goods: at least within the context of measuring capabilitarian well-being, different kinds of goods, including the values constitutive of or relevant to those goods, cannot in general be captured by procedures that always give a total ordering of items, especially procedures that treat such values as if they can all be traded off against one another.[4] In the next section I will defend three ways in which capabilities are heterogeneous according to Sen's reconstructed characterization. I then turn to Sen's representation of capabilities, analyzing how the representation fits to core commitments of the characterization.

In the introduction to this thesis I presented the idea that incoherent measures of capabilitarian well-being typically include a vector representation – often called the 'Vector View' - of different kinds of capabilities. Such measures, furthermore, tend to incorporate procedures that always provide total orderings. In this chapter I highlight two kinds of incoherence for such measures in relation to Sen's characterization and representation of capabilitarian well-being. I first provide a critique of the misguided assumption that the Vector View mirrors Sen's characterization. In fact, Sen defends a 'settled' representation in terms of sets of co-possible sets (see below) that differs from the Vector View representation. I then explain how procedures that use the Vector View to always provide total orderings are incoherent with respect to further features of Sen's characterization – especially Sen's support for 'incomplete' cases for which there should be no fact of the matter regarding rankings of those items in terms of capabilitarian well-being.

[3] By 'different kinds of phenomena' I refer to objects that feature different kinds of qualities, in the same way that beds and houses feature different kinds of properties, or, similarly, in the same way that different elements on the periodic table feature different kinds of qualities with respect to pH, viscosity, reactivity, etc. Note also that even at this early juncture we can see how Sen's recognition of diverse kinds of perspectives/values is in keeping with the previous chapter's discussion of *phronêsis*. Diverse perspectives that are themselves determined by diverse kinds of experiences are, according to Aristotle, a necessary element for a deliberative body to make a practical decision.

[4] As I've explained in the previous chapter, I bracket the general question of whether *asummetria* or 'non-commensurability' truly holds for different kinds of well-being values.

Following the first challenge, I provide an introductory sketch of Sen's 'settled' representation of capabilities as the following:

> **Sen's Settled Representation**: Capabilities are represented as sets of the different co-possible sets of functionings that an individual, subject to constraints like skills and resources, has the opportunity to do or be and has reason to value.

I return to a more nuanced explication of Sen's representation in section three of this chapter. For now let's note that Sen's 'functionings' are states of doing and being such as voting, being properly nourished, pursuing an eighth-grade education, etc.[5] Given certain constraints, a person chooses among co-possible sets of such functionings, also often referred to as possible 'lives worth living.' Here the representation follows Sen's characterization of value-ladenness in that co-possible sets of functionings can be *evaluated* when, for example, a person is to be able to deliberate over them. Evaluation can occur both intrapersonally (within one individual) or interpersonally (across multiple individuals).[6] For example, one individual can evaluate *intrapersonally* the set of possible functionings that include her becoming a professional athlete as better or worse than an alternate set of possible functionings that include her becoming a nurse. Or a deliberative body may *interpersonally* evaluate sets of possible functionings available to citizens that include, say, opportunities to become well-paid midwives as better or worse than alternate sets that merely include opportunities to become non-compensated domestic caretakers, etc.

In either the intrapersonal or interpersonal case, Sen's representation is meant to capture the *holistic* evaluation of sets of co-possible functionings, taken to stand for the way that people evaluate entire lives worth living against one another. Most actual implementations of the Vector View representation for measuring capabilitarian well-being do not fit with this holistic emphasis of Sen's settled representation. Measures based on the Vector View representation typically exhibit a

[5] We get to the representation after a discussion of the economic and philosophical conversation surrounding Sen's characterization of capabilitarian well-being, including the recognition that different kinds of capabilities are valued in diverse ways – that they are heterogeneous. See section three for further analysis of the formal representation of capabilities as sets of sets of possible functionings.

[6] Sen holds that evaluation measures the level of freedom available to the individual, where freedom can take forms of agency or well-being freedom. I return to this subtlety in the next section and through the rest of the chapter as it directly bears on whether Sen's capabilitarian well-being can be measured by a single quality like well-being freedom, which could ensure total well-being orderings.

'multidimensional' set of proxy indicators, each standing for one or many different functionings or some common feature thereof, but not standing for co-possible sets of opportunities. This is a different emphasis of evaluation. So why, we might ask, is the Vector View so readily used by contemporary metrics alleging to follow Sen's characterization? As we shall see in section three of this chapter, I show that the Vector View representation *is* given limited support by Sen in his arguments against single-grade, utilitarian measures of well-being. However, Sen's support given within a utilitarian framework does not directly translate into support for the same representation to be used within Sen's capabilities framework.

Many contemporary measures of capabilitarian well-being, furthermore, evaluate the Vector View representation according to procedures that always provide total orderings, rendering the measure incoherent with respect to Sen's (or Nussbaum's) characterization and respective representations. I argue one likely source of this confused treatment is a misinterpretation of Sen's cautious support for practical comparisons, for example in situations in which policy makers compare and choose among different policies. Sen's recognition of the inevitability of practical decision-making comparisons might be misinterpreted as grounds for practitioners to incorporate procedures that always provide total orderings of capabilitarian well-being, and so I call this misinterpretation the Justification for Total Orderings on the Basis of Practical Comparison (JTOBPC), introduced in the previous chapter. By explaining how such total orderings may provide outputs for which there should be no fact of the matter in terms of capabilitarian well-being, I argue that such incoherent measures at least warrant further investigation.

Even when JTOBPC-type measures of capabilitarian well-being clearly do not cohere with Sen's (or Nussbaum's) representation and characterization, their application may still be deemed *useful enough* according to one or another practical purpose. For example, a Vector View representation may, when evaluated according to procedures that always provide total orderings, provide a rough estimate for tracking very large differences in capabilitarian well-being among populations. Or a Vector View may be the best of a limited set of available measurement options relative to some intended decision-making purpose. Or the total orderings may function as some very rough standard used for international

conversations around development, and so on.[7] But this short list of uses is exhausted by the actual set of applications to which the capabilities approach is currently lent, suggesting that practitioners who allege to follow Sen or Nussbaum's characterization need to better consider the significance of incoherence. Even in the above 'roughly useful' cases, incoherence at least warrants further investigation both to clarify the external reasons incorporated into the measure and to consider the significance of harms that can result if the measure is off-target, especially in cases for which there should be no fact of the matter in terms of capabilitarian well-being yet the measure produces a total ordering anyhow.[8]

[7] Regarding this last point, see section five of this chapter and, especially chapter three section five.
[8] See especially chapter four.

2.2 Sen's Characterization: Presuppositions Related to Value-Ladenness and Against Procedures that Always Provide Complete Rankings

In the previous section I gave a shorthand for Sen's characterization of capabilities that is in keeping with how authors typically refer to Sen's concept. For example, Ingrid Robeyns describes Sen's capabilities as a person's "real opportunities to do and be what they have reason to value," (Robeyns 2016: 1) while Lord Nicholas Stern similarly discusses "the capabilities and freedoms of individuals to live a life they have reason to value" (Stern 2007: 41). Though Sen develops his characterization through layers of philosophical discussion, he does repeatedly offer a shorthand of his concept as the capabilities "of persons to lead the kinds of lives they value" (Sen 1999: 18) or, more particularly, a person's capabilities "to achieve functionings that he or she has reason to value" (Sen 1992: 5). Each of these quotes presupposes that capabilities measure what people can actually, or 'substantively,' accomplish, though Sen's concept emphasizes the less directly measurable *opportunities* people have to accomplish these doings and beings, and not what is typically measured: the doings and beings that people actually achieve.

Each of these quotes also indicates that capabilities are value-laden inasmuch as people have 'reason to value' their opportunities to achieve different kinds of doings and beings. These quotes do not, however, directly state that capabilities are resistant to being measured according to procedures that always provide total orderings - what Sen calls 'complete rankings' - of capabilitarian well-being. Further pre-suppositions in Sen's work ground this commitment. These presuppositions support a reconstructed characterization of capabilities, given in the previous section as the following: *Capabilities are the extent of the freedom individuals have to achieve different kinds of doings and beings they have reason to value in diverse ways.* I note here, in keeping with the discussion of *asummetria* in the previous chapter, that the reconstructed characterization makes it clear that capabilities admit of different kinds such that they are resistant to being rendered as total orderings of one another.

Let's further unpack these presuppositions by considering a recent, less short-winded version of Sen's characterization.[9] In his (2005) keynote lecture (quote i), Sen characterizes the approach to measuring capabilities as the following:

> an intellectual discipline that gives a central role to the evaluation of a person's achievements and freedoms in terms of his or her actual ability to do the different things a person has reason to value doing or being. These achievements (like being well-nourished, avoiding premature mortality, taking part in the life of the community, and so on) are called human functionings, and we have to concentrate on those functionings which we *value and have reason to value*
> (Sen 2009b: 16, *emphasis added*).

The quote includes an important distinction regarding Sen' characterization of valuation: functionings are both valued *and* reasons can be given for why those functionings are valuable. Functionings can be interpreted as *valuable simpliciter*, as evidenced by related emotions, expressions, or other extra-linguistic indicators of value. And, of course, functionings can be interpreted as valuable according to expressions of reasons. There is some overlap here regarding these two modes of expression. Expressions can include reasonable arguments that translate into what might be called 'formal propositional content' of those arguments that can be expressed impartially among different people. Expressions can also include the surrounding set of gestures, utterances, or other extra-linguistic features or actions that supplement the formal content of the reasons themselves. These modes of expression are often given simultaneously by the same individual during processes of deliberation regarding the value of one or another set of capabilities.

An important implication of this distinction regarding modes of expression is that extra-linguistic features of expressions often are not captured by the purely formal, impartial content of expressions stated as reasons.[10] For example, a person might value one or many functionings inasmuch as the thought of them evokes certain emotions or further forms of valuation simpliciter: she might value, simpliciter, the feelings of happiness and tranquility associated with being able to spend more time with loved ones rather

[9] Sen's characterization of capabilities covers a writing span of nearly thirty years. In this chapter I focus on a number of watershed pieces from that period, but I generally focus on more recent written work that represents Sen's more settled view.
[10] So said John Dewey and Ernst Cassirer, among many others. See chapter three section four. For more on this point, see chapter three sections three and four regarding how values are recognized by prudent agents.

than working overtime. She can also, simultaneously, value those functionings inasmuch as she can give and assent to personal or even publicly expressed reasons that at least partially capture the way she values those functionings. For example, she may publicly express her reasons for why she or, more generally, anybody else should value time spent with loved ones, and such reasons could be re-expressed, impartially, by others. Moreover, her personal expression might agree with or be influenced by expressions given in surrounding social discussions or interpreted from surrounding social norms. Such reasons and norms might be expressed or otherwise influence deliberation carried out in community venues such as, for example, coffee shops, churches, bus stops, university classrooms, or through public media such as newspapers, radios, print books, televisions, social media accounts, etc.

Though this somewhat subtle distinction regarding modes of expression is only briefly addressed by Sen, we shall see in the following sections that the distinction is part of what grounds his pluralism regarding the kinds of reasons that can be given for how and why certain functionings are valued relative to one another. I bolster this point in chapter three when I discuss Nussbaum's characterization of value-ladenness and Elizabeth Anderson's pragmatic theory of value. At this introductory juncture, it is enough to note that Sen clearly differentiates 'valuing' simpliciter from 'having reason to value.'

The recognition that valuation simpliciter can be indicated by extra-linguistic expression is especially important in the context of measuring capabilitarian well-being, as it demarcates the practices by which groups of deliberators can publicly express different kinds of values. For example, a person might not be able to give a well-articulated reason to explain how she feels marginalized or oppressed relative to perceived opportunities for some functionings – say, for example, those that feature limitations significant occupational, educational, or domestic limitations -, yet she can at least *express* the negative values, simpliciter, as some set of acts that relays her perspective, perhaps on the level of shared emotion. Such expressions can, in turn, provide a basis for articulation as sets of reasons in a public setting by

supportive third parties, for example through consequentialist, virtue, or rights-based reasons, etc.[11] For instance, in her Justice and the Politics of Difference (2011), Iris Marion Young explains that when marginalized or oppressed individuals attempt to express their perspective and experiences, such attempts at explanation are often ignored or suppressed because 'irrational' or 'unreasonable' or 'crazy.' Young argues that even if others in the public domain label such expressions as irrational, unreasonable, or crazy, those expressions are undeniable indicators of value, and supportive third parties *can* help to turn such expressions into articulate, reasonable dialogue that can defensibly survive public scrutiny.

An important philosophical point here regards the difference between a mere 'expression' of value simpliciter versus a 'reasonable explanation' that can defensibly survive public scrutiny.[12] Sen explains that such sets of reasons 'survive' within community discussion when they are given more impartial forms such that they can be recognized and discussed among other individuals. For example, claims regarding the injustice found in typical experiences at the hands of border patrol officers might survive in public discourse as indications of oppression even when those who have experienced such injustice or are especially vulnerable to it are themselves afraid or otherwise hesitant to publicly articulate those claims.

Regarding the kinds of reasons that are recognized in public discourse, Sen repeatedly supports diverse forms of ethical and moral reasoning as indicators of value.[13] In his 2009 The Idea of Justice, for example, Sen argues that to be what he calls a 'defensible' indicator of value, a person's expression must survive what Sen calls 'critical scrutiny': it must be able to be articulated in terms of at least compelling, often impartial reasons that defend the value expressed (Sen 2009a: 199-201). Yet to be compelling does not mean that the set of reasons has to win some zero-sum game. Consider, for example, Sen's suggestion that altruistic reasons survive critical scrutiny even when they continue to be challenged by reasons that favor personal survival (ibid: 202). Sen's sense of critical scrutiny does not entail there is

[11] They are also the resonant grounds for further modes of expression. It is not merely the propositional content that travels, but also the relevant wave of emotion. I return to this idea as it is encompassed by Nussbaum's discussion of *prohairesis* in chapter three.
[12] I return to this particular point and how it relates to the measurement of capabilitarian well-being in chapter four.
[13] See for example Sen *The Idea of Justice* (2009: 194-207) and Sen *On Ethics and Economics* (1987b: 61-65).

only one accepted set of moral principles or arguments that 'wins out' in public discussion as the most defensible expression of value. Rather, Sen supports a pluralism of possible kinds of compelling reasons and related modes of expression.

So how does this pluralistic public expression of reasons – what Sen also calls 'deliberation' – work according to Sen? Let's consider a more complex example, one that I introduced during the discussion of *phronêsis* in the previous chapter. In his earlier Development as Freedom, Sen considers a challenging hypothetical case for a care worker, Annapurna, who wants to give a gardening job to one of three unemployed laborers: Dinu, Bishanno, or Rogini. Dinu is the poorest and so has a strong value claim; awarding the job to Dinu will satisfy the most economic need. Bishanno is the most depressed and also has a substantial value claim; awarding the job to Bishanno will create the most happiness. Rogini has a chronic ailment that could be relieved by further income; awarding the job to Rogini will create the most health (Sen 1999: 54).[14] Who to select for the job? The contrasting values in this example indicate the difficulty of finding some rational or even *reasonable* set of arguments that can defensibly indicate both how we value and how we might evaluate different kinds of well-being goods. Annapurna's consideration of the situations of Rogini, Bishanno, or Dinu evoke complex mixtures of different kinds of pleasures and emotions (happiness, health, pity, empathy, pride, sorrow and joy, to name a few) that, moreover, relate to further values like fairness, equity, or justice, all supported by reasons given according to concomitant systems of morality or ethics.[15] Moreover, different cultural lenses might affect how Annapurna feels and reasons about the different situations that Rogini, Bishanno and Dinu face. Consider, for example, the difference between the culture that might influence Annapurna's decision as a UN development worker versus the culture that might influence her as a local matriarch.

In light of this example, we can recognize that deliberators acting according to Sen's process of critical scrutiny would at least need to be sensitive to diverse sets of coherent, impartial reasons that

[14] Sen uses the Annapurna example to show that different values become entangled when we attempt to value functionings through processes of evaluation. I will return to this point about entanglement shortly as it relates to how groups of capabilities are evaluated.
[15] I note here, following the discussion from section one of the previous chapter, that I consider expressions of emotion to be expressions of value, following Ruth Chang's 'ordinary' definition of values.

Annapurna - or some similar public interpreter - could plausibly give regarding how to value the gardening job, just as deliberators would have to allow for the possibly less articulate but certainly compelling expressions that Rogini, Bishanno, or Dinu might give regarding how the job is valuable. The example illustrates how Sen's pluralism gives a wide scope to what is meant both by valuing simpliciter and what people have reason to value. Annapurna or any of the laborers could, for example, give an argument that emphasizes values of fairness and efficiency related to greater relative wealth (perhaps favoring Dinu), or could give a more consequentialist argument that emphasizes maximization of total pleasure (favoring Bishanno), and so on. Note also that Annapurna might give such sets of reasons as an attempt to explain what she interprets from the expressions of value given by the laborers, again as influenced by surrounding cultural values and norms.[16]

From this discussion of valuing and having reasons to value we are introduced to the importance of public deliberation presupposed by Sen's characterization. In his (2005) keynote (quote ii), Sen argues that

> the necessity of scrutiny and critical assessment is not just a demand for every person's self-centred evaluation, but an *insistence on* public discussion and interactive public reasoning. This is because social evaluation is not just a matter of separated and sequestered individual reasoning. Public discussion and deliberation can lead to a better understanding of the role, reach and significance of particular capabilities (Sen 2009b: 18, *emphasis added*).

As previous examples have shown, capabilities are value-laden not only inasmuch as individuals can express or give reasons, but also as such expressions are interpreted and re-expressed by others within the process of public deliberation. The earlier examples of a woman expressing how she values more time spent with loved ones shows that surrounding social norms and moral or ethical arguments can influence expression and re-expression. The same holds for the Annapurna example we just discussed.

[16] I return to Sen's pluralism regarding values and their connectedness to public processes of selecting and measuring functionings later in this section and thesis, as both Martha Nussbaum and Elizabeth Anderson give reasons and emotions (Anderson calls their combination 'expressive reasons') central roles as indicators of value. I note here, as a preface, that support for the expressive role given to both emotions and reasons dovetails with the argument that capabilities are of different, or 'qualitatively distinct' kinds. See in particular the remainder of this section, chapter three sections two and three, and chapter four sections three and five. The term 'qualitative distinctiveness' is, as mentioned in the previous chapter, taken from Sen and Nussbaum Wider Paper 30 (1987).

Yet in a more technical sense, capabilities can be valued according to how groups of individuals *deliberate* about them in public settings, for example on community panels, in town-hall meetings, or through further political actions such as voting, posting on social media, protest marches, etc. For Sen, multiple forms of reasoning and expression become indicators of value when they can defensively survive critical scrutiny in the public setting, and so the public setting must be one that best allows for critical scrutiny to occur. The goal of this thesis is not to defend exact forms of expression or deliberation that can survive critical scrutiny, but it is important to Sen's concept that valuation is best recognized through public deliberation and reasoning, and, as we shall see, when different kinds of capabilitarian well-being values are recognized as distinct, what Sen calls 'heterogeneous.'[17]

Thus far we have determined that Sen's capabilities are intrinsically value-laden inasmuch as individuals or deliberating groups (or both) give expressions, including reasons, as indications of the values constitutive of or relevant to different kinds of doings and beings that constitute various lives worth living. The earlier keynote characterization quote (2005-ii) also gives a central role to 'social evaluation,' indicating some *process* by which people's values and reasons influence the measurement of opportunities to achieve various sets of functionings. Following the discussion above, this process can be one of internal reflection for some individual, but also can be – and, according to Sen, should be - a *social process* of public deliberation that allows for critical scrutiny. I explain how evaluation is a social process further below, as Sen's characterization of this process informs how features of value-ladenness and heterogeneity can coherently be represented or operationalized when practitioners attempt to measure capabilitarian well-being.

One further presupposition relevant for Sen's characterization regards how the process of evaluation measures the extent of one's *freedom* to achieve valuable functionings – also called one's 'agency.' Sen repeatedly claims, for example, that it matters for measuring capabilitarian well-being whether a person is free to choose to fast or is forced to starve. Accordingly, the measure should not

[17] I return to a discussion of the importance of deliberation for the process of *measuring* capabilitarian well-being at the end of chapter four.

focus only on the available outcome (lack of nutrition) but also on how the consequence is achieved (free choice or not). Similarly, it matters to what extent one really is *free* to actively pursue the functioning. Consider, for example, the particularly acute expressions of value that may accompany partial loss of agency in the context of pursuing one's domestic partner or religion as opposed to those that generally accompany partial loss of agency in the context of pursuing extra compensation, say by working extra hours at a job.

Following this example, let's take a moment to get a deeper sense of what Sen means by evaluating one's extent of freedom to actually achieve valuable functionings. Sen recognizes multiple aspects of freedom in his characterization of capabilities: *opportunity freedom*, *procedural freedom* (also called the *process* aspect of freedom), *agency freedom*, and *well-being freedom*. These aspects work for the most part as contrastive pairs so I'll summarize them accordingly. The opportunity aspect of freedom has a more consequentialist orientation in that it focuses on "the actual opportunities a person has," while the process aspect "deals with having arrangements and institutions that can be seen as components of procedural freedom" (ibid: 27). As this latter definition is somewhat circular, let's clarify by example. A person might have opportunities to achieve certain groups of functionings, but the *process* by which those functionings become available might seriously inhibit her 'procedural' freedom. She might, for example, find herself compelled to 'fast' because surrounding social norms enforce the notion that it is the role of a virtuous mother to give more food to her family, or she might fast because she really, freely chooses the action as a religious practice that brings her closer to her faith. Both cases of 'fasting' feature similar groups of opportunities related to lack of nourishment, but compellingly different surrounding norms and related procedures. In the former case, for example, social norms might determine, as a subtle, pervasive form of coercion, procedures that proportion the allotment of food for 'virtuous' female members of the household.

In a similar example, Sen explains that a society might seek to provide equal opportunities for health, and so people distribute more healthcare to men who on average live shorter lives than do women

(ibid: 27). Here the 'outcome' aspect of freedom might support a gendered way of valuing health that overlooks values related to institutional fairness or equality, while the process aspect of freedom includes these further values in favor of a different distribution of resources. What Sen indicates by making the distinction between the opportunity and the process aspect of freedom is that evaluation of opportunities for functionings should be sensitive to both outcomes and processes; such sensitivity entails that relevant values may provide a basis with which to support or challenge surrounding norms and procedures. For example, closer scrutiny given to the 'virtuous' fasting of mothers or wives might allow for the recognition and influence in deliberation of negative values associated with historically oppressive domestic roles.

Sen also makes a helpful distinction between aspects of agency freedom and well-being freedom. Agency freedom features a more personal emphasis in that it "encompasses all the goals a person has reasons to adopt, which can *inter alia* include goals other than the advancement of his or her well-being" (Sen 2009a: 287, original emphasis). Rather than attempt a full definition of personal agency, I simply note here, following Sen, that agency freedom includes values related to goals that follow personal judgments and priorities - what might be called one's 'personal meaning.' For example, an agent might attach more personal meaning to the set of possible functionings that favors her becoming a successful firefighter or corporate lawyer rather than to the group that favors her having to accept caretaking work along with, perhaps, a greater likelihood of spending more time at home raising her children. This example indicates that agency freedom reinforces the process aspect of freedom; it can matter substantially whether opportunities for functionings include procedures that limit social roles or distribution of resources. Consideration of agency freedom along with process freedom might, for example, support public recognition of the value of legally-enforced institutional procedures that provide extended pregnancy leaves at full pay, and so on.

As mentioned above, the evaluation of capabilities is not typically an exercise of personal reflection regarding, inter alia, the extent of one's agency. Many capabilities measurement systems -

especially those critiqued by this thesis - are used for community-level assessments of entire populations. Such assessments tend to invoke 'well-being freedom,' a less agent-oriented aspect of freedom that promotes increased freedom of individuals-within-communities. Well-being freedom includes some aspects of agency freedom (it certainly matters whether, for example, an individual is free to pursue higher education or further personally-meaningful vocational opportunities) but also can include a more holistic orientation towards initiatives that better serve community-recognized forms of well-being rather than agency only. It may, for example, be more relevant to the well-being of the community to create a homeless shelter rather than to build a sports park, even if many agents would claim significant advances for their agency freedom were the latter option pursued. Sen explains, for example, that "in determining the extent to which a person is deprived in a way that calls for community assistance from others or for the state, a person's well-being [freedom] may be, arguably, more relevant than his or her agency [freedom]" (ibid: 288). Broadly speaking, the freedom to pursue well-being may be seen through a collective or a community lens that contrasts with or even diminishes valuation according to individual agency freedom.

Consideration of well-being freedom may, for example, lead a community to value increased access to food, education, and healthcare over increased funding for the pursuit of personally meaningful projects like one's erecting a monument to a personal hero (ibid: 288). Here Sen's example of a meaningful project might seem a bit abstract, but it emphasizes that what individuals take to be personally meaningful will, at least in some cases, clash with what the greater community recognizes as constitutive of well-being. This clash is important to the capabilities approach. In some cases, it might be that parochial values held by the majority, for example gendered norms regarding acceptable gender roles, support 'community-level' valuations of well-being that are challenged by valuations favoring agency or other procedural senses of freedom. Similarly, an individual may express and give reasons for the value of publicly sharing her political or religious opinions, and such expressions may be supported by communities even when they challenge dominant well-being freedom valuations within that community.

In such cases it matters how the process of evaluation recognizes and allows for the defense of particular values within forums of public deliberation. Such examples of the complex interplay among aspects of freedom illustrates how challenging it can be for a social process of evaluation to capture the relevant valuations. I return to this point shortly, first taking a moment to consider a further aspect of freedom presupposed by Sen's characterization.

According to Sen, one's 'substantive freedom' emphasizes whether or not one can *actually achieve* different functionings (Sen 1999: 75). Substantive freedom arguably subsumes other aspects of freedom. It can, following an earlier example, capture the procedural freedom difference between an individual who fasts and one who is forced to fast, given that "the first *can* choose to eat well and be nourished in a way the second cannot" (ibid, original emphasis). Sen repeatedly emphasizes, in various works, substantive freedom as a key feature of capabilities. This emphasis provides the opportunity for practitioners to misinterpret substantive freedom as 'the' central feature of capabilitarian well-being, perhaps one that can be represented as a single grade or that can be captured by procedures that always provide total orderings of capabilitarian well-being. As I explain in section five of this chapter and again in the next chapter, not only Sen's but also Nussbaum's characterizations do not support such a misinterpretation.

For now, I argue that while an emphasis on measuring substantive freedom might allow for the capture of some values related to procedural, agency, or well-being aspects of freedom, it cannot capture all relevant aspects of freedom. Let's first go back to the forced fasting example. Yes, the focus on substantive freedom may capture differences in opportunities available to the woman who faces social pressure to 'fast' as opposed to those available to the woman who chooses to fast to get closer to her faith. Yet in such cases the concept of substantive freedom overemphasizes the opportunity aspect of freedom rather than the procedural. For example, substantive freedom does not capture further relevant values like justice or fairness that might be used to highlight institutions of oppression within individual households; public recognition of such further values might influence how items are evaluated, for example favoring a

policy that emphasizes education and consciousness raising in conjunction with food distribution rather than allowing that dominant, parochial norms determine the distribution of food within individual households.

Similarly, consider how 'the' central focus on substantive freedom does not provide a sufficient means for Annapurna to distinguish why she might care more for Bishanno's depression over Rogini's chronic illness, or vice versa. Not only do different options support different kinds of doings and beings that may not be captured in terms of their relative 'substantive freedom,' but also different kinds of values relevant to each option may not be not fully captured in similar terms. On Sen's view, substantive freedom is another aspect of freedom, a category that better helps practitioners to follow the subtler presuppositions embedded in Sen's characterization. Substantive freedom is not, for Sen, a foundational emphasis in that sense that it could somehow capture all other relevant aspects of freedom, much in the same way that a single sense of 'the good,' for Aristotle, is not foundational in that it does not readily capture different kinds of well-being virtues.[18]

Let's return to the topic of how Sen sees evaluation as a social process through which people attempt to capture the complex interactions of different kinds of values relevant for measuring capabilitarian well-being. Previous examples might seem to suggest functionings can be evaluated against one another as isolated comparisons, perhaps according to the something like the 'level of opportunity' that can be measured for each. However, an important nuance of Sen's characterization is that entire sets of opportunities for functionings must be evaluated holistically. For example, it would be a misinterpretation of Sen's characterization to argue that when measuring capabilities, one's level of nourishment can be evaluated in isolation against one's level of maintaining important personal relationships within a community. These and other functionings would be evaluated together, with an emphasis not only on how much opportunity an individual has to actually achieve different combinations of those functionings but also the diverse ways in which those groups of functionings are valued. For

[18] See previous chapter.

example, an individual's ability to achieve a certain level of nourishment would be considered along with other functionings such as available vocational paths, level of happiness, feelings of safety or belonging in public, etc. A complex mix of values will determine, both for the sequestered individual and the deliberative group, how to evaluate the entire functioning set.[19] This characterization point about holistic evaluation is further discussed in the next section as it coheres with Sen's representational focus on sets of possible sets and *not* the incoherent representation of a Vector View of different capabilities indicators.

At this juncture all relevant presuppositions for how Sen characterizes value-ladenness have been considered. We can now recognize that Sen's capabilities are meant to be the extent of the freedom people have to achieve valuable doings and beings, and that capabilities are intrinsically value-laden inasmuch as they are indicated by personal and public expressions, including reasons. Furthermore, evaluation is, on Sen's view, a social process that considers various aspects of freedom and other values relevant to co-possible sets of opportunities for functionings, considered holistically.

Let's now pivot to how Sen characterizes capabilities as 'heterogenous' such that they cannot generally be measured by procedures that always create total orderings of capabilitarian well-being. As noted earlier, shorthand characterizations of Sen's characterization do not explicitly include this intrinsic feature, though the above keynote characterization does helpfully suggest that functionings are (2005-i) 'different' and (2005-ii) 'particular;' they are of *different kinds* that are valued in *diverse ways*. Much of what is meant by these two italicized phrases has already been introduced by the foregoing examples, and so I return to those examples in defending the following three claims regarding Sen's characterization: (1) doings and beings admit of different kinds, (2) doings and beings are recognized and evaluated according to different kinds of values, and (3) evaluation as a pluralistic, social process includes diverse ways of valuing.

[19] I return to this point as it relates to Elizabeth Anderson's practical defense of keeping different kinds of values distinct during deliberation regarding those values in the second half of the next chapter.

The first claim is almost intuitive but worth a moment's reflection. Sen typically characterizes functionings by providing the kinds of general examples given in the (2005-i) keynote characterization above. Following these examples, we can recognize, for example, that functionings related to nourishment, health, and community participation are of different kinds such that – following the discussion of *asummetria* in the previous chapter – what is good about them is resistant to being rendered in terms of one another, much as it would be difficult to measure the pH of fluorine in terms of the viscosity of mercury. In particular, as we saw with Aristotle's discussion exchange value, total orderings of items that exhibit *asummetria* of different kinds of qualities will often fail to capture those qualities.

To be fair, some functionings could arguably be defined according to some simpler, single-grade basis that might capture different kinds of doings and beings. Consider, for example, abstractly described functionings like 'feeling happy,' 'finding meaning,' 'feeling satisfied,' or other similarly vague subjective states of being. Recall also the discussion in the previous chapter of *chreia* as a plausible single-grade basis for measuring the value of different kinds of economic objects. On such a basis, one could argue that some homogenous, subjective 'goodness' underlies all possible values as they apply to different possible sets of functionings. One could even argue that those abstract states are actually 'root-level functionings' that can be used to measure all other possible sets of functionings. For example, the set of functionings that one experiences in a given week could, perhaps, be roughly captured by a certain feeling of 'meaningfulness' or 'happiness' that could be compared against similar feelings for other sets of functionings, say the those experienced in the 'less meaningful' prior week. As Sen notes, such arguments are in keeping with similar utilitarian claims that all forms of value can be measured in terms of a single grade of something like happiness, desire, or preference satisfaction (Sen 1999:64).[20]

Yet as I discussed in the previous chapter, Sen follows authors like John Stuart Mill, Marx, and Aristotle to argue that most functionings are of different, 'heterogeneous' kinds.[21] As Sen explains in his article on "Plural Utility," pleasures "can be distinguished in many different ways" and resist being

[20] See the discussion in the previous chapter for further information on single grade metrics.
[21] See chapter one section three.

measured according to a common unit of utilitarian pleasure (Sen 1980: 194). The pleasure that one might feel while eating an éclair or training for a marathon, for example, simply is of a *different kind* from the kind of pleasure one might feel while singing a hymn in church, even if both can be understood according to some vague feeling of 'pleasure.' Sen goes on to argue that different kinds of pleasure, and by extension other 'root-level' values relevant to or constitutive of different kinds of capabilities, resist being rendered in a total ordering.

The second claim (2) is partially justified by the support given for the first. Recall from the previous chapter that *values* taken in Ruth Chang's 'ordinary' sense include action-guiding pleasures and desires, which Sen, following Mill, Marx, Ricardo, and Aristotle, takes to be of different kinds. Recall from chapter one that values in this 'ordinary' sense also include virtues, ideals and norms. Inasmuch as they are action-guiding or can be translated into sets of reasons that would survive Sen's critical scrutiny, virtues, ideals and norms are of different kinds in a manner similar to how pleasures and desires are of different kinds. For example, in contexts of public deliberation, 'fairness' concerns a different set of expressions and action-guiding principles than does 'justice,' even if there may be significant overlap between such sets. Examples given earlier in this section further highlight how different kinds of values related to freedom, justice, or well-being become relevant to the holistic evaluation of different possible sets of functionings, for example how they relate to procedural freedom, well-being freedom, or in Annapurna's case the diverse experiences of Dinu, Bishanno and Rogini. The discussion of *phronêsis* in the previous chapter provides an outline for the conditions that enable this kind of analysis, given that those who involve *phronêsis* in their decision-making processes evaluate different options according to the recognition of distinct kinds of values.

The third claim (3) follows from Sen's support for pluralism and public deliberation as relevant presuppositions regarding value-ladenness of capabilities. Returning to the Annapurna example, we can see that Dinu, Bishanno and Rogini have different ways of valuing the gardening job and related functionings. It might be possible for these valuations to be described by sets of impartial reasons that

survive critical scrutiny within public deliberation, but Sen notes, in keeping with the previous chapter's discussion of the importance of experience, that the *perspective* of the individual also matters. Sen claims, for example, that although "much can be done through the deliberate use of open impartiality, the hope of proceeding smoothly from positional views to an ultimate 'view from nowhere' cannot hope to succeed fully" (Sen 2009a: 169). Impartial reasons are an insufficient basis for recognizing and evaluating values expressed from particular perspectives, especially when the sources of those perspectives can give their own expressions in public. For example, Annapurna or some supportive interpreter may be able to articulate some set of reasons that indicates how Rogini values the job as a means to fix a chronic ailment, but the actual *value* of reducing Rogini's stoically-borne pain is best expressed by Rogini. Similar arguments apply in support of that claim that people – especially those who are stakeholders for various policy initiatives - express values in diverse ways that depend on the position of the expressive agent.

Moreover, even the most articulate expression of reasons might not sufficiently capture relevant values when such reasons are expressed impartially, even by a well-meaning ally within a context of public deliberation. For example, the experience of being a black man walking down the street in a white-dominant neighborhood, eloquently related by Frantz Fanon's Black Skin White Masks, necessarily involves a complicated phenomenology that cannot really be impartially summarized by the even a well-meaning, 'colorblind' observer who has not had the same experience:

> I turned beseechingly to others. Their attention was a liberation, running over my body suddenly abraded into nonbeing, endowing me once more with an agility that I had thought lost, and by taking me out of the world, restoring me to it. But just as I reached the other side, I stumbled, and the movements, the attitudes, the glances of the other fixed me there, in the sense in which a chemical solution is fixed by a dye. I was indignant; I demanded an explanation. Nothing happened. I burst apart. Now the fragments have been put together by another self.
> (Fanon 1952: 82).

What would it mean to attempt to put the particular perspective given by Fanon's description into some impartial though politically functional category, for example 'identity threat' or - perhaps worse - a

single-grade 'lack of happiness?' The impartial perspective would not capture some of the valuation indicated by Fanon's expression were it not expressed by someone who at least has really *experienced* the position of receiving a racially-discriminatory gaze, along with related injustices. The best that we can do, as interpreters, is to imagine that particular experience and attempt to feel and understand the relevant set of emotions and values. Note also that this example of Fanon's particular experience helps to illustrate a broader point that particular, even if abstruse, perspectives are value-laden such that the *act of their expression* influences public deliberation as valuation simpliciter. Accordingly, the same set of functionings takes on diverse forms of meaning, and so is evaluated differently, depending not only on the experience of the expressive agent, but also the way in which the agent is given voice within the relevant deliberative forum.

For the sake of philosophical rigor, let's consider a challenge from Plato's Socrates, introduced in the previous chapter, who insists that, no, even particular perspectives *can* be measured in terms of one another because there is some other homogenous stuff that, deep down, indicates the true level of 'well-being' of *every* particular functioning. Such a challenge could be given in support of something like 'root-level functionings' considered above, or, in utilitarian terms, a system like Von-Neumann and Morgenstern's cardinal utility function, which represents preferences for different options according to procedures that always give a total ordering of items. Such a challenge could, of course, also be used to defend a similar measurement system that always provides total orderings of capabilitarian well-being. Yet the Platonic challenge contradicts Sen's characterization. Sen repeatedly insists that different kinds of capabilities are, in fact diverse in the sense that they are 'non-commensurable':

> capabilities are diverse, indeed they *must* be since they deal with different features of our life and our freedom. This is, of course, a most unremarkable fact, but there is such a long tradition in parts of economics and political philosophy of treating one allegedly homogenous feature (such as income or utility) as the sole 'good thing' that could be effortlessly maximized (the more the merrier), that there is some nervousness in facing a problem of valuation involving heterogeneous objects … Capabilities *are clearly non-commensurable* since they are irreducibly diverse
> (Sen 2009a: 239-240, emphasis added).

In numerous earlier works, Sen posits that capabilities are 'heterogeneous,' again referencing this notion that they are 'irreducibly diverse,' also what Sen and Nussbaum (1987) call 'qualitatively distinct' as introduced in the previous chapter. The Platonic position simply does not accord with Sen and Nussbaum's concept.

Sen gives further arguments in support of the diversity of capabilities that start to establish the commitment to avoid procedures that always provide total orderings of capabilitarian well-being. Let's look to one of Sen's watershed articles, his 1979 Tanner Lecture "Equality of What." The main thrust of Sen's argument in this article is that the process that uses weighted comparisons of intensities of pleasure or desire to create total orderings – what Sen calls 'complete rankings' – of utilitarian well-being is problematic in that it cannot reflect both the diversity of human needs and the diversity of other important moral demands, especially those related to equality.[22] I will show that Sen's argument further supports claims (2) and (3) above.

Regarding (2), Sen considers how a cripple (let's proceed with the contemporary term 'disabled person') who has a different set of needs than some 'average' citizen is mis-measured by a single-grade utilitarian scale. The disabled person gets less overall utility from a given amount of income than does her healthier average counterpart and so, on the basis of total intensity of desire or pleasure measured across individuals for the respective state of affairs, she is denied that distribution of income because it could be put to a more 'effective' use in terms of overall utility. Further constraints, in particular, affect the disabled person's ability to derive pleasure or satisfaction from resources relative to other persons. For example, a disabled person might simply consume a lot more resources moving from A to B than would an average citizen.

But then it might be possible for the defender of a metric that measures total intensity of utilitarian pleasure or desire to update the measure to better take individual needs into account. For example, the disabled person could have a better claim for access to resources if further egalitarian

[22] This argument is repeated in many of Sen's writings. See, for example, Development as Freedom (1999) 55-85.

principles that rank states of affairs according to bases other than total utility maximization were added to the measure. One such principle is Sen's *leximin*, according to which one would rank different states of affairs according to the utility level of the least well-off individual in each state (or, in the case of a tie, according to the utility level of the second least well-off, etc). Given the implementation of the leximin for measuring total utility across individuals, the disabled person's low utility level relative to others would become the basis for awarding extra resources or assistance.[23] The single metric defender could then claim that a single-grade metric *can* capture the deep-down well-being value of one or another set of functionings; the measure just needs further sets of principles that reflect diverse needs. Application of further principles like the leximin might, mutatis mutandis, allow the measure to even reflect the diversity of other moral demands related to more complex values like fairness, justice, human rights, etc.

Sen argues, however, that single metric intensity measures of total pleasure or desire prove insufficient on account of *adaptive attitudes*, favoring (3) above. Sen (1979b) has us consider the plausible case of a disabled person who has an adaptive attitude such that her level of subjective well-being, perhaps as observed through survey or interview data, indicates that she is doing 'well enough' in spite of the obvious challenges that she faces on account of her disability. Such an individual would not be given extra assistance even if the leximin were operating; her utility measure, based on her adaptive view, would look the same as, or perhaps even better than, that of her healthy counterpart (ibid: 217). The example illustrates that because they have a subjective basis, complete rankings of total pleasure or desire are unable to reflect well-being according to further moral demands that can be used to challenge adaptive attitudes. Taken together, Sen's disabled person examples indicate that single grade measures of total pleasure or desire are problematic because *different kinds* of moral demands force us to recognize – according to common-sense intuitions - that the measurement outcome is not tracking what people generally mean by 'well-being.'

[23] Note that these examples assume that the measure is responsible for the fine-tuned judgement, ie the allocation of resources. Sen's later writings such as the quotations from his (2009) The Idea of Justice back away from the claim that the measure is singularly responsible for the complex decision, for example the allocation of resources. In the case of these examples from 1979, it might be more charitable to note that the measure provides an informational basis, and this informational basis can help to frame the subsequent policy decision. Sen seems to defend this more subtle framing perspective in Development as Freedom (1999: 54-86).

Recall again the examples given earlier in this section. Were procedures that always provide total orderings of capabilitarian well-being to be implemented for such cases, relevant aspects of freedom or other well-being values might fail to be recognized not only because of procedural coercion from surrounding social norms, but also according to the internalized, adaptive attitudes of affected individuals. Consider, for example, the woman who fasts because she considers it a virtue to sacrifice for her family, and so does not report low levels of agency satisfaction, or even much unhappiness. Similarly, Rogini might stoically claim that her chronic ailment is really unimportant, that she is actually quite happy and doesn't really need the extra income. One goal of a capabilities metric that focuses on opportunities to achieve valuable functionings is to challenge these kinds of adaptive perspectives. Just as with policies that make sure food aid is fairly distributed within a household might provide greater capabilitarian well-being, Annapurna may greatly value awarding the gardening job to Rogini, despite Rogini's adaptive attitude, on the justification that it provides, on the whole, a better life worth living that it does for the other laborers.

These examples illustrate, as does Sen's example of the disabled person, Sen's commitment to the view that relevant well-being values cannot be captured by single-grade subjective metrics of relative intensities of pleasure or desire. This recognition supports the above claims that for Sen, (2) doings and beings are recognized and evaluated through different kinds of values (i.e. diverse moral demands), and (3) evaluation as a pluralistic, social process necessarily includes diverse ways of valuing (i.e. that of Rogini versus that of Annapurna).

2.3 Sen's Representation of Sets of Co-Possible Sets, The Vector View, and Incompleteness

In the previous section I explained how, regarding the commitment to the value-ladenness of capabilities, Sen emphasizes a complex, social process of evaluation. The process must, furthermore, respect Sen's commitment that different kinds of capabilities are resistant to being rendered into total ordering. Yet here there seems to have been some misinterpretation of Sen's characterization by practitioners working within the capabilities approach. Many contemporary measures that cite Sen's work represent capabilitarian well-being according to a vector of separate indicators - for example the vector of Human Development Index (HDI) indicators related to health, education level, and income. Such measures also tend to incorporate procedures that always provide total orderings - what Sen calls 'complete rankings' - of capabilitarian well-being for different populations. One commonly used set of such procedures includes those that weight and sum numbers assigned to each item in the vector. Such measures are incoherent with respect to Sen's characterization and settled representation, indicating that the measure is in danger of being off-target.

Practitioners might, as mentioned in earlier sections, justify such practices by arguing that similar evaluative procedures are often used in decision-making processes, so why not incorporate them into the context of measuring capabilitarian well-being, at least as a proxy for the central process of evaluation that Sen describes in his characterization? Recall that Sen, after all, contributed to the development of the HDI, which incorporates an incoherent set of procedures to produce complete rankings of capabilitarian well-being. Given that Sen contributed to the development of a measure that is inconsistent with a core commitment presupposed by his own characterization, perhaps one could take a JTOBPC-type position[24] that heterogeneity of different kinds of capabilities really is a trivial feature, one that is easily transgressed, without significant consequences, in favor of procedures that always provide total orderings?

[24] Justification for Total Orderings on the Basis of Practical Comparisons. See previous sections.

Questions like these have some traction, first of all because the vector representation used by the HDI and other similar measures might seem like a good-faith attempt to respect Sen's commitment to the heterogeneity of different kinds of capabilities. Such measures appear to treat different kinds of indicators - themselves proxies for different kinds of valuable functionings - as distinct, at least up to the point of incorporating procedures that always provide complete rankings. Because they start with a representation that holds different kinds of capabilities indicators distinct, such measures might, perhaps, provide a rough if incoherent representation of the holistic process of evaluation central to Sen's characterization. I return to this point later in the section.

Secondly, as noted in previous sections, Sen did defend what he calls a 'Vector View' representation of well-being. However, Sen's Vector View representation does not actually cohere with Sen's settled representation of capabilitarian well-being. This is because, firstly, Sen's earlier defense of the Vector View is given as part of his critique of the utilitarian approach to measuring well-being, an approach that differs from the capabilitarian approach. For example, there just is a significant conceptual difference regarding the amount of pleasure or desire that can be recognized for different kinds of well-being objects as opposed to the substantive freedom and well-being freedom, among other freedoms, that can be recognized as constitutive to or relevant for such objects. Secondly, even if we could assume that Sen's critique of the utilitarian approach provides a sufficient basis for how practitioners should represent capabilitarian well-being, Sen gives further reasons for why procedures that always provide complete rankings remain incoherent when applied to such a basis. I focus on Sen's watershed (1980) article "Plural Utility" to explain how even in the limited context of supporting the Vector View as an improved basis for measuring well-being over a utilitarian account, Sen limits the role procedures that render different kinds of indicators into complete rankings.

We shall also see in "Plural Utility" that Sen supports a related point, which is that when sets of co-possible sets of functionings are evaluated against one another as rankings, the measurement outcome

is likely to be what Sen later calls 'incomplete.'[25] According to Sen, the social process of evaluation tends to incorporate ranges of assigned value. Rankings determined on the basis of ranged assignments can be partial such that the ranking is incomplete (Sen 1992: 46 – 49). For example, a ranged value of .5 to .7 assigned for one indicator will not clearly rank above or below a ranged value of .3 to .6 assigned for some other indicator - there may be no fact of the matter regarding how to assign the ranking in such a case. By contrast, an assignment of .5 does always rank above one of .4, so long as these numbers are assigned along the same scale of measure. Sen's defense of incompleteness due to ranged assignments is further supported by features of pluralism and the heterogeneity of capabilities pre-supposed in his characterization. In this section I clarify these supporting links. I furthermore argue that it is a serious misinterpretation of Sen to treat his discussion of limitations associated with ranged assignments as support for general use of procedures that make assignments in a way that allows that complete rankings of capabilitarian well-being can always be provided.

Let's continue by considering Sen's 'settled' representation of capabilities as sets of co-possible sets, after which I contrast Sen's settled representation with the earlier Vector View representation. Sen represents capabilities as the following: *different kinds of co-possible sets of functionings that an individual, subject to constraints like skills and resources, has the opportunity to do or be*. For example, let's allow that individual *k* can, subject to certain resource constraints, attain different co-possible combinations - what Sen (1999) calls 'culmination outcomes' – of education, political involvement, a career, social affiliation, and leisure time to pursue subjectively meaningful projects.[26] The sets are co-possible in that different choices for *k* support different combinations of functionings. For example, *k* might choose to pursue less leisure in favor of a different co-possible set of functionings that includes more political involvement or education, etc. Different culmination outcomes in this example would be represented as co-possible functioning sets within the overall capability set for *k*. Note that the

[25] See, as one example among many, Sen's account of 'incompleteness' in *Inequality Reexamined.*
[26] See Sen 1999: 27 or Sen 2009a: 22-23, 215. It's worth noting there is a difference for Sen between the more consequentialist 'culmination outcome' and the more process-oriented 'comprehensive outcome.' Sen's favor for the latter provides further support for the argument that even a weighted vector view representation is unlikely to fully capture the relevant set of values that feature in the process of social evaluation.

representation of capabilities as sets of co-possible sets of attainable functionings also captures limitations related to resources or the surrounding social structure: middle-class *k* might be able to either travel the world when young or look after her children while they are little, but likely not both during the same time period. The set representing the combination of all these co-possible functioning sets is the formal representation of individual *k*'s capability set.

One might wonder where in this settled representation is the social process of evaluation? Sen's settled representation is taken to indicate, among other values, aspects of freedom including the substantive freedom that people have to achieve different co-possible sets. A person who has more choices and can simultaneously attain greater levels of different kinds of functionings than another person has greater substantive freedom (among other aspects of freedom), and this recognition affects the evaluation of the relevant sets of co-possible sets.[27] Evaluation, however, is not a straightforward process on Sen's view. Among other methods discussed by Sen, evaluation could, for example, focus on the following: comparing the number of different functioning options a person has, comparing the value of the most important functioning within one or another co-possible set, comparing the negative value of the worst option in the relevant sets, considering the level of constraints imposed by the process through which the person is able to achieve valuable functionings, or on some combination of the former methods, etc.

Sen's recognition of a wide array of options does not, however, entail that any method that can be formally postulated is sufficient to the task of evaluating capabilitarian well-being. Though Sen considers different possible options for evaluation, he also gives significant restrictions on how evaluation should proceed. One key limitation is that evaluation of sets of lives worth living in terms of their capabilitarian well-being is resistant to isolated comparisons of functionings, for example as if such functionings or their concomitant sets of values could be traded off in terms of one another. As explained in earlier sections, functionings and the values constitutive of or relevant to them are heterogeneous in that they are

[27] See the discussion of 'substantive freedom' in the previous section.

resistant to being rendered in terms of one another, especially when we recall that their evaluation respects their holistic entanglement within sets of co-possible sets of functionings. Consider, for example, that adding a given functioning to one co-possible set may increase that set's value, while adding it to another may decrease the value of that set. From such an example can start to see how it is problematic for practitioners to start from a Vector View representation with different kinds of indicators in isolation and then implement procedures that assign numbers to each indicator, weight them, and then combine them in order to always provide a complete ranking. Such procedures problematically treat different kinds of capabilities indicators as if they can be compared or even traded off against one another in isolation.

And yet, there is, within Sen's settled representation, support for the possibility for *complete ranking* outputs in the special case of what Sen calls 'complete set dominance.' For example, if *k*'s set contains sets of functionings $\{\{a_1\},\{a_2\},\{a_3\}\}$, where $\{a_i\}$ represents a specific culmination outcome of possible functionings, while *l*'s set contains sets of functionings $\{\{a_1\},\{a_2\}\}$, then *k*'s set dominates *l*'s, and *k*'s capability set clearly exhibits more substantive freedom than *l*'s such that *k*'s set ranks higher than *l*'s. So long as the capability set of each individual is subsumed by those each set of equal or higher ranking, outputs as complete rankings of capabilitarian well-being remain coherent with respect to Sen's characterization and representation, in this limited kind of case.

But then there remains the problem of how to provide complete rankings for more common cases that feature non-dominant sets. If both *l* and *k*'s otherwise equivalent capability sets contain co-possible sets of functionings not found in the other's set, then there may be no fact of the matter regarding which outranks the other. In such a case, a complete ranking can only be achieved via some further operational procedure that overwrites partial ranking gaps. Consider, for example, a case in which *k*'s set contains the sets of functionings $\{\{a_1\},\{a_2\},\{a_3\}\}$ and *l*'s set contains sets of functionings $\{\{a_1\},\{a_2\},\{a_4\}\}$, where functioning set $\{a_3\}$ and $\{a_4\}$ are the same except that $\{a_3\}$ contains the functioning of voting while $\{a_4\}$ contains the functioning of attaining a high school degree. Some further process is needed in order to

make any ranked comparison of *k*'s set to *l*'s.[28] For example, let's allow that set $\{a_3\}$ might get assigned a value of 5 within deliberation, while set $\{a_4\}$ gets assigned a value of 4. Though a numerical value can, as a practical matter, be assigned to either set, this possibility does not substantiate the claim that such a process actually *measures* the capabilitarian well-being of $\{a_3\}$ over $\{a_4\}$, or for that matter for *k*'s lives worth living over *l*'s. As I clarify in later chapters, such cases may be incoherent when they treat different indicators standing for different kinds of qualities exhibited by $\{a_3\}$ or $\{a_4\}$ as if they can be traded off in terms of one another, for example as intervals along the numerical scale according to which the weights of 5 and 4 are assigned for *k*'s and *l*'s capability sets, respectively.

Moreover, even if there are procedures by which weights could be assigned within processes of deliberation, Sen repeatedly defends an evaluation process that applies *ranges* of weights, if it is to apply weights at all. When procedures allow for ranges of weights, complete rankings can be given only inasmuch as ranges of weights dominate one another according to what Sen calls 'partial ratios of comparison.'[29] How exactly to establish the degree of partial dominance and what to interpret from such a scale remains a challenging technical question, one that need not be pursued further here.[30] Whatever ranking could be created from the ranged values may remain partial or incomplete; if so, if the deliberative body were to agree on such ranges, some sort of overwriting process would be needed in order to force a complete ranking.

Note, moreover, that even the above set of examples is misleading as it assumes the felicity of evaluation processes used to render different kinds of capabilities indicators – at least according to ranges of weights and typically according to point weights - into total orderings of capabilitarian well-being. As the following example of the application of a Vector View representation demonstrates, such an

[28] Here a pairwise trade-off between {C} and {D} is possible (though that a proportional weight can be given for the comparison is less likely), but consider how complicated the example becomes as soon as *k* and *l* have not only {C} and {D} but also {E} and {F} as possible functioning sets that do not directly dominate. Pairwise trade-offs become exponentially more complicated in that now there are four pairwise permutations, and so on.

[29] See Sen 1970: 404 for the definition of the Quasi-Ordering Comparability Ratio used to proportionally evaluate different sets.

[30] The problem becomes significantly more complex as soon as there is a third item, because ratios of comparability between A:B and between B:C may not hold as A:C. I return to this problem when I consider an extended example of the problem of forced tradeoffs in chapter four.

assumption is misguided. Building from the previous example of middle-class *k*, one co-possible functioning set for *k* on Sen's settled representation - call it set $\{a_5\}$ - might look like

$$\{a_5\} = \{\{\text{college degree in x}\}, \{\text{political activity y}\},$$
$$\{\text{comfortably have z kids in 20s}\}, \{\text{etc.}\},$$

while another co-possible functioning set – call it $\{a_6\}$ - might look like

$$\{a_6\} = \{\{\text{college degree in x}\}, \{\text{political activity y}\},$$
$$\{\text{travel the world for w months in 20s}\}, \{\text{have z kids in 30s}\}, \text{etc.}\}.$$

However, on the Vector View, such a set is to be represented by proxy as a *vector* of indicator values, each taken to stand for different functionings, or even more broadly for different types of well-beings. For example, *k*'s co-possible functioning sets $\{\{a_1\}$ through $\{a_n\}\}$ might be reduced on the Vector View representation to a single vector that includes something like

<educational attainment value x, political activity value y, affiliation level z, etc.>

and that single vector would be given numerical assignments as

<x, y, z, etc.>

where x, y, and z stand for different indicator values, say

x = years of education, y = political activity survey score, z = social affiliation survey score.[31]

The process of evaluation typically used by many multidimensional well-being measures then proceeds by weighting x, y, and z across different vectors (and *not* sets of co-possible functionings) and then assigning single numbers by multiplying assigned indicator values by weights and summing. The resulting scores produce complete rankings of capabilitarian well-being, say to compare *k*'s total capability set against *l*'s.

[31] Where the survey giving the score for z might include and weight responses to questions regarding level of satisfaction for family planning, staying involved in the community, traveling to foreign countries when young, etc

As the example illustrates, this process requires making numerical assignments to different well-being indicators in the vector. I want to emphasize here just how *reductionist* this use of the Vector View representation really is, not only as a representation of Sen's capabilities but also as a means by which to provide an evaluation in terms of capabilitarian well-being. First, the obvious point: the rich array of functionings that might be included in *k*'s set, for example different kinds of political activities or educational experiences, has been reduced to a numerical score for that indicator. Second, when evaluated according to weighted combinations of vector elements into complete rankings of items, the Vector View representation precludes the holistic evaluation of co-possible sets, for example how middle-class *k has the opportunity to choose* between starting a family while young or traveling the world at the same age versus merely attaining some degree of the relevant indicator standing for one or the other. As I've indicated in the previous section, two important pre-suppositions of Sen's pluralistic characterization of capabilities (among many) are the process and agency aspects of freedom as values that affect the evaluation of sets of co-possible functionings. The complex interplay of such values may not be captured by evaluation processes that always render proxy indicators into total orderings, especially, as we shall see in the next chapter, when procedures treat such indicators as if they can be traded off against one another.

In his watershed "Plural Utility," Sen (1980) maintains that a *disaggregated* Vector View representation has more descriptive strength than does some single grade utilitarian metric of pleasure, desire, or choice, often created by processes that collapse the Vector View into numerical values along a single scale. Sen's limited support for the disaggregated Vector View hinges on the recognition that different kinds of moral demands - demands that we intuitively recognize as important -, simply cannot be captured by a single grade metric and should be represented as distinct. Following John Stuart Mill, Sen argues that different types of utility are "non-comparable in objective terms," (Sen 1980: 195). Sen's and Mill's point was demonstrated in the previous section: the pleasure or desire associated with completing a marathon *just is* of a different kind than that of eating an éclair. Ergo, different kinds of pleasures or well-

beings are better represented – at least as a measure suited to descriptive purposes - as separate elements in the Vector View representation.

But are examples featuring eclairs and marathons really sufficient to delimit the process of evaluation used to measure capabilitarian well-being? After all, even if different well-beings are of different kinds, many forms of 'evaluation' could render them comparable in terms of one another such that they can be ranked. For example, couldn't a person claim to prefer eating eclairs "roughly thrice as much" as running marathons, and so the appropriate representation could be a 3:1 trade-off, ranking the former above the latter? Sen's move in "Plural Utility" is to distinguish the Vector View representation from what he calls a 'secondary' process of evaluation that weights and then combines indicators, often as if they can be traded off in terms of one another, into a measure of total utility. Sen argues that the secondary notion of 'total utility' is insufficient to the task of capturing different kinds of well-being values, and so 'total utility' amounts to what Marx calls a 'metaphysical abstraction' (ibid: 198).[32] Sen argues that we can avoid the metaphysical abstraction. The Vector View, in a disaggregated, pre-evaluated form, remains useful because it provides a descriptively-strong representation of well-being. I return to this point in the next section.

Restrictions then apply regarding how evaluation as a secondary process may proceed. In the same article, Sen defends – just as he does in the context of measuring capabilitarian well-being - the likely *incompleteness* of rankings of utilitarian well-being according to weighted aggregations of Vector View indicators (ibid: 205). Sen does allow that in rare cases, all the well-being elements of vector <x> representing one state of affairs might rank higher than all the well-being elements of vector <y> representing another state of affairs such that "x must yield higher total utility than y" (ibid: 205). If all elements in a vector dominate those of another vector, then there is justification for complete ranking without the application of any weights.[33] Yet in the more likely case that vectors do not dominate one another, Sen gives the familiar line that, in general, "weights may be partially specified as lying within

[32] I've illustrated an analogical point in the discussion of Aristotle's analysis of 'exchange value' given in the previous chapter.
[33] The same dominance structure could hold for a vector view that includes capabilities rather than utilities, though, as noted above, this kind of vector would already be a reductionist version of Sen's settled representation of capabilities as sets of co-possible sets.

some *ranges*" (ibid: 205). So even in the context of a critique of measures of utilitarian well-being, complete rankings are unlikely.

Here however, we can start to see how practitioners working within the capabilities approach may misinterpret Sen when they cite his work. Sen (1980) proceeds to argue, again in the context of a critique of utilitarian measures of well-being, that "the narrower the ranges, the more extensive the partial ordering of [total] utility" (ibid: 205). As mentioned above, the idea that ranges of weights can be 'narrowed' (what I've referred to above as an overwriting process) up to some threshold that allows for complete rankings could be misinterpreted as Sen supporting processes that always give complete rankings of utilitarian well-being, and, perhaps, can be incorporated in order to always give total orderings of capabilitarian well-being. Yet what Sen's cautious support for ranged weighting really indicates is that even with respect to the reductionist Vector View representation, surrounding moral concerns generally are either insufficiently definite or are insufficiently agreed upon (or both) when groups of people deliberate about how to evaluate well-being according to those moral concerns. Ranges of weights are likely, at best, and Sen does not provide further clarification regarding how - or why - the narrowing of ranges of weights is to proceed.[34]

In summary, procedures that always provide complete rankings of items do not cohere with Sen's intended process for evaluating capabilitarian well-being. Such operational procedures tend to follow from a Vector View representation that is at best a reductionist proxy of Sen's settled representation of sets of co-possible functioning sets, and is more likely a misinterpretation of Sen's earlier critique of utilitarian measures of well-being. For Sen, evaluation of capabilitarian well-being as sets of co-possible sets is a pluralistic process intended to allow for holistic consideration of options in terms of further values, for example those related to process and agency aspects of freedom, justice, fairness, or other norms and values referenced in the previous section. Weighted comparisons of isolated indicators on the Vector View representation are operational procedures that become incoherent with respect to Sen's

[34] I return to this quandary in the next section, when I consider the possibility that weights might be made more definite as a practical matter. But then, as a practical matter, practitioners should weigh the concomitant danger that the measure may be off-target.

characterization and settled representation of capabilitarian well-being, especially when such procedures are used to always provide complete rankings. Such procedures also transgress Sen's presupposition that capabilities are heterogeneous.

Sen explains, moreover, that rankings are likely to be incomplete because individuals engaging in public deliberation – when they use such a process to assign weights – tend to be able to settle on only ranges of acceptable weights. In such cases, there may simply be no fact of the matter regarding how different items rank in terms of their capabilitarian well-being. One acceptable procedure for Sen allows for ranking sets according to dominance, but Sen explains such rankings will also likely be incomplete. Even so, one might give the JTOBPC-type argument that the same kind of representation often used as a means for decision-making might still be sufficiently useful for measuring capabilitarian well-being, at least in specific contexts or with respect to certain purposes, say for roughly identifying a significant lack of capabilities in contrast with significant wealth. On a similar basis, one could argue that ranges of weights could be narrowed up to the point of producing total orderings. I turn now to a casual discussion of these arguments as they relate to Sen's characterization and representation.

2.4 Misinterpretations of Sen's Position on Practical Comparisons as Justifications for Total Orderings of Capabilitarian Well-being

In the previous sections I have explained how commitments to value-ladenness and heterogeneity of capabilities are supported by pre-suppositions of Sen's characterization and reflected by his settled representation. Measures that always provide complete rankings of capabilitarian well-being are incoherent with respect to these commitments. Such measures tend to incorporate a reductionist Vector View representation along with further procedures of evaluation that fail to reflect Sen's holistic orientation of evaluating co-possible lives worth living. Even when weights – perhaps as a practical matter - are to be used as a means for evaluation, Sen argues that deliberation is likely to result in ranges of weights, producing incomplete rankings in cases for which there should be no fact of the matter in terms of capabilitarian well-being. Yet, as has been noted earlier, different contemporary measurement systems that cite Sen – measures such as the Human Development Index (HDI) – incorporate procedures that always produce complete rankings of capabilitarian well-being. How can this be so?

I've indicated that it is possible to misinterpret the subtlety of Sen's position regarding representations or operational procedures that always produce complete rankings. In the previous section I explained that Sen's critique of utilitarian measures of well-being could be misinterpreted as support for the claim that the Vector View could be a sufficient representational basis for capturing capabilitarian well-being. I've shown that, in fact, the Vector View is at best a highly reductionist proxy, even before 'secondary' evaluation procedures that weight and combine the elements of the Vector View to always produce complete rankings are incorporated. Such procedures are, furthermore, incoherent with respect to Sen's commitment to the heterogeneity of different kinds of capabilities and, as we shall see in the next chapter, with respect to Nussbaum's commitment to avoid tragic tradeoffs. Yet it can perhaps be maintained, especially by one who argues on the basis of the Justification for Total Orderings on the Basis of Practical Comparisons (JTOBPC), that procedures that mimic decision-making processes by always providing complete rankings of capabilitarian well-being are roughly useful, at least for some purposes

and in some contexts. Such a position can, perhaps, be misinterpreted from Sen's cautious acknowledgement that, yes, comparisons often must be made in practical choice situations.

As I've explained in earlier sections of this chapter, Sen argues that capabilities are 'clearly non-commensurable,' - i.e. 'heterogeneous' or 'qualitatively distinct' - most recently in his (2009) *The Idea of Justice*. However, on the next page in that same work, Sen seems to consider the argument from the JTOBPC when he notes the following:

> having a medical intervention and enjoying a visit to a foreign country are two quite non-commensurate achievements, but a person may not have much of a problem in deciding which would be *more valuable* in her condition… The choice and weightings may sometimes be difficult, but there is no general impossibility here of making *reasoned choices* over combinations of diverse objects
> (ibid: 241, emphasis added)

Sen's quote allows that, at least in some cases, reasonable preference stands as an indication of relevant value. Moreover, since practical choices among options inevitably have to be made, preferences for such options could, again as a practical matter, be forced into complete rankings of items. For example, a survey could force respondents to completely rank three or more policy options according to the level of 'capabilities' they believe each option is likely to generate, perhaps on the condition that only one policy can ultimately be implemented. The proponent of the JTOBPC might argue that since such procedures arguably produce an indication of relative value, perhaps they serve well enough for measuring capabilitarian well-being?

Firstly, I note in passing that Sen's lynchpin example here is merely a pairwise *comparison* leading to a choice *between* adding one or another functioning to some life worth living, rather than a more complicated assessment *among* three or more different kinds of functionings. As we shall see in chapter four, issues of transitivity arise when we choose among three or more options, and such issues make it all the more challenging to claim that decision-making procedures that always render such options into complete rankings truly capture the different kinds of capabilitarian well-being values constitutive of or relevant to such options. The fact that we *must* make choices does not imply that such

preferences capture all relevant values and so are a sufficient basis for actually measuring capabilitarian well-being.[35]

In the same work, Sen provides a further response to the proponent of the JTOBPC: "The presence of non-commensurable results only indicates that the choice-decisions will not be trivial (reducible to counting what is 'more' and what is 'less') but it does not at all indicate that it is impossible – or even that it must always be particularly difficult" (ibid: 241).[36] For Sen, comparisons of choices in terms of what is 'more' or 'less' are often unable to fully capture the relevant set of values, as reductive quantification 'trivializes' the relevant values. In fact, different choices often are 'non-commensurable' on Sen's view, even though, as Sen recognizes, people do not always find it difficult to make choices in 'non-commensurable' cases. Recall the discussion of the 'incomplete' Goode-Homolosine and the 'complete' Mercator projections in the previous chapter. Sen is making a distinction here between our getting a descriptively accurate if incomplete picture as opposed to a forced, typically reductionist, and yes, complete picture that may suffice for some further purpose.

In his earlier "Plural Utility," Sen challenges the use of choice as a basis for measuring utilitarian well-being. In particular he faults using choice as a *definition* of utility, for such use entails that revealed preferences fully capture all the relevant values that together constitute the utilitarian well-being a person experiences relative to a given situation. Sen gives a counterexample of a ship captain whose expressed values seem to belie those indicated by his honorable choice: "You drown with your ship doing your duty but desiring escape, and the choice-based definition of utility declares that you have once again – as always! – succeeded in maximizing your utility" (Sen 1980: 206). This somewhat wry example does clarify the issue at hand: relevant values interact in complex ways, the reality of which may not be captured by reduction to the three preference options of more, less, or equal to. Similar to his argument in the above (2009) quotes, Sen allows that choice might provide *evidence* for utility - that it provides the

[35] Moreover, such choices are 'tragic' in that they often reduce or externalize important dimensions of value in order to generate a settled decision. For further discussion of tragic choices and how recognition of related problems with reduction or externalization supports intrinsic non-commensurability as a feature of Nussbaum's characterization and representation of capabilities, see chapter three section three.
[36] Firstly, it is important to recognize that Sen is claiming non-comparability of the kind he worries about precludes *counting* as a sufficient representation of choice-decisions; other important values or principles are invoked. This claim foreshadows the analysis given below of the well-being measurement quandary between countable quantities and non-commensurable qualities.

basis for generating a picture that might be of limited use. Yet such a basis is likely to be too distorted to serve many general purposes: "Choice may say too little about utility when someone else makes the choice; too much when the chooser's secondary utility ranking is incomplete; and be too far off the mark when the choice is wholly or partly based on non-utility motives" (ibid: 206). Note especially the second point. In cases for which there is no fact of the matter in terms of the chooser's 'secondary utility ranking,' choice overwrites the gaps in favor of a 'complete' picture. A descriptively-accurate, coherent representation would instead call for 'incompleteness.'[37]

Let's re-consider an example from section two as a demonstration of this Mercator-esque kind of overwriting. On the basis of which choice most improves well-being, Annapurna must, again, choose the appropriate recipient for an important wage-earning opportunity, and the importance she gives to different values like health, happiness, or desire satisfaction might influence her decision. In many cases she may find that she had to omit, reduce, or overemphasize the importance of some values in order to select a laborer, even when she does her best to carefully to weigh the relevant values against one another in order to generate a complete ranking of the different options. For example, it might be the case that awarding the job to Rogini (who has a chronic health ailment) produces the most health and the least desire satisfaction, while awarding the job to Bishanno (the most depressed) produces less health but the most happiness, while awarding the job to Dinu (the poorest) produces the least health but the most utilitarian desire satisfaction. In this kind of case in which no laborer's situation clearly dominates the other in terms of the relevant values, there may simply be no fact of the matter regarding which outranks the other. And yet, Annapurna will have to make a choice!

Recall that Sen's characterization and representation are linked to a philosophical history. In the previous chapter, for example, I considered Aristotle's defense of *asummetria* as it relates to the recognition that exchange value is often unable to capture the different kinds of values – including the relevant kinds of well-being values - of economic objects. Moreover, as noted in the previous section, Marx would similarly call any single grade metric of total utility a 'metaphysical abstraction,' that

[37] See chapter one section three for discussion and comparison of the Mercator and Goode-Homolosine projections.

overlays a 'forced uniformity' onto different *kinds* of human activities (ibid: 198). Arguments given by these luminaries might provide a conceptual basis for worrying about the incoherence of measures that allege to follow Sen or Nussbaum's characterization yet proceed to incorporate procedures that always provide total orderings of capabilitarian well-being. But still, the practical challenge from the follower of the JTOBPC remains: if people can and do make decisions using such procedures, and can assent, as Annapurna surely does above, that such decisions regard something like the 'well-being' of options considered, then isn't the measurement output roughly useful, even if incoherent and in danger of being off-target?

Let's return to "Plural Utility" where, right after recognizing Marx's concern, Sen offers the following reply:

> even those who see no difficulty at all in aggregating utility in terms of intensity, may find it *useful* to consider the disaggregated picture. This usefulness may be related to the *moral* rejection of intensity-aggregates… or to seeking *descriptive richness* (as in A.C. Pigou's distinction between 'economic welfare' and other types of welfare), or *predictive value* (if, for example, people's choices are taken to be related not just to the sum-total of utility but in a more complex way to the vector of utilities). Non-comparability is sufficient but not necessary for the usefulness of the vector view
> (ibid: 198, original emphasis).

Like so many of Sen's arguments, the point made here is both subtle and pithy. For some purposes, the disaggregated vector representation is even more useful than is the 'secondary' picture of total, or, relatedly, complete rankings of capabilitarian well-being. Commitment to descriptive strength is one such purpose, but interestingly even commitment to *predictive value* might, for Sen, favor the need to represent different kinds of indicators as separate; the disaggregated and thus 'incomplete' picture might provide a more useful basis for recognizing important features regardless of whether they can be rendered into a total ordering of items according to some 'secondary' process of evaluation. In fact, that rendering process can mute the importance of such features. For example, if some feature of a situation, say a person's inability to sleep because of acute, chronic pain, were rendered less poignant because other features could 'compensate' for that feature in terms of intensity of utility, then the decision made on such

a basis would be one that fails to accurately capture the situation in terms of the relevant values. I return to this point in the next chapter, foreshadowing, for now, the related problem for procedures that evaluated different kinds of capabilities as if different features can be traded off against one another.

It may in many cases be both moral and useful to keep different indicators, and by extension different values, separate so that values cannot be understood as if they can be traded off against one another. For example, it might be important to leave in the indicator that stands in for how Rogini experiences her disability, *even if* some amount of another indicator like income could be used to relieve Rogini's chronic ailment and so seems like a sufficient substitute. The picture provided with the indicator left in is simply more faithful to the relevant narrative. I'll return to the importance of this implication in the next chapter: Nussbaum argues that different kinds of qualitatively distinct values must, in general, remain distinct if we are to accurately represent the relevant states of affairs in terms of capabilitarian well-being, even if we often must make a tragic decision for one or another option. In the same chapter, I explain how Elizabeth Anderson supports keeping different kinds of values separate for similar reasons, both moral and practical.

2.5 Simplified Stand-ins: Contextual versus Universal Covering Considerations

Let's consider a further line of practical argument that practitioners can give in support of using incoherent measures of capabilitarian well-being, even though such measures are in danger of being off-target. In the previous section I indicated, following Sen, that a disaggregated or incomplete view is more useful with respect to descriptive purposes and even some predictive purposes than is one that forces a complete picture, say by means of procedures that always provide total orderings of items. Such procedures can overwrite gaps in the rankings when there should be no fact of the matter in terms of their capabilitarian well-being, and the picture provided could be harmful depending on the purpose to which the measure is lent. Accordingly, I have argued that the incoherence warrants further investigation, especially to determine what further reasons have been brought into the measurement system in order to justify the creation of a complete picture. I do, however, allow the practical point that the complete picture may still fit well enough with the actual, incomplete picture to be of limited use, especially in cases of great urgency or for which a very rough picture may be sufficient. However the JTOBPC-type claim that such practices might work for specific contexts or purposes does not justify the general acceptance of such procedures.

Let's proceed by re-considering the possibility of creating a single grade metric that functions as what I have earlier called a 'simplified stand-in' for a concept that admits of much greater diversity of valuation than does the simplified stand-in. For example, we saw in chapter one that Aristotle considered the possibility that the Greek concept *chreia* – roughly translated as need or desire - might stand in for the different kinds of values – including well-being values – featured by different economic objects. Importantly, Aristotle noted that *chreia* works 'sufficiently' in some cases, but often fails to capture all relevant values. On a similar note, let's consider how some central feature of capabilitarian well-being might be treated as a simplified stand-in for the different kinds of capabilitarian well-being values that feature for different kinds of capabilities. As in the case of Aristotle's *chreia*, the relative level of such a feature could be treated as if that feature is a common currency of exchange, according to which different

numerical values can be assigned to different kinds capabilities indicators and then, as in the case with Aristotle's exchange value, total orderings of items can be created in terms of that common currency, *as if* those orderings represent the capabilitarian well-being of such items.

Consider again, for example, the different kinds of capabilities that are needed by Dinu, Bishanno, and Rogini. Annapurna's choice to award the job to one of the three laborers is what philosopher Ruth Chang (2017) calls a 'hard choice' such that the situation of each laborer may not truly stand in a relationship of greater, lesser, or equal to one another, given that different levels of different kinds of values are relevant for each option. As I mentioned in the previous section, in such a case there may be no fact of the matter in terms of rankings of capabilitarian well-being. Yet in such a case, Annapurna can still manage the hard choice by comparing the different situations according to what Chang (ibid) calls a 'covering consideration,' perhaps by preferring to award the job to the laborer who will better be able to provide for others, or who stands to get more Quality-Adjusted Life Years (QALYs), or (less plausibly) who will earn more gross income, etc. Furthermore, consider that in some cases comparison by means of covering conditions might admit of degree, for example as tradeoffs of intervals of, say Rogini's physical health for Bishanno's mental health in terms of QALYs, or even the former in terms of the latter in terms of cost of treatment. In such cases, the overall 'comparison' is admittedly rough given that the covering consideration does not fully capture the relevant constitutive qualities of either item or among three or more items. Call this kind of comparison - one that admits of degree and that does not fully capture the relevant constitutive qualities of either object of comparison - 'covering comparability.'

Covering comparability could be applied, as a further example, when one compares the 'nutritional value' of a chicken (*c*), a salad (*s)*, and a dessert (*d*) by means of their caloric content to determine the relevant comparison relationship which, if transitive, can be represented as a complete

ranking.[38] The covering consideration (calories) here admits of degree, and practitioners can treat the overall comparison as a *ranking* of the greater 'nutritional values' of the objects *c*, *s*, and *d* on behalf of the covering consideration. Say the desert has 1,000 calories, the chicken 500 calories, and the salad 200 calories. Practitioners would admit there is no real basis for claiming that the dessert has five times the 'nutritional value' of the salad. Furthermore, the chicken should generally be preferable in terms of nutrition, even if it has fewer calories, and in some cases the salad might be preferable. The measure clearly misses its target of determining appropriate 'nutritional value' in such cases. However, given an urgent purpose and context, say to determine how to distribute food to starving people, a measure of rankings in terms of total calories may be roughly useful for capturing the 'nutritional value' of the relevant items.

Regarding how covering considerations might justify procedures that always provide complete rankings of capabilitarian well-being, I consider two different senses of covering considerations as they might relate to Sen's notion of substantive freedom. Given that two or more phenomena are comparable according to a covering consideration, either the set of qualities constituting that covering consideration is highly sensitive to contextual differences such that limited-scope practical comparisons of different phenomena are possible, or the covering consideration is roughly universalizable with respect to some underlying value such that practical comparisons among phenomena are possible with respect to a wide array of contexts and purposes. Call the former CCC for contextual covering comparability and the latter UCC for universal covering comparability. I argue that while scales of degree provided by CCCs - for example in the calorie-based case of *c*, *s*, and *d* above - are often defensible relative to the purpose of the measurement, scales of degree provided by UCCs - for example the universal application of *chreia* as a general covering consideration for economic value - often are less defensibly on-target with respect to complex concepts that have different kinds of features.

[38] I leave the term 'nutritional value' undefined in this instance, noting in passing that 'nutritional value' is, like 'well-being,' what philosopher Nancy Cartwright, following Otto Neurath, calls a *ballung* concept in that takes on different sets of features depending on context, even when people tend to agree to the meaning in many contexts. See below.

Especially in the context of measuring capabilitarian well-being, UCCs that always provide total orderings by treating different kinds of features as if they can be exchanged according to a degree of common currency will tend to produce measurement outputs that strongly challenge our intuitions regarding the target concept. For example, given a choice for a local municipality between putting public funds into schools or the police force, a CCC may provide information useful for quantitative analysis, perhaps in terms of money saved or crimes prevented, that supports a practical decision to grant one or another allocation of funds. And such analysis may generally track what a group of people deliberating about such options would call the 'capabilitarian well-being' for each. However, that the same covering consideration may not be applicable in contexts in which further values and related covering considerations become relevant - say, for example, a nationwide choice between supporting impoverished schools or subsidizing local businesses in order to better improve capabilitarian well-being. In such a case, a monetary standard or even some more complex index of indicators as UCC may provide highly unintuitive rankings relative to further features that are important to local contexts.

Note that a UCC could – as a practical matter - be used as a general standard for ranking or even for proportioning resources for policies. Given that capabilities indicators are typically given as observable, interpersonally-comparable units like currency, years of education, or mortality, it might be plausible to claim that indicators of capabilities could be represented according to some universal covering consideration, perhaps a key feature of the characterization like (but not necessarily limited to) substantive freedom. One might, for example, claim that improved healthcare generally ranks higher on a scale of 'substantive freedom' than does improved schooling, and that, moreover, the former exhibits a greater *degree* of substantive freedom. However, while substantive freedom could be practically justified as CCC – for example because some local policy decision must be made -, it is a much greater leap to claim substantive freedom as UCC justifies measurement procedures that treat different kinds of capabilities indicators as if they can be traded off in terms some shared degree of common currency.

In fact, capabilitarian well-being or even a simplified stand-in like substantive freedom is complex and contextually-sensitive such that it exhibits different sets of qualities in different contexts - what Nancy Cartwright et al. (2011, 2016) call a *ballung* concept, following philosopher of science Otto Neurath.[39] For example, following Sen's pluralism and Nussbaum's 'multiple realizability,' the set of capabilitarian well-being values that would be relevant for evaluating options in terms of capabilities for people in Kigali differ significantly from those that are relevant for evaluating the same for people in Toronto. The more universal a covering consideration becomes with respect to a *ballung* concept, the less likely it is to track important qualities across different contexts, even if it may have provided roughly useful approximations in some.

Practitioners may bite the bullet and argue it is just too useful to use substantive freedom or some similar simplified stand-in as a UCC in spite of incoherence, much in the same way it is arguably just too useful to proceed as if exchange value truly measures different sets of constitutive qualities in different economic objects even though it often doesn't. In fact, there are cases in which exchange value tracks away from important valuations such that we have good reason to be concerned when the measure is off-target with respect to those valuations. We just may, for example, decide that caretaking services are worth more than what is often registered as minimum wages determined by the market. Regarding similar issues, Cass Sunstein (1994) notes that there are well-justified restrictions against representing human organs (or even human beings) in terms of their economic exchange value, even if such trade-offs might be possible. Those restrictions are there because of the danger of using exchange value as a simplified stand-in for other values, for example those that are relevant to practices involving human organs. I return to this problematic issue of treating different kinds of capabilities as if they can be traded off against one another in the next chapter, as it relates to Nussbaum's characterization.

To take a less extreme example, if Annapurna omits an important kind of value or treats different features of the situations of Rogini, Bishanno and Dinu as if they can be exchanged for one another,

[39] See Efstathiou "Is it Possible to Give Scientific Solutions to Grand Challenges? On the Idea of Grand Challenges for Life Science Research" (2015: 52). See also chapter four section two.

Annapurna may end up making a regrettable decision on this basis. If Annapurna is to make a prudent decision in terms of capabilitarian well-being, she may need to favor a means of representation by which different kinds of capabilitarian well-being values are treated as distinct. Though procedures that always provide for total orderings of capabilitarian well-being do provide a method by which a decision can be made or other purposes of measurement pursued, incoherence of the measurement system should be investigated, especially with respect to context and purpose. I grant that even if incoherent, the system may provide outputs that are roughly useful, as in the limited nutrition case above. Yet the recognition that that system may provide a useful-enough output in some contexts is not sufficient to justify general use as a UCC. Procedures that always provide total orderings of capabilitarian well-being cannot be justified as generally acceptable practices simply because they may provide roughly-useful outputs relative to specific contexts and purposes.

Chapter Three:

Nussbaum's Characterization and Representation:
Prohairesis and Pragmatism as Justifications for the
Commitments to Value-Ladenness and Qualitative Distinctiveness

3.1 Introduction

In this chapter I analyze Martha Nussbaum's characterization, representation, and intended operational procedures for measuring capabilitarian well-being, many of which differ from those of capabilities approach co-founder, Amartya Sen.[1] Nussbaum's characterization has a stronger basis in Aristotelian political and moral philosophy, especially the Aristotelian process of prudent decision-making known as '*phronêsis*,' introduced in chapter one. In particular, Nussbaum defends a process of *phronêsis* according to *prohairesis* – what she calls 'desiderative deliberation' - in which ethical agents recognize different kinds of values including at the level of deep emotion. Nussbaum's characterization of *phronêsis* dovetails with Sen's commitments to both the value-ladenness and the 'heterogeneity' of capabilities, discussed in the previous chapter. But Nussbaum's characterization goes further: she argues directly against 'tragic trade-offs' in which capabilities are represented as if some amount of one can be traded off against another.

As with her characterization, Nussbaum's representation of an objective list differs significantly from Sen's settled representation of sets of co-possible sets. Nussbaum's representation gives an outline of different 'dimensions' of capabilities - for example health, affiliation or education - that are needed to live a flourishing or dignified experience. This important difference between Nussbaum's and Sen's representation tends to be misinterpreted by practitioners working within the capabilities approach: although both Sen's and Nussbaum's representations can be operationalized according to sets of proxy indicators, practitioners who use indicators to represent different 'dimensions' of capabilities more clearly follow Nussbaum's representation, even though such practitioners tend to cite Sen as the relevant

[1] I further discuss the representational theory of measurement in the following chapter section two.

reference for the concept of 'capabilities.' Because they actually follow this core aspect of Nussbaum's representation and characterization, such practitioners commit to even stronger arguments against procedures that treat different kinds of capabilities as if they can be traded off against one another. Such procedures are commonly used by measures that always provide total orderings of capabilitarian well-being.

Nussbaum's representational structure also plays a distinctive role in guiding deliberation about capabilities, one that mirrors her characterization of desiderative deliberation. The commitment to having distinct dimensions – what I call 'central features' - of capabilities within Nussbaum's list frames how agents express, recognize, and evaluate in order to attain what she calls a 'reflective equilibrium,' a general consensus between our more intuitive valuations and our reasons for valuing different kinds of capabilities.[2] Nussbaum's support for prudent decision-making in which people recognize and give influence to expressions of deeper emotions and desires further clarifies the usefulness of framing deliberation according to some list-type outline and, relatedly, to providing access in deliberation to persons whose experiences are relevant.[3]

Nussbaum's representation has been challenged for defending a specific – sometimes called 'perfectionistic' - vision of the good life.[4] I acknowledge that at first glance, Nussbaum's objective list does appear to make a particular normative claim regarding what central features should constitute a life of Aristotelian flourishing.[5] Yet at least with respect to a broader purpose of measuring capabilitarian well-being, Nussbaum's representation remains both open-ended and 'multiply realizable' such that it can be interpreted according to different context-relative sets of values.[6] Here we find a conundrum: the structure of Nussbaum's list provides for distinct kinds of central features, yet that structure is open-ended and can shift depending on context. Recall from previous chapters the JTOBPC-type position of

[2] See, for example, Nussbaum (2011: 77)
[3] See in particular section four of this chapter. See also sections three and five of chapter four regarding responses to social power relations in deliberation. See also chapter one section four regarding the necessary role of experience in *phronêsis.*
[4] Regarding the label of 'perfectionism' as applied to Nussbaum's characterization, see, for example, authors cited by Claassen (2018: 24). See also Claassen's extended argument against Nussbaum's 'perfectionism' in the same chapter.
[5] See chapter one regarding *eudaimonia.*
[6] See chapter one section three and also section three of this chapter.

practitioners who support measures that always provide total orderings of capabilitarian well-being, even when they recognize such practices are incoherent with respect to Sen and Nussbaum's concept.[7] Such practitioners can point to the conundrum as evidence in support of the following relativist challenge: because Nussbaum's list is open-ended such that central features can be interpreted differently depending on context, different kinds of capabilities can be rendered into total orderings of capabilitarian well-being depending on the decisions of the deliberative body. As discussed in chapter one, such a view is arguably that of Chowdhury and Squire (2006), who accept a wide array of methods for assigning weights used to render total orderings of capabilitarian well-being.

In fact, though Nussbaum's list is open-ended, it does provide a structure rigid enough to support treating different kinds of central features as qualitatively distinct. In this chapter I show how Nussbaum's list is derived according to a process of engaged discussion, reflection, and – ideally - deliberation that defensibly holds across cultures, even as the somewhat rigid outlines of central features of the list remain sensitive to different sets of cultural values.[8] I also offer a pragmatic response to the JTOBPC-type challenge from authors such as Chowdhury and Squire: on the basis of best available evidence and argument, there are substantive reasons to avoid procedures that always provide total orderings of capabilitarian well-being.[9] Foremost among these reasons is the practical concern that such total orderings produce a reductive picture that is likely off-target in that it fails to accord with people's common intuitions, particularly their intuitions regarding how amounts of different kinds of capabilities might be traded off against one another.[10]

I further the pragmatist response by surveying related philosophical arguments regarding how people should deliberate regarding valuations (and evaluations) of different kinds of doings and beings.[11] I consider both Hilary Putnam's recent account of pragmatist inquiry and Elizabeth Anderson's

[7] Again, the Justification for Total Tradeoffs on the Basis of Practical Comparison. See chapter one section one.
[8] See Claassen (2018) p. 24.
[9] See chapter one section three.
[10] See the next chapter for a demonstration that emphasizes these kinds of intuitions.
[11] See chapter two section two regarding the distinction between 'valuation' and 'evaluation.'

expressive theory of value as outlined in her (1993) Value in Ethics and Economics. Anderson's theory mirrors Nussbaum's account regarding how emotionally-informed agents of *phronêsis* recognize – either individually or in public deliberation – different kinds of functionings and values. Anderson calls such items 'incommensurable,' which I take to accord with Nussbaum and Sen's (1987) use of the term 'non-commensurable' as it relates to the qualitative distinctiveness of different kinds of capabilities, which is what I mean by it throughout. Recall chapter one section four in which I reference Sen and Nussbaum as saying that the measurement "procedure insists on treating each of the values involved as a qualitatively distinct item, not reducible to any other item, not conceivable as simply a certain quantity of something else. This commitment to the qualitative integrity of each value is one of the greatest advantages of this procedure over other approaches that might be used … in assessing traditional cultures." (1987: 25-26).[12] I note at the outset that though Anderson does not justify distinctiveness of different kinds of values in the exact same way that Nussbaum does, Anderson's expressive theory still supports the claim that practitioners should be concerned whether procedures that always provide total orderings of capabilitarian well-being are sufficiently on-target relative to the given context and purpose.

Yet as discussed in previous chapters, a commitment to the 'pedestrian middle-grounds' of pragmatism does not justify the complete avoidance of any and all procedures considered to be incoherent with respect to the target concept.[13] Pragmatist arguments can be given to justify – relative to a limited set of purposes and contexts - incoherent measures that still might be *roughly useful*. Anderson, for example, offers a limited practical justification for such procedures within the scope of what Anderson calls 'goodness-of-a-kind' judgments, for example those that determine the ranked scores given in Olympic decathlon or figure skating competitions. I argue that, relative to this limited kind of purpose, incoherent measures that always provide total orderings of capabilitarian well-being may be treated as roughly useful performance standards, even if in danger of being off-target. Yet, as a practical matter, the incoherence

[12] See also Comim, Qizilbash, Alkire (2008: 164-167).
[13] See chapter four section two regarding Neurath's famous boat metaphor. Sometimes in order to 'keep the boat afloat' it is practical – and arguably in keeping with the *actual* advancement of the sciences - to work with whatever are the better options we've got, even if they're far from ideal.

warrants further investigation regarding the significance of the harms that may result if the measure is significantly off-target.

Nussbaum's characterization and representation provide a further distinction that influences whether incoherent procedures that always provide total orderings of capabilitarian well-being may be considered roughly useful. Nussbaum's objective list representation posits a context-sensitive threshold for each central feature of capabilities such that if an individual is below the threshold for any central feature, that individual experiences a *tragic lack* that significantly inhibits the attainment of a flourishing or dignified existence. Tragic lacks of different kinds of central features are, according to Nussbaum, not comparable and cannot be treated as if they are intersubstitutable. Above the relevant thresholds, however, different kinds of capabilities may be less resistant to procedures that render them as if they are intersubstitutable. I consider this practical point, among others, in the following sections.

3.2 Nussbaum's Characterization: The Value-Ladenness and Qualitative Distinctiveness of Central Features of Capabilitarian Well-being

Following her background in the history of philosophy, Nussbaum develops her earlier characterization from an exegesis of Aristotle's views on well-being.[14] Recall from chapter one that, on the Aristotelian view, different kinds of virtues - also called 'goods of the soul' - are relevant for prudent decision-making, or, more broadly, for pursuing a flourishing life of *eudaimonia*.[15] Those different kinds of goods of the soul are the basis upon which Nussbaum defines her list of different central features of capabilities.[16] Like Sen, Nussbaum provides her characterization by developing a comprehensive overview of relevant features and concepts rather than giving a single analytic definition. In this section I distil those features and concepts into two different purpose-sensitive characterizations, both of which retain the commitments to value-ladenness and the qualitative distinctiveness of different kinds of capabilities.

In her monograph Creating Capabilities, Nussbaum (2011) states that capabilities are the following:

> [they are] what Sen calls 'substantial freedoms,' a set of (usually interrelated) opportunities to choose and to act… not just abilities residing inside a person but also the freedoms or opportunities created by a combination of personal abilities and the political, social, and economic environment
> (Nussbaum 2011: 20).

Recall Sen's shorthand characterization of capabilities *as the extent of people's freedom, also called 'opportunities' to achieve doings and beings they have reason to value.* Nussbaum's characterization is generally similar: capabilities are different kinds of opportunities and freedoms. One significant difference, however, is that Nussbaum focuses on the *abilities*, both internal and external, that people are

[14] See Nussbaum (1986a: 89-121, 290-317) and Nussbaum (1986b).
[15] See chapter one section three.
[16] The definitions provided in Nussbaum's list are arguably part of her characterization. Still I analyze that content, along with Nussbaum's call for thresholds for each of those central features, more directly as part of her representation.

able to develop.[17] Without abilities like using practical reason or knowing how to interact and build affiliation with others, a person cannot achieve a flourishing existence. Although Sen's characterization recognizes abilities as kinds of doings or beings that influence the extent of people's freedom, Nussbaum's brings the importance of abilities to the forefront.

This emphasis on abilities – what Nussbaum (2011) also calls 'basic capabilities' - provides a more rigorous precedent for challenging parochial values that might bias the valuation and evaluation of such abilities, for example those regarding the education of persons hindered from achieving such activity on account of their belonging to one or another social category of gender, race, ethnicity, etc. Note the related emphasis on what Nussbaum calls people's substantial freedoms – what Sen also calls 'substantive' freedoms. The capabilities measure is intended to capture what opportunities are actually available relative to external conditions such as people's economic, political, and social situations. Nussbaum's characterization also fits with Sen's emphasis on 'procedural freedom' in claiming that relevant opportunities cannot be nominal, for example when they are merely proclaimed to be accessible for all; they must be opportunities that people actually can and do achieve.[18]

On the basis of the preceding analysis, let's entertain a shorthand characterization of Nussbaum's basic capabilities as the following:

> **Nussbaum's Basic Capabilities (Shorthand):** Basic capabilities are situationally-sensitive basic abilities to freely achieve opportunities for valuable doings or beings.

We could follow Sen's shorthand in claiming that what makes doings and beings – also called 'functionings' - valuable are the diverse kinds of *reasons* that people can give within pluralist deliberation, but Nussbaum's account of 'reasonable' valuation and evaluation has a different philosophical basis than does Sen's. Even in her earlier writings, Nussbaum claims that valuable

[17] There is, of course, room in Sen's account for abilities, also called 'basic capabilities,' yet their cultivation is less directly defined within Sen's characterization to the extent that it is in Nussbaum's.
[18] Consider, for example, Steve McQueen's film *Red, White and Blue*, a historical fiction account in which a young police officer faces intense racial discrimination and overt racism even though he is nominally 'wanted' for the force.

functionings should follow from a vaguely outlined list of heterogeneous "features of importance" – what I call 'central features' - that define a flourishing human existence (Nussbaum 1992: 216). What really distinguishes central features from other functionings people may have reason to value is Nussbaum's claim that the "a life that lacks any of [the central features], no matter what else it has, will be lacking in humanness" (Nussbaum 1992: 222). An important implication of this claim – really a point of return throughout the chapter - is that different kinds of capabilities, especially central features, are qualitatively distinct because, for the most part, no amount of one central feature can be traded off against the severe lack of another.

More can be said regarding the value-ladenness of central features. In the same (2011) monograph, Nussbaum considers an important question regarding human autonomy as it relates to capabilities:

> What is each person able to do and be?' In other words, the approach takes *each person as an end*, not asking just about the total or average of well-being but about the opportunities available to each person. *It is focused on choice or freedom*… It thus commits itself to respect for people's power of self-determination. The approach is resolutely *pluralist about value*
> (ibid: 18, original emphasis).

Here Nussbaum's commitment to value pluralism overlaps Sen's pluralism in that a person's particular valuations warrant a special kind of respect for the person as an 'end' when such valuations are expressed, especially within deliberation.[19] On account of this respect for the autonomy of the individual, the valuation survives what Sen calls 'critical scrutiny,' as discussed in the previous chapter. Respect for autonomy also constrains the way that central features are outlined and evaluated, for example in providing a basis through which to challenge valuations that favor historically-sedimented norms.[20]

Consider, for example, the life outcomes that might be valued according to the mono-theistic and somewhat apolitical worldview of an Amish farmer versus those valued by a politically active, liberal

[19] Regarding Sen's pluralism, see especially the previous chapter section two.
[20] The commitment to defending the autonomy of agents allows that individuals might interpret and give support to the valuations of others, much in the same way that allies in a process of Sennian pluralist deliberation might help to defend or otherwise render an expression of value more articulate for the deliberative body. See the previous chapter, section three.

atheist. Though the former may not value certain forms of state-sponsored education and the latter may not value certain modes of religious practice, the ways those persons express the values constitutive of or related to their intended life outcomes deserve support in the form of public recognition and, in some cases, protection.[21] Nussbaum's support for a Kantian notion of *respect* for autonomy might, furthermore, be used to justify measuring capabilitarian well-being in terms of substantial opportunities both for state-sponsored education and free religious practices, even though different groups of people will evaluate those functionings in different, often clashing ways.

In later work such as Women and Human Development, Nussbaum (2000a) also defends the central importance of a related concept of human *dignity*. Nussbaum fuses that concept with a political purpose of establishing constitutional principles that structure the politically-guaranteed safety net of a given society:

> the best approach to this idea of a basic social minimum is provided by an approach that focuses on *human capabilities*, that is, what people are actually able to do and be – in a way informed by an intuitive idea of a life that is worthy of the dignity of a human being. I shall identify a list of *central human capabilities*, setting them in the context of a type of *political liberalism* that makes them specifically political goals and presents them in a manner free of any specific metaphysical grounding
> (Nussbaum 2000a: 5, original emphasis, underscore added).

First off, let's grant that there is significant conceptual overlap between this and Nussbaum's earlier work regarding autonomy: respect for *dignity* of persons strongly resonates with respect for persons as autonomous ends. As related to the purpose of measuring capabilitarian well-being, principles derived through deliberation according to either conception will likely be similar. Even so, it is worth noting the shift in Nussbaum's conceptual language relative to a more specific political purpose of establishing a constitutional basis for a social minimum. The quote also illustrates that Nussbaum seeks to remove any deeply metaphysical assumptions – for example those related to respect for Kantian 'ends in themselves' -

[21] Here there is room for an entire thesis or more regarding related questions of paternalism, state protection, taxation, libertarian vs. liberal rights, etc. The goal of this thesis is not to argue for the primacy of one or another conception, but to get a sense of Nussbaum's characterization and also a sense for the reader of surrounding arguments. In this case, then, what protections a state would need to guarantee for its citizens depend on how much the relevant religious or political practices inhibit people from attaining the life outcomes that they consider the most valuable. See also the related discussion regarding individual perfectionism and general perfectionism in the comments in this section, as well as Sen's characterization of agency freedom and well-being freedom in the preceding chapter.

from her narrower view such that it can instead be justified by best available intuitions and evidence in processes of intra and inter-cultural deliberation, resulting in what Nussbaum calls a 'reflective equilibrium,' following Rawls.[22]

The above two quotes present both a broader and a narrower view of Nussbaum's characterization of capabilities, each of which is oriented toward a different set of purposes.[23] The broader view follows Nussbaum's earlier work with the World Institute for Development Economics Research of the United Nations University (WIDER) toward developing a progressive, philosophically-based welfare measure. The characterization that follows the WIDER – call it the 'broader' - emphasis can be given thusly:

> **Nussbaum's Broader Characterization**: Capabilities are people's situationally-sensitive freedoms and abilities to achieve opportunities for doings and beings that are valuable according to a list of different kinds of central features that together constitute a sufficiently autonomous human life.

As is discussed below, the characterization follows the Aristotelian attempt to ascertain the best possible functionings for humans as autonomous social entities who use practical reason.[24] The broader characterization focuses on measuring substantial freedoms and opportunities according to a multi-feature view of a flourishing life of *eudaimonia*, one that respects individual self-determination and pluralist valuation.

Yet Nussbaum also gives a more constrained political emphasis to her characterization. Capabilities are recognized and measured – ideally within public deliberation - in order to establish a

[22] The drive to remove the philosophical scaffolding, so to speak, brings with it the question of whether the Aristotelian commitments to value-ladenness and qualitative distinctiveness remain. I return to this point below.
[23] This classification of two different purposes also follows Claassen's nomenclature of 'Nussbaum-I' and 'Nussbaum-II'. See Claassen (2018: 31). See also Chad Kleist's excellent thesis on *Developing Capabilities: A Feminist Discourse Ethics Approach* (2016) Chapter 1 for a discussion of the differences between Nussbaum's 'flourishing' and her 'dignity' versions of capabilities.
[24] See chapter one for further discussion of this emphasis as it relates to Sen's characterization.

basic minimum target of what a society should attempt to provide for citizens. Call the more constrained characterization the 'narrow' emphasis, which can be given thusly:

> **Nussbaum's Narrower Characterization:** Capabilities are people's situationally-sensitive freedoms and abilities to achieve a basic social minimum of opportunities for doings and beings that are valuable according to a list of different kinds of central features that together constitute a life of human dignity.

Let's note here that the social minimum generated by the list is intended to be publicly accepted much in the same way that Rawls' principles of justice are accepted from the Rawlsian original position, as the attainment of a reflective equilibrium in which our intuitions generally fit with reasons given. Nussbaum claims that respect for dignity does not need metaphysical justification, for example some deeper moral theory, because it is sufficiently established by checking our intuitions along the way to reaching the relevant consensus.[25]

Nussbaum's move away from metaphysical assumptions opens the door for the relativist, JTOBPC-type challenge introduced in the previous section: without the grounding provided by a moral theory, how defensible in disparate cultural contexts can reasons given in support for the narrower characterization really be? We shall see in the next section that though Nussbaum argues her narrower characterization lacks a metaphysical basis, the characterization does defensibly hold - albeit in an open-ended form - across different contexts. Similarly, it should not surprise us that both the broader and the narrower characterization settle on the same general list of central features. On the Aristotelian view that Nussbaum follows, the good life is the good *political* life, and the structure of good government should be the one that provides fair opportunities for citizens according to reasons that best fit with deeper intuitions. In the remaining sections I argue that Nussbaum's characterization *phronêsis* according to

[25] Though unlike Rawls, Nussbaum does not require that such intuitions are checked according to a veil of ignorance in the original position. The Rawlsian method is certainly one way to reach such an equilibrium for Nussbaum, but, as explained below, it is one of many plausible methods that fall under the scope of *phronêsis* according to *prohairesis*.

prohairesis helps to explain why Nussbaum's list is defensible across cultural contexts, according to either the broader or the narrower characterization.

At any rate, for both of Nussbaum's characterizations the two commitments to value-ladenness and qualitative distinctiveness of different kinds of capabilities hold. Regarding the commitment to value-ladenness, as Nussbaum notes, the capabilities approach "does not read norms off from innate human nature" but instead "is evaluative and ethical from the start; it asks, among the many things that humans might develop the capacity to do, which ones are the really valuable ones, and which are the ones that a minimally just society will endeavor to nurture and support?" (Nussbaum 2011: 28). I focus on the importance of this emphasis on value-ladenness 'from the start' in the next chapter.

Regarding the second commitment, not only are Nussbaum's capabilities "qualitatively distinct" such that "aspects of individual lives cannot be reduced to a single metric without distortion," (ibid: 18), but "all are distinctive, and need to be secured and protected in distinctive ways" (ibid: 35). A person's lack of any central feature significantly inhibits that person's attainment of a flourishing or dignified existence. For this reason, Nussbaum's capabilities are resistant to being rendered so as to provide total orderings of capabilitarian well-being, especially through procedures that treat central features as if they can be traded off against one another. Nussbaum does allow that in the case of what she calls a 'tragic trade-off,' one central feature may need to be traded off for another, say relative to the purpose of selecting among policy options that each exhibit different kinds of central features. However, trading one distinct kind of capabilities for another 'externalizes' the second, reducing it to something other than its particular nature, and can lead to biased policy recommendations that only partially address the relevant situation.

Recall the discussion from chapter one of Aristotle's claim that exchange value is generally a poor measure of the different kinds of well-being values featured by different economic objects. Nussbaum's discussion of tragic tradeoffs follows this Aristotelian emphasis. Let's consider Nussbaum's example regarding how Americans work more than do many Western Europeans, and so have fewer

leisure hours. On the standard economic reduction of leisure into income, lack of leisure can be given a numerical value in terms of extra accumulated wealth, and so Americans might seem well off. Yet the reductive substitution of leisure for income entails that "the full measure of this tragic situation has not yet been taken" (ibid: 39). Reduction to an amount of income does not capture how people value the lost leisure hours in terms of mental, social, physical, or other kinds of well-being, especially when such losses put them below the threshold for what could be called a minimally-functional human existence. Similarly, the challenging work of caretakers - often women, often unpaid - is marginalized when it is either externalized as leisure or is under-represented as deflated wages that are 'determined' by the contemporary labor market, including the surrounding politics that govern wages. Inasmuch as it supports the qualitative distinctiveness of different kinds of central features of capabilitarian well-being, Nussbaum's characterization supports the commitment to avoiding procedures that treat such features as if they are intersubstitutable.

3.3 Nussbaum's Representation: The Open-Ended and Multiply-Realizable List, Tragic Trade-offs, and a Prelude to Phronêsis as Prohairesis

In Nussbaum's *representation* what would seem to be a panoply of different functionings and abilities – perhaps akin to Sen's sets of co-possible sets - is instead posited as an objective list (OL) of ten distinct central features – also called 'spheres' - of human experience. Each of the spheres outlines what Nussbaum calls a 'thick vague' definition of that kind of human experience, the content of which has been refined over time from Nussbaum's 'broader' characterization of a flourishing or *eudaimonic* existence to the 'narrower,' politically-oriented characterization of a life of dignity (see Table 1 below). Each central feature can also be defined according to a threshold of sufficiency, established with different levels of precision relative to purpose and context. For example, thresholds can be determined by a population as some level of generic functionings deemed minimally sufficient for achieving a flourishing existence, say as basic levels of health and education ensured and made accessible to all citizens. Or thresholds can be rendered in terms of more specific functionings relevant for individuals or smaller groups, for example functionings such as availability of different kinds of vocations, being able to have certain kinds of relationships with other people, having access to certain political capacities, etc. I return to a discussion of thresholds later in the section. For now, let's look to how Nussbaum has refined her objective list representation:

Table 1: Aristotelian Domains of Human Experience, and 1992 and 2011 Capabilities Lists

Uniquely Human Spheres (1992):	**Human Functional Capabilities (1992):**	**Central Capabilities (2011) :**
[1] *Mortality	Not dying prematurely/ adequate life expectancy.	**Life :** Not dying prematurely.
[2)] *Human Body (Hunger, Shelter, Sex, Mobility)	Good health, lack of disease, mobility, lack of fatigue, etc.	**Bodily Health :** Adequate health, nourishment, shelter.
[3] *Capacity for Pleasure/Pain	Abilities to avoid unnecessary or nonbeneficial pain, defined contextually.	**Bodily Integrity :** Freedom of movement and from violent or sexual assault. Freedom to pursue sexual satisfaction and reproduction.
[4] *Cognitive Capability	Use the five senses in culturally important ways.	**Senses, Imagination and Thought :** Imagine, think and reason through science, music, literature, defending freedom of expression, freedom of speech, avoiding nonbeneficial pain.
[5] *Early Infant Development	Attachments to external things and persons.	**Emotions** : Ability to have attachments to people/things outside ourselves. Emotional development not hindered by fear/anxiety.
[6] *Practical Reason (NE I, II)	Form a conception of good, engage in critical, action-guiding reflection.	**Practical Reason :** Form a conception of the good, engage in critical, action-guiding reflection and planning.
[7] *Affiliation with Other Human Beings	Live for and with others, interact, show empathy. Also, have social bases of self-respect and non-humiliation.	**Affiliation :** **(A) :** Living with and for, interacting with others, showing concern and ability to imagine one another. **(B) :** Having social bases of self-respect, being treated with dignity and nondiscrimination.
[8] *Relatedness to Other Species, Nature	Live with concern for animals, plants, and the surrounding environment.	**Other Species :** Live with concern for and in relation to animals, plants, and the world of nature.
[9] Humor and Play	Laugh, play, and enjoy recreational activities.	**Play :** Laugh, play, enjoy recreational activities.
[10] Separateness	Live your own life in your own context.	**Control over One's Environment :** **(A) Political :** Right of effective participation in political choices governing one's life. **(B) Material :** Equal rights to holding property, equal employment rights, ability to works as a human being exercising practical reason, relationships of mutual recognition.
*Spheres of human experience and choice presented in Nussbaum "Non-Relative Virtue"s 1987. (NE II.7)		

To get a sense of how the Aristotelian view of well-being grounds at least the earlier emphasis of Nussbaum's representation, let's consider row [4] above. From her exegesis of the virtues presented in Aristotle's Nicomachean Ethics along with analysis of culturally-relevant literature such as the Greek tragedies and histories, Nussbaum establishes that a flourishing human life must be one that allows for – *inter alia* - sufficient 'cognitive development' such that the organism is able to engage, in a self-determined manner, in cognitively-supported relationships with both her internal and external world.[26] An earlier (1992) outline of this sphere is that we have the basic capability to develop our senses in culturally-important ways. Note the deliberate vagueness of this outline. It allows for open-ended interpretation, especially in deliberation, regarding what are the appropriate kinds and sufficient thresholds of cognitively supported relationships within a given culture.

The second (2011) outline fills out the sphere with more specific basic capabilities and functionings, for example 'reasoning through science' and, in a more political sense, 'defending freedom of expression.' Here we can see the refinement of substantive content relative to the purpose of Nussbaum's 'narrower' characterization and its less overtly Aristotelian support for a dignified rather than a flourishing existence. Yet even as Nussbaum's characterization has shifted in practical scope, the orientation of the outlined content has not undergone major conceptual shifts except in domains [3], [5], and [10]. As far as measurement is concerned, Nussbaum's representation maintains the same general features, including commitments to value-ladenness and qualitative distinctiveness, in 2011 that it did in 1992 or even 1987.[27]

Regarding the commitment to value-ladenness, Nussbaum's objective list is clearly value-laden in that some sufficient level of each central feature is needed in order to lead a life valued as flourishing or dignified. Values also influence the setting of thresholds for each central feature, for example thresholds of educational achievement that might, on average, be taken to represent a sufficient level of [4]-type

[26] Nussbaum argues that literature from *any* cultural perspective is fair game for analysis. The point is that the analysis helps to broaden and challenge our intuitions regarding the set of central features that constitute human well-being as measured by capabilities. See, for example, the quote on the Cyclopes just below.
[27] See Nussbaum (1987) 'Non-Relative Virtues.'

cognitive functioning. Values furthermore influence the prior *selection* of the central features. As Nussbaum explains in her earlier (1992) "Defense of Aristotelian Essentialism," when we look to the functioning of the human being, we can look not only at our own intuitions or the set of expressions and reasons others give in deliberation, but we can also look to literature:

> [consider] the myths and stories that situate the human being in some way in the universe: between the beasts, on one hand, and the gods, on the other; stories that ask what it is to live as a being with certain abilities that set it apart from the rest of the living beings in the world of nature, and with, on the other hand, certain limits that derive from our membership in the world of nature
> (Nussbaum 1992: 215).

Literature provides a basis for recognizing qualities that are uniquely human. Consider how the Cyclopes of Homer's *Odyssey* are not quite human because they lack "all sense of community and affiliation" (ibid: 216), a distinction that help us to define central feature [7] on the list, given that it separates a human life from one that is not. Further contrasts and philosophical arguments together start to provide the kind of outlines that make up the list.

Note that Nussbaum's support for literary analysis and philosophical argument is vulnerable to the JTOBPC-type, relativist challenge presented in previous sections: can such analysis really contribute to a list that holds *across different cultures*? As the title of Nussbaum's 1992 paper indicates, Nussbaum believes that there is an 'essentialism' that manifests when we attempt to define the different central features on the list, especially through the appropriate kind of deliberation. Regardless of one's cultural background, a reflective, ideally collaborative inquiry that includes inter alia discussion of literature, myth, tradition, and personal intuition generates a basic picture that corresponds – with some contextual variance - to the second column of Table 1.

Here we can start to see, on account of its own open-endedness, how Nussbaum's representation features a commitment to pragmatism: the list provides the basis for its own falsification in that one only needs to provide sufficient counter-evidence - say from literature, experience, or even personal reasons that are informed by deep intuition - in order to overturn a proposed outline for a central feature. Even so,

Nussbaum's view has been challenged as 'perfectionistic' in that it relies too heavily on culturally-specific normative content, and so the outline arguably won't hold across a wide array of contexts.[28] Recall the concern from chapter one that a life of *eudaimonia* is a concept introduced by one philosopher whose views are steeped in a single culture (ancient Greece) and so that concept cannot be expected to capture the wide variety of human lives worth living. Regarding the central feature of 'affiliation,' for example, individuals in some cultures prefer large community gatherings and three-day wedding festivals, while others prefer semi-solitary walks in the forest, etc. How could a general outline of a single feature hold for such a wide range of practices?

Furthermore, the relativist might find it suspect that the second list (2011 in Table 1 above), developed according to a wide-scoped and metaphysically-thin sense of human dignity, is largely the same as the earlier list (1992 in Table 1 above) based on a mono-cultural account of Aristotelian *eudaimonia*. I argue we should not be surprised by the overlap. If the method first offered by Aristotle does what it is supposed to do, the same vague outline should hold whether we base it on a deeply reflective account from one culture or some set of metaphysically thin intuitions from multiple cultural perspectives.[29]

I offer two further rebuttals to the relativist's challenge that Nussbaum's list is too perfectionistic. Firstly, Nussbaum directly characterizes her list as open-ended, meaning it can and should be re-interpreted. Let's consider one of Nussbaum's many defenses of the open-endedness of the list:

> Since the intuitive conception of human functioning and capability demands continued reflection and testing against our intuitions, we should view any given version of the list as a proposal put forward in a Socratic fashion, to be tested against the most secure of our intuitions as we attempt to arrive at a type of reflective equilibrium for political purposes… Some items on the list may seem to us more fixed than others… the list remains open-ended and humble
> (Nussbaum 2000a: 77).

[28] See, for example, arguments from Claassen (2018) regarding the perfectionism of Nussbaum's list. See also concerns from Arneson (2000) regarding "Perfectionist Capabilities."
[29] In fact there is much evidence that validates the thick-vague outline. See, for example, the results of surveys done by Wolff and deShalit in Disadvantage (2013) or the excellent summary of various list-based accounts done by Alkire (2008a) in *Choosing Dimensions: The Capability Approach and Multidimensional Poverty* esp. tables 1-5.

Note that the number and exact outlines of central features are not set; Nussbaum's version is a revisable – if well-defended - benchmark. Moreover, the list can *and should* be altered by checking it against our 'most secure intuitions,' ideally through processes of deliberation that respect the relevant purpose, whether at the level of an individual life or in the aggregate, across lives.

As the quote indicates, Nussbaum envisions deliberation as a collaborative discussion in a 'Socratic fashion' rather than - as a contrast - a zero-sum debate. In an earlier paper on "Internal Criticism and Indian Rationalist Traditions" Nussbaum and Sen claim that cross-cultural deliberation involves not only some consensus-building "commitment to non-contradiction," but also commitment to toleration, including a pluralist "protection of the right to diverse choices of the good" (Nussbaum and Sen 1987: 30-31). While Nussbaum recognizes that values do conflict, deliberation need not be framed as 'winner take all.' Consider, for example, the way a traditional conservative might value state supported voting practices in 2020 versus the way a vocal member of the ACLU might value them. As discussed in the previous chapter, pluralist deliberation allows clashing sets of values to survive critical scrutiny, especially in non-ideal situations that call for ranges of political compromises. Moreover, especially in cases in which traditional values might support suppression of persons on the basis of socially-constructed categories, tolerance in deliberation provides a space for the autonomous development of the abilities of individuals to recognize central features and other functionings as valuable. A key feature of this kind of collaborative process is that agents of pluralist deliberation should be provided with opportunities to develop their own understanding in light of best available evidence and argument.

Nussbaum argues that when it proceeds in this Socratic fashion, a collaborative inquiry at the level of selecting relevant central features forms a stable - if open-ended - representation across cultures.[30] Here we can also recognize the feminist push from Nussbaum's position. Some parochial practices, for example female genital mutilation (FGM) or exclusion from educational opportunities on the basis of gender, will be vigorously defended from the perspective of culture A against that of culture B. There

[30] I return to this next point in section five regarding the principles of pragmatic inquiry.

would then seem to be an impasse regarding how to propagate the list, as culture A might hold that certain central features just are not relevant for different socially-constructed categories of personhood. Here, however, an open-minded inquiry provides a space for recognition of deeper values that can shift perspectives. For example, community leaders, when made aware of the particular details of FGM, often turn against the practice, changing their valuations. Similarly, the expression of the injustice of limitations of capabilities on the basis of gender or other social identity backgrounds provides a space for inquiry and even challenge where before there was tacit suppression or general ignorance. As I clarify in section five from a pragmatist perspective, what matters from a Sen/Nussbaum perspective regarding the selection of functionings is that the relevant voices are given a forum and influence, and that both sides are willing to attempt to understand the deeper reasons behind the valuations that are expressed within deliberation.[31]

Secondly, although Nussbaum's list does have a general form that, she argues, holds in an essentialist sense across different cultures, the relevant interpretation of each central feature is multiply realizable. For example, recall from chapter one that the virtue of pride has value when expressed - and recognized - according to a somewhat aggressive sense of Greek *megalopsuchia*, yet the same virtue also has value when expressed and recognized according to a contemporary Christian sense of humility.[32] Though cultural expression differs, the root virtue is the same. In a similar fashion, let's re-consider the wedding festival versus walking in the woods example above and acknowledge that what it means to 'affiliate' with other human beings (sphere [7] in Table 1) in a dignified manner certainly entails the setting of culturally-dependent conditions satisfying that particular end.[33] However, though the exact set of functionings that constitute sphere [7] might differ with respect to different cultural values, the root-level outline of the central feature can still hold across cultures. For example, despite the immense array of relevant relationships that an anthropologist might point to as indicative of 'affiliation,' we can intuit

[31] See Nussbaum Women and Human Development (2000a) pp. 34-59.
[32] See chapter one section three.
[33] See Comim et al. (2008: 166-167) and Nussbaum (1992: 24-26). It is possible according to Nussbaum's theory that even the OL dimensions themselves may change slightly depending on context, as suggested by consideration of Megalopsuchia. See Nussbaum (1987b: 11).

that a common basis holds, in a rough sense, as something like the sharing amongst humans of space, gestures, ideas, perspectives and emotions, etc., perhaps with a tendency to engage in collective practices that take on roughly similar, if vague, forms across cultures. That rough sense is the starting outline of some sphere of experience that grounds the relevant central feature.

One notable implication of multiple realizability is that different spheres may exhibit the same kinds of features in different contexts. For example, the attainment of a certain degree of practical reason ([6] in Table 1) may overlap with or even be defined according to other aspects of human affiliation ([7]), cognitive capability ([4]), or any number of the other domains of the list in Table 1. This overlap provides another argument against the charge that the list is perfectionistic, as the same rough set of valuable functionings and abilities can be established according to different structures. The list might, for example, be populated with a different number of central features, say six instead of ten.[34] And even then, such features are themselves multiply realizable.

As a further response to the relativist challenge, however, I note that though constitutive functionings of central features may overlap, once central features have been defined at the representational and operational levels by procedures of collaborative inquiry (and ideally deliberation) that are sensitive both to context and purpose, they then remain *distinct* for purposes of measurement *in that context*. The conceptual overlap that potentially exists among central features – say at the level of intercultural consideration of how the list might be formed - certainly does not allow the relativist to justify the in-practice rendering of central features as if they are intersubstitutable, say by procedures that always provide total orderings of capabilitarian well-being. Once they have been selected, central features remain qualitatively distinct. That there can be overlap and yet qualitative distinctiveness might seem puzzling. To get a better understanding, let's turn to discussion of how qualitative distinctiveness really features in and is supported by Nussbaum's representation.

[34] See, for example, a helpful survey of these kinds of capabilitarian lists by Sabina Alkire (2008) in "Choosing Dimensions: The Capability Approach and Multidimensional Poverty."TRAVIS—diodn't you already say this in a previous ftnote

As explained in earlier sections, Nussbaum's commitment to the qualitative distinctiveness of central features entails that the different kinds of values constitutive of or relevant to those features are resistant to being rendered in terms of one another, especially as if they can be traded off against one another. Nussbaum does recognize, following Sen, that as a practical matter such features can often be compared, for example with respect to some urgent need or a limited set of policy options, etc.[35] And so a familiar JTOBPC-type argument can be given to justify procedures that always give total orderings of capabilitarian well-being, perhaps as a practical matter. However, Nussbaum's representation and characterization highlight a further incoherence for such procedures. For Nussbaum, following Aristotle, a life lacking in any central feature faces significant if not insurmountable barriers to achieving a flourishing or dignified existence.[36] When they are below the relevant threshold, central features – and, by extension, indicators standing for them - are not even comparable, nor should they be treated by a measure as if they are intersubstitutable. In other words, there should be no fact of the matter regarding how to render the lack of different kinds of central features into a total ordering of capabilitarian well-being.

Consider the following argument by analogy: looking at an airplane, could one claim that features essential to the wing hydraulic system can really be substituted for features essential to the computer control system? If either system is below the relevant threshold for functioning, the plane will no longer fly. What purpose can there be, then, for the comparison? To claim that features can be substituted, or, similarly, that weights can be assigned to different features as a means of expressing relative importance with respect to the overall 'excellence' of the plane, is to deny the deep necessity of each central feature. Note that we can more readily compare and perhaps even rank ancillary features of the plane against other central features and, in some cases, against other ancillary features. For example, we can plausibly give some very low weight to the value of the in-flight television system versus that of the computer control system, and we could perhaps rank the value of the television system against that of having windows for

[35] See further below in this section and also the last section of the previous chapter regarding this distinction.
[36] See Nussbaum (1986b).

passengers, etc. But central features that are *necessary* for proper functioning – in this case, for safe transit through the air – tend to remain non-intersubstitutable inasmuch as we seek to measure their value relative to that functioning.

Similar difficulties become apparent, mutatis mutandi, for comparisons of infant development versus health, affiliation versus cognitive capacity, or other isolated comparisons of Nussbaum's central features given in Table 1 in the previous section. Each central feature of capabilitarian well-being represents a deeply-needed kind of human excellence, the lack of which prevents the attainment of a flourishing or dignified existence. Accordingly, on Nussbaum's view, sub-threshold *comparisons* of capabilitarian well-being should be limited to the same central feature when viewed through the same cultural lens, while sub-threshold comparisons across different central features should be avoided. There just is no fact of the matter regarding some total orderings made on the basis of such comparisons.[37] This is what Nussbaum means by calling each central feature 'qualitatively distinct.'

Nussbaum offers further support for the commitment to qualitative distinctiveness of central features by discussing more practical cases of what she calls 'tragic tradeoffs,' in which agents must choose to support one central feature at the expense of another. Let's return to the now familiar example of Annapurna, who must give a single gardening job to one of three unemployed laborers: Dinu, the poorest, Bishanno, the most depressed, or Rogini, the most physically unhealthy. To create a better correspondence between workers and dimensions on Nussbaum's list, let's add that Dinu[a] has a young child who needs proper nourishment. The situation of Dinu[a] strongly resonates with dimension [5], Bishanno's with [9] and perhaps [7], and Rogini's with [2]. Granted, there could be significant overlap among dimensions such that the situation of each laborer arguably could be evaluated according to many other dimensions on the list, especially when we consider that the list is multiply realizable. Even so, this example demonstrates at least that there are different *kinds* of central features, as we can intuitively feel the different kinds of deep values relevant for each laborer's situation. We can also see the tragic nature

[37] See for example Enrica Chiappero-Martinetti, "Complexity and Vagueness in the Capability Approach: Strengths or Weaknesses?" in Comim et al (2008: 268-309).

of the choice that Annapurna must make; no matter what, awarding the job to one laborer creates different kinds of significant losses for others such that it does not seem Annapurna can truly trade off the gain experienced by one laborer getting the job against the lack experienced by either of the other two.

The example also indicates that many attempts to represent the relevant set of values according to total orderings of Nussbaum's capabilitarian well-being would be incoherent insomuch as they would involve treating one deeply-needed central feature as if it can be compared with and even substituted for another. Let's acknowledge, as in previous chapters, that a choice could certainly be forced by means of one or another reductive abstraction, say for example by reducing the situation to a single grade utilitarian metric of pleasure, desire, or preference and then optimizing. Yet, as discussed in the previous chapter, Sen (1980) argues that a single-metric utilitarian basis often overlooks valuations - especially those constitutive of or relevant to different kinds of capabilities - that are influential in context. In fact, Sen (1980) supports a disaggregated, 'plural' view over secondary evaluation processes that can always produce total orderings of well-being.

As has also been explained in previous chapters, one common method by which procedures can always produce total orderings of capabilitarian well-being is a decision-making method that weights different elements of a vector view representation and combines them, often by treating those different elements as if they are intersubstitutable.[38] In some cases the method may proceed according to the assumption that some shared feature sufficiently functions as what I have called a 'covering consideration' following philosopher Ruth Chang's (2017) use of the term.[39] For example, such a method could certainly be applied to Annapurna's difficult decision: Rogini's improved physical health does not seem any more or less valuable than Bishanno's mental health in terms of what better contributes to a flourishing human existence, but the two could be compared according to some covering consideration, perhaps the amount of QALYs created by rewarding the job to either laborer. Yet if Annapurna – or the relevant deliberative body - makes the decision that Rogini's improved physical health outweighs

[38] See, in particular, chapter one section one.
[39] See previous chapter section five.

Bishanno's improved mental health (or vice versa), the fact remains that one worker remains in a tragic situation even when the abstraction indicates the most 'efficient' solution has been chosen. The gain does not truly compensate the loss. This is the practical problem that follows from representing different kinds of central features of capabilitarian well-being as intersubstitutable in the context of making tragic policy decisions.

However, further factors – some of them practical - affect whether such incoherent procedures might still be alleged to provide an approximation that is roughly useful. Consider that the situation of Dinu[a] does have greater pull on our intuitions; the proper development of her young child has a longer-lasting effect in terms of years well-lived – say again as QALYs – than does the benefit given to either of Bishanno or Rogini, assuming they are adults of similar age. In fact, the more obvious choice here to award the job to Dinu[a] supports a pragmatism that allows for recognition of *all possible values* relevant to that context including, for example, the effects on the life of Dinu[a]'s child.[40] As explained with the airplane example above, it may be the case that some necessary features can more readily be compared with and perhaps even traded off against other ancillary features, and so some contexts might exist for which incoherent procedures really are roughly useful. I'll return to this caveat just below, but the example supports the claim that some kind of weighted substitution among central features – perhaps in terms of a context-sensitive covering consideration - can seem sufficient for some purposes.[41]

Note that if we alter the situation slightly such that Dinu[b] has no young child but instead has missed out on a primary-level education (dimension [4]) and could use the wages to get that education, we can again see from the lack of any obvious choice the intuition that deprivations of different central features are of qualitatively different kinds. Furthermore, because of their tragic nature, deprivations of

[40] Chang (2017) makes a similar point that even 'on a par' items are amenable to rough comparisons. To wit, we may prefer the full slice of apple pie to the thimbleful of lemon sorbet, even if both desserts are 'on a par.'

[41] See, again, previous chapter section five and the example of using caloric content as a rough covering consideration for 'nutritional value' of different food items to be distributed. As Ruth Chang (2017) notes, items that are not comparable because they are features of hard choices that are 'on a par' can still sometimes be compared by shifting the choice situation, often with respect to a further 'covering consideration.' I use the phrase 'not comparable' rather than Chang's 'incomparable' or 'non-comparable' because Chang's argument against comparabilism stands outside of the scope of this thesis. What I claim is 'not comparable' would, however, best fit with what Chang calls 'on a par.' Whether they are 'on a par' or, to use other authors' terminology, 'incomparable,' the relevant point is that the capabilitarian well-being of different individuals – or, in the aggregate, the capabilitarian well-being of different populations – is not comparable in the sense required to support total orderings of capability states for all sets of states.

different kinds of central features should not be comparable. Yet the examples just above have also indicated that – at least relative to certain practical factors – some comparisons and even substitutions might manage to track our deep intuitions regarding capabilitarian well-being. Here we can re-invoke the familiar refrain from JTOBPC-type practitioners who ignore, for practical reasons, the commitment to avoid incoherent procedures that always provide total orderings. Tragic or otherwise, a choice must be made, and that choice *at least* can be used to rank options featuring different kinds of capabilities, so why not use a weighted index or some similar measure that always provide total orderings, often by treating different kinds of capabilities as if they can be traded off against one another?

Yet, as argued earlier, the incoherence of such procedures with respect to Sen and Nussbaum's characterizations and representations at least warrants further investigation regarding the significance of the associated danger that the measurement outputs may be off-target. As Nussbaum notes, "it will be crucial to represent in the weightings the fact that each and every one of a plurality of distinct goods is of central importance, and thus there is a tragic aspect to any choice in which citizens are pushed below the threshold in one of the central areas" (Nussbaum 2000a: 81). Consider, as a further example, a community or municipality attempting to measure the capabilitarian well-being afforded by different policy options. The awarding of some resources, say to improve schools or increase commercial output, might significantly reduce affordability of living spaces for low-income residents or, as the area becomes more gentrified, reduce access to medical care, etc. If a growth-oriented policy agenda improves some dimensions but others remain below the social minimum, Nussbaum argues that policy makers should focus on how to address the tragedy *rather than* allow that the improvement of some set of central features outweighs or, even worse, substitutes for the lack of improvement of others. In other words, the incoherence at least warrants further investigation given the significance of possible harms that can result if the measure is off-target.

Nussbaum's point about recognizing tragedy is grounded by a philosophical investigation and defense of how people can make ethical decisions that are best informed by the relevant set of values.

Nussbaum emphasizes a particular view of *phronêsis* according to *prohairesis*. According to this view, practitioners should make transparent that use of incoherent procedures, even in urgent cases, can lead to omission of the expression, recognition, and consideration of important values that would otherwise influence the relevant decision. For example, when Annapurna chooses to award the job to Dinu[a], it is important that Annapurna – or, by extension, the relevant deliberative body – understand and seek to address the further suffering of Rogini and Bishanno. As is the case in so many moral dilemmas, a further option often emerges once the relevant values have been recognized. As is discussed in the next section, there is a shared *experience* of deliberation that tracks the set of values relevant for best navigating the tragic decision.

The deceptive ease with which the representation of capabilitarian well-being as total orderings allows the policy maker to act as if some surrounding values can be ignored is a significant problem as justifies a technocratic kind of politics rather than the kind of pluralist deliberation that both Sen and Nussbaum support. If tragic decisions are like Gordian knots that we need to untangle, the solution is not to simply cut through the rope by using the simplified view, though such an action may, metaphorically speaking, free the knot. Recall from chapter one the distinction – given according to the analogy of the Goode-Homolosine and Mercator projections of the three-dimensional globe - between a descriptively accurate yet incomplete measurement representation and one that is 'collapsed' into a complete picture relative to a more specific purpose.[42] The incomplete perspective is important because, as we take the time to properly untangle the knot, we become more likely to recognize that different kinds of capabilities "need to be secured and protected in distinctive ways" (Nussbaum 2011: 35). On the reductive view, we would not even entertain some of these distinctive ways of securing and protecting what people value. I return to this point in the contexts of Elizabeth Anderson's expressive theory of value, considered in section five.

[42] See chapter one section three and also chapter four section two.

Let me now attempt to clarify what might have seemed 'puzzling' regarding multiple realizability, given that it allows for overlap of aspects of central features at the conceptual level yet supports qualitative distinctiveness once central features have been selected as distinct. Practitioners might look to the open-endedness and multiple-realizability of Nussbaum's list and argue that the interactions of different kinds of constitutive values, external conditions, norms, and functionings are simply too complex to use as a basis to create some *n*-dimensional list. Given that different constitutive functionings will overlap and that, moreover, further values such as equality or justice are relevant for the selection of central features, how can we justify the qualitative distinctiveness of those dimensions? In response I say that those dimensions are justified as distinct because, even though the list is open-ended, they have been *recognized through* the proper kind of deliberation as those that are deeply needed for the relevant kind of dignified or flourishing existence.

What the dismissive practitioner neglects to consider is that purpose and operational procedure have the function of *sedimenting* the qualitative distinctness of values relevant for that context. Thus, not only at the micro level of deliberating groups and communities, but even at Nussbaum's intended macro level of, say, populations of entire nations, the list – if generated by the kinds of procedures supported by Nussbaum's (or Sen's) characterizations - will include something like "a sufficient level of cognitive functioning, however outlined" such that it really is a tragedy if citizens cannot attain the relevant threshold, as they have defined it. With respect to more particular populations - say a village or community - the relevant dimension might, for example, contain more specific functionings and thresholds relative to special forms of education like indigenous language and history, relevant vocational training, access to education regarding citizens' rights, etc. Different degrees of exactness hold relative to purpose and context, but this complexity does not amount to extreme relativism in the way that the supporter of the JTOBPC would apply the term.[43]

[43] See particularly chapter one section four.

This point about central features illustrates one further, important element of Nussbaum's representation, which is that the objective list appears ready-made to be translated into a Vector View representation, with each element of the vector corresponding to a different dimension of the list. In fact, it would seem that contemporary measures that represent different *dimensions* of capabilities are not, in fact, following Sen's representation even though they often cite Sen and not Nussbaum. Rather they are following Nussbaum's representation, especially when they measure different *dimensions* of capabilities. Recall that there is little if any discussion of different central features or dimensions in Sen's characterization and representation, and also that Sen's settled representation does not readily (if at all) mirror a Vector View representation. Measures that use procedures that always render indicators standing for different dimensions of capabilities into total orderings of capabilitarian well-being are generally starting from Nussbaum's characterization and representation. Inasmuch as they do so, practitioners using such measures must recognize the strong case put forward by Nussbaum against treating different dimensions as if they are intersubstitutable, especially when such tradeoffs are below relevant thresholds.

3.4 Keeping Operational Procedures Coherent: Phronêsis according to Prohairesis as 'Desiderative Deliberation'

Nussbaum's 1985 book The Fragility of Goodness and her 1986 paper on "The Discernment of Perception" are two investigations of Aristotle's conception of well-being that together provide a philosophical defense of capabilitarian well-being as a set of qualitatively distinct kinds of goods. Nussbaum explains that the Aristotelian view refutes what she calls a Platonic, universal conception of the good, introduced in chapter one.[44] The Platonist argues that well-being has, deep down, a homogenous essence that can be measured on a single grade.

Nussbaum explains that in Plato's Protagoras, the character Socrates defends a rational, quantitative science of the universal good that can be used to measure the amount of the good for every particular case. According to the Pythagorean epistemology that underlies Plato's cosmological view, universal forms can be grasped only by means of numerical expression (Nussbaum 1985: 106). Plato's Socrates claims that a rational *techné* (science) of the good allows moral agents to mitigate *tuché*, or brute luck, because moral agents can use the *techné* to give numerical values to the good, turning practical decision-making into a matter of universal degree rather than particular kind. Recall, for example, Annapurna's difficult decision regarding how to award the gardening job. By means of this universalist *techné*, agents such as Annapurna have an exact moral compass that can be used to navigate difficult moral choices, for example by choosing the option that scores best (ibid: 90). Nussbaum argues that the problem for the Platonic view is that the moral *techné* at best pans out as a metric of pleasure, perhaps something more refined than simple Benthamic pleasure, but a mono-concept nonetheless (ibid: 98).

Nussbaum's Aristotle defends a more context-sensitive, heterogeneous view of the good, as summarized in chapter one. Aristotle's process of prudent decision-making, or *phronêsis*, is committed to allowing that different kinds of goods each have their own essences and must be considered in context (ibid: 296). On Nussbaum's view, an agent of *phronêsis* can recognize the relevant set of moral goods

[44] See also chapter one section four.

according to a complicated form of *prohairesis* that combines both intellectual and emotional understanding to gauge each situation (ibid: 307). Nussbaum calls this process "either desiderative deliberation or deliberative desire," meaning agents of *phronêsis* deliberate according both to reasons and to the more desire-driven – what Nussbaum calls 'passional' – component of their beings (ibid: 308).

Like many animals that interact in social orders, humans register and act according to a mixture of deeply intuitive emotions and thoughts that have been honed by experience to be sensitive to particular contexts. A process of habituation allows that the organism becomes able to efficiently - almost automatically - grasp the more prudent option. Humans who are better able to make prudent decisions – who act according to *phronesis* – do so at least partly in accord with *prohairesis.* Ergo, on Nussbaum's view, a "well-formed character is a unity of thought and desire, in which choice has so blended these two elements, desire being attentive to thought and thought responsive to desire, that either one can guide and their guidance will be one and the same" (ibid: 308). Nussbaum notes that even our imaginations, which we use to consider the experiences of others and hypothetical consequences of different actions, operate according to this complex basis of *prohairesis*.

Nussbaum's account of *phronêsis* according to *prohairesis* illustrates how valuation and deliberation should proceed at the operational level of measuring capabilitarian well-being. Deliberation according to *prohairesis* includes recognition of deep moral ends, or values, including relevant central features of capabilities (ibid 284). These values, following Ruth Chang's 'ordinary' sense of the term introduced in chapter one, are often recognized at the level of deep emotional intuition, much in the same way that animals recognize salient features of a situation to make almost instinctive or 'natural' decisions that serve deep ends related to survival and communal interaction. Consider, for example the way that most animals have a deeply intuitive, arguably innate sense for how to care for their offspring, even in novel situations. Many animals also intuitively make decisions on behalf of or along with their pack, flock, community, etc. Combinations of deep emotion, instinct, and action all appear to be at work when fish school, when geese or butterflies migrate, when wolves hunt together, etc.

Against the intuitions that support the *techné* defended by Plato in the Protagoras, Nussbaum argues that a detached, objective perspective lacks access to the same kind of content that we get from *phronêsis* according to *prohairesis* (ibid: 291). Nussbaum further argues that "to effect the commensurability [homogeneity] of values is to do away with them all as they currently are, creating some new value that is not identical to any of them" (ibid: 296). As demonstrated in the previous section, the detached – and often reductive – perspective of the *techné* is unable to recognize the specialized, context-relevant ways that different kinds of values pull on and direct our decision-making. The abstract representation often unhelpfully *cuts through* that which should be carefully untangled.

As a preface to the next section, I note that other authors defend a similar view regarding the practical importance of *untangling* the gordian knot rather than simply slicing through it. As we shall see, Elizabeth Anderson's expressive account certainly stresses the importance of recognizing different kinds of values as a guide to making decisions. Another supporting account is provided by Sir Stuart Hampshire in his short (1977) essay "On Having a Reason." Hampshire argues that agents of *phronêsis* don't take the time to consciously recognize and weigh every value that might apply in a given situation. In many cases, there simply is not enough time; values are recognized as the appropriate basis for action because agents have what Hampshire calls pre-developed 'schema' that operate on a cognitive level different from that which forms strings of propositional content. Hampshire argues that even when we attempt to create a rational string of propositions meant to capture the thought process that underlies one or another complex decision, that string is at best an abridgement, what Hampshire calls a 'Cartesian error' (Hampshire 1977: 88). By extension, abstract analysis according to some Platonic mono-concept constitutes an even more egregious error because it further reduces the complex kind of processing that is relevant in that decision-making moment.

Hampshire goes on to explain that it *is* possible for agents of *phronêsis* to engage, post-decision, in a reflective exercise in which they find *reasons* that express what they valued such that they acted in such and such a way (ibid: 89). Recall from the previous chapter that this exercise follows the way that

agents of deliberation can, on Sen's view, express reasons for valuations given in deliberation. Hampshire describes a reflective process of slowly untangling the different values, say in order to determine what *would have been* the best action in a dilemma that one recently faced. Hampshire's example is that of a military officer who is unsure how to act among subordinate officers in a bar, but we could just as easily consider a group deliberating over Annapurna's decision. In those kinds of moments, agents of deliberation need to have a deep respect for the ways that different values are expressed in context, some of which cannot without significant reduction be captured as a sequence of propositional expressions. Here Hampshire makes a relevant point that parallels the emphases of both Nussbaum and, in the next section, Anderson. We "ought not … to pursue an unattainable ideal of rationality in practical matters as requiring an explicit weighting of arguments before moral decisions are made and opinions are formed" (ibid: 98). Though it can provide a basis for justifying the total ordering of different options, the moral 'weighting' is problematically reductive.

In order to deliberate well in the context of measuring capabilities, Nussbaum's desiderative deliberator must recognize the relevant ends, often expressed as valuations in the form of deep emotions or intuitions. When groups of agents deliberate, a similar process must be allowed.[45] In order to enable that process, we must recognize the way that the commitment to qualitative distinctiveness *frames* the decision. Nussbaum mirrors Hampshire's position when she argues that "to make objects of desire commensurable [homogenous] is to remove, already, one source of our passional intensity about them" (Nussbaum 1985: 307). If the passional intensity is to be recognized, evaluated, and processed as a relevant component of the measure, then different kinds of deep values must remain qualitatively distinct.

Nussbaum recognizes that, if done well, deliberation through practical reason attains a deep kind of insight not unlike sense perception: "[p]ractical insight is like perceiving in the sense that it is non-inferential, non-deductive; it is, centrally, the ability to recognize, acknowledge, respond to, [and] pick out certain salient features of a complex situation" (ibid: 305). To link *phronêsis* with *prohairesis* is to

[45] For a similar account, see chapter two section two for Sen's characterization of public deliberation, in which agents both express valuation simpliciter and also are able to give reasons for why they value different items.

argue that humans have the ability to *perceive* the situation in a deeply intuitive manner, much in the same way that deliberation can be framed according to schemas in Hampshire's discussion above. Nussbaum also argues that "deliberation… accommodates itself to what it finds, responsively and with respect for complexity" (ibid: 301). If deliberation is to be sensitive to contextual complexity, it must be structured so as to best capture and allow for the influence of particular kinds of values.

From this recognition we can distil an important principle: commitment to qualitative distinctiveness provides the best possible deliberative frame – or schema - for recognizing, evaluating, and measuring capabilitarian well-being. Yet though Nussbaum's philosophical analysis of *prohairesis* provides an argument, we are left, seemingly, without sufficient evidence to warrant the conclusion supported by the principle. There is still the challenge from practitioners who use procedures that are incoherent with respect to the commitment to qualitative integrity, seemingly because such measures are *roughly useful*, especially given certain practical considerations. In response, I turn to further pragmatic analysis regarding what it means to recognize, deliberate over, and even make decisions on the basis of different kinds of deep values that we sometimes express as deeply complex 'knots' of emotion, intuition, principles, and other action-guiding abstracta.

3.5 Further Pragmatic Considerations: Putnamian Inquiry and Anderson's Expressive Theory

In the preceding sections I have explained how Nussbaum's characterization and representation include commitments to value-ladenness and the qualitative distinctiveness of different kinds of central features. According to Nussbaum, agents of *phronêsis* who act according to *prohairesis* are best able to recognize and deliberate over the different kinds of functionings and values that are constitutive of central features, especially when agents allow that central features are distinct, context sensitive, and generally non-intersubstitutable. However, practitioners may argue that even if procedures that transgress either commitment are incoherent with respect to Nussbaum and Sen's characterization, they may be roughly useful for certain purposes, relative to certain practical constraints.

Incoherent procedures that always provide total orderings of capabilitarian well-being – for example those used by the Human Development Index (HDI), the Better Life Index (BLI), and other capabilitarian measures - may be justified by this kind of practical argument. For example, as discussed in previous sections, it may be that the combination of context and purpose only necessitates the comparison of above-threshold central features, or the comparison of some ancillary features to a single below-threshold feature. In both cases, incoherent procedures that always render total orderings are likely to provide approximations that roughly fit with people's intuitions regarding the capabilitarian well-being of relevant items, just as, on Sen's view, procedures that force complete rankings will sometimes produce outputs that sufficiently approximate incomplete rankings. I've also argued that in some cases, practical considerations such as urgency or availability of data relevant to the task to which the measure is lent may justify the use of incoherent procedures, even if the measure is in danger of being off-target.

There are further justifications that can warrant the use of incoherent procedures that always provide total orderings of capabilitarian well-being. In this section I claim, following arguments from Elizabeth Anderson in *Value and Ethics in Economics*, that such procedures may be justified when they are used by 'performance standards' that score capabilitarian well-being of different populations in a

manner similar to how scores are given for competitive figure skaters or decathlon teams in the Olympics. Yet even in such cases it is important that practitioners recognize the 'measure' is still incoherent with respect to Sen and Nussbaum's characterization and consider the practical significance of the possibility that the measure may be off-target.

Before we get to Andersonian performance standards, it is helpful to consider a few further pragmatic arguments that bolster the case for remaining committed to the qualitative integrity of different kinds of capabilities. One such argument is Hilary Putnam's defense of different kinds of values in his essay "The Fact/Value Dichotomy and the Future of Philosophy." Putnam supports pragmatist philosopher John Dewey's claim that, at least in the social sciences, there cannot be an absolute dichotomy between facts and values. Economists often claim that some 'facts' like GDP per capita or total profit for a firm can be measured as separate from other 'evaluative' content such as people's valuations of well-being. Yet the Deweyan claims that there simply are *facts about value* such that values can be measured as facts (Putnam 2017: 28). This claim dovetails with the pragmatist commitment that prudent decision-making is granted through a complex recognition of reason and desire. According to Putnam following Dewey, values are *educated* into human agents by experience – much in the manner that they are habituated for Nussbaum's agents of *phronêsis* according to *prohairesis.* Yet for the Deweyan, 'happiness' can be recognized as an objective 'act' that incorporates the appropriate relationship among the agent and the relevant external objects, rather than existing only in the subjective mental state of the agent (ibid: 36). This commitment to the *fact* of the objective act follows the commitment of Sen and Nussbaum to measuring capabilities according to the opportunities that people have to attain different kinds of *doings and beings* rather than according to their subjective mental states.

The Deweyan, moreover, holds that different kinds of pleasure correspond to different kinds of objective relationships among the agent and relevant external objects, such that "'qualitative differences are not intrinsic, but are due to the different objects with which they are associated (as pleasures of hearing, or vision) … *they disappear when the pleasure is taken by itself as an end"*" (Dewey 1908: 257

in Putnam 2017: 37, *emphasis added*. Note that Dewey argues there are different kinds of relevant ends, as I discuss below). The Deweyan defense of different kinds of pleasures-qua-relationships parallels the way in which Sen's characterization defends the commitments to value-ladenness and heterogeneity of capabilities: doings and beings are valued in diverse kinds of ways because of how an individual connects those functionings to other functionings in her life, producing what Sen calls the particular 'culmination outcome' that must be evaluated holistically as one of many co-possible lives worth living.[46] It also parallels the way that Nussbaum's desiderative deliberator treats different kinds of deep values – or, by extension, different kinds of central features – as qualitatively distinct in order to better grasp the complexity of the entire situation. For the Deweyan, there are different kinds of pleasures that can be thought of as 'ends,' and the Deweyan view warns against *reducing* an objective 'fact' well-being to a mono-concept like a simple state of pleasure; relationships among external objects are relevant as indicators of different *kinds* of ends.

Putnam also defends a set of conditions that together provide what he calls a 'pragmatist inquiry' for deliberation in complex decision-making contexts. Citing both Dewey and pragmatist philosopher Charles Sanders Peirce, Putnam explains that ethical decision-makers must, *inter alia,* (i) seek to make full use of human reasoning, (ii) avoid 'relations of hierarchy and dependence,' (iii) make careful evaluations based on experience that respect new descriptions, (iv) avoid having an absolute 'metaphysical foundation outside,' (v) investigate in a cooperative and democratic manner, and (vi) maintain a consistent commitment to fallibilism, recognizing that the currently-accepted standard is likely faulty (ibid: 39). Together these conditions (i) – (vi) provide a blueprint for a process of deliberation that accords with Sen and Nussbaum's characterizations of processes for evaluating capabilitarian well-being.

For example, much has already been said regarding (iii) the value of experience such that deliberative groups must support 'new' descriptions on the basis of particular experiences. Let's again consider Annapurna's decision. To make a prudent evaluation in terms of capabilitarian well-being,

[46] See previous chapter section three.

Annapurna needs to have had the relevant kind of experience such that she can make the right kind of evaluation. Conditions (ii) and (v) further maintain that the experiences and expressions of those whose valuations are relevant are admitted into and remain influential within deliberation. I return to an extended discussion of this point in the next chapter. Furthermore, the commitment to fallibilism in condition (vi) provides a safeguard against the dominance of one or another cultural or parochial set of norms, as does the injunction against an absolutist metaphysical standard. These two commitments are mirrored by Sen's defense of pluralist deliberation and Nussbaum's commitment to an open-ended and multiply realizable list. In short, the conditions of pragmatist inquiry provide a standard that best allows for the recognition of the appropriate schema within groups of deliberating agents.

The commitments of both Deweyan pragmatism and Percian inquiry provide further support for the commitment to recognizing that there are different kinds of capabilities and capabilitarian well-being values. Yet even granting these principles, there are some cases in which we might be compelled to allow for procedures that render different kinds of capabilities into total orderings, and not merely because there is extreme urgency or a favorable grouping of features in some limited context. In her article "Practical Reason and Incommensurable Goods," Elizabeth Anderson (1997) follows Putnam in offering a pragmatist theory that "asserts the autonomy of practical reason, which takes no orders from a supposedly external structure of values" (Anderson 1997: 108-109). I note that Anderson's definition of 'incommensurability' - itself following from that of Joseph Raz (1986) - is slightly different in terms of intended application from that of the 'non-commensurability' used by Sen and Nussbaum when they argue that one of two 'non-commensurable' phenomena "cannot be conceived as simply a certain quantity of something else" (Sen and Nussbaum 1987: 25). Anderson's theory applies to a wider scope of value-driven decision-making, but does include consideration of values relevant to Sen and Nussbaum's characterization of capabilitarian well-being. As such, Anderson's 'incommensurability' is a sufficient synonym for 'qualitative distinctiveness' as concerns the measurement of capabilitarian well-being.

According to Anderson's Expressive Theory of Rational Choice (henceforth ETRC), goods that have separate intrinsic natures (what practitioners following Nussbaum might call 'dimensions' of well-being values) can be compared simply if there is sufficient reason to compare them, i.e. that the comparison is useful for guiding our actions. At first glance, Anderson's position seems to support that of practitioners such as Chowdhury and Squire who ignore the commitment to qualitative distinctiveness and accept weights used to construct total orderings because such procedures are, indeed, useful for guiding our actions. Anderson explains that in deliberation, pragmatists simply "construct the structure of values through the sorts of reason-giving principles that reflectively make sense" (ibid: 96). JTOBPC-type practitioners might also claim that use of incoherent procedures 'reflectively makes sense' simply because the deliberative body assents, say by majority vote, to the need for such procedures. On Anderson's view, reasons can be given to support the claim that some covering consideration might reflectively 'make enough sense' to justify the transgression of the commitment to qualitative distinctiveness. For example, some central value such as Sen's 'substantive freedom' or Rutger Claassen's (2017) 'navigational agency' might be deemed a common currency sufficiently shared by different kinds of capabilities such that it can be used by deliberators to weight those indicators relative to one another and then apply further procedures that always provide total orderings of capabilitarian well-being indicators.

Yet it would be hasty to claim that Anderson's view generally supports such operational procedures – especially in the context of measuring Sen and Nussbaum's capabilitarian well-being. In the same article, Anderson (1997) does allow that some intrinsic values may admit of different degrees of emotional intensity - say, for example, relative to a common currency - when reflected on by deliberative agents. However, Anderson argues it is a *mistake* to translate degrees that are recognized in reflection into systems of weights that can be used to provide total orderings of items. Anderson's argument mirrors Hampshire's view that though a deliberative agent - say a military officer in a bar - can provide

formal analysis as a process of reflection, that formal analysis is a reduction of the complex process that actually occurs during deliberation.

Anderson claims that reduction of values to intersubstitutable degrees of intensity implements a zero-sum game in which maximization is the goal. However, values have a distributive form in which agents orient themselves toward each relevant value and consider it as a *full particular*, regardless of zero-sum rationality. Anderson gives the example of parental love, which is not oriented towards a zero-sum aggregate of love to be doled out to whichever child, but is distributed to each child given the particular consideration of the child and the moment (ibid: 98). This example parallels the way that Nussbaum would have a deliberative body consider different central features of well-being. To trade one central feature for another according to a zero-sum game – especially if one or the other is below the relevant threshold - is to ignore the deeper sense in which such a reduction is a tragedy. Ultimately, Anderson rejects decision-making based on analyses that render different kinds of values intersubstitutable in terms of a common currency because it provides "no room for the free play of these other motivations" (ibid: 100).

In her earlier (1993) Value and Ethics in Economics, Anderson further argues against a monist view – similar to that of Plato's Socrates - that reduces different standards of value to a single measure "which commensurates all goods" (Anderson 1993: 46). Here Anderson follows Nussbaum in arguing that the universalist *techné* that might be used to guide an agent of practical wisdom would actually provide mis-direction for important cases: "an agent who regarded the meanings of her actions as contributing quanta of value to them, and who then chose to act with maximum expected value, would by this process destroy any meanings she thought she had incorporated into the optimum act" (ibid: 81). The process of deliberation must allow for the recognition of the appropriate kind of particular.

Anderson also offers a now-familiar defense of qualitative distinctiveness: "goods differ in kind if people properly enter into different sorts of social relations governed by distinct norms in relation to these goods" (ibid: 12). This claim follows the Deweyan line of argument for different kinds of well-beings

according to relationships among subjects and external objects; Anderson argues that "treating a good as a particular kind of good is as much a way of realizing and expressing appropriate relations among people as it is a way of properly valuing the good itself" (ibid: 14). Anderson further recognizes, following both Sen and Nussbaum's characterizations, that "a fundamental implication of the thesis that goods differ in kind is that people should deliberate about them in different ways, according to different frames" (ibid: 24). Here we see not only that there are diverse ways to valuate and evaluate goods as indicated by Sen's view, but we can also (again) see the recognition of *prohairesis* that informs Nussbaum's.

In light of these arguments, Anderson's Expressive Theory of Rational Choice provides substantive reasons to retain, at least for purposes of measuring capabilitarian well-being, the commitments to value-ladenness and qualitative distinctiveness. However, Anderson's (1993) discussion of 'incommensurability' does provide justification, with respect to a specific kind of purpose, for decision-making-type procedures that always provide total orderings. Anderson explains that total orderings can always be given in a special set of cases she calls 'Goodness-of-a-Kind Judgments' in which different kinds of valuable components are weighted and rendered into total orderings, indicative of a further kind (ibid: 51). Anderson considers the example of figure skating scores, each of which is a weighted composite of a more subjective valuation of aesthetic performance along with weights given to combinations of athletic feats, for example completing a certain number of spins while jumping (ibid: 52). In a context such as this, the component-weighting strategy tracks the overall *performance* such that individual skaters can be ranked against one another.

Relative to the specific combination of context and purpose, a standard of performance can be holistically taken as a certain *kind*, even though it is a combination of different kinds of components. The set of weighted components can be compared - as measurable items - across individuals; what is then rendered into the total ordering is the performance. Anderson argues that the same kind of component-weighting process is justifiable in other sports competitions – for example Olympic decathlons - "because it serves the functions assigned to it: it fairly, decisively, and precisely ranks different athlete's

performances across different events according to a process that experienced devotees of the practice find reasonable and relevant to what they care about (ibid: 50, Anderson citing Luban 1990)." By extension, Goodness-of-a-Kind Judgments can defensibly stand as measures for a wide variety of standardized performances that feature different kinds of values. The performance standard captures those different kinds of values in the appropriate kind of way relative to that kind of performance.

The procedures used to generate the kind of standard that is used in figure skating or decathlon are similar to the incoherent procedures used by many contemporary capabilities measures to provide total orderings of capabilitarian well-being. Consider, for example, a measure such as the Human Development Index (HDI) as a performance standard providing a 'score' given by weighting different kinds of components – in the HDI case these are education level, health, and income – and then implementing further procedures – like taking a weighted average - to always provide total orderings. Such orderings do provide a score that arguably captures the right kinds of values in an approximate kind of way. It is possible to claim that, in this limited context, the total ordering is the 'reasonable' view of the different participant populations, and so is a roughly useful approximation of capabilitarian well-being, even if incoherent and in danger of being off-target.

However, it must be noted that a performance standard that always provides total orderings will best fit a specialized set of purposes, one exhausted by the range of purposes to which a measure like the HDI can reasonably be lent. As has been noted in previous sections, a performance standard that always gives total orderings is likely to provide outputs in cases for which there should be no fact of the matter in terms of Sen and Nussbaum's capabilitarian well-being. If, for example, the standard were used to inform tragic choices among policies featuring different kinds of capabilities, the forced ordering could justify a tragic allocation of resources that may have been put to better use in terms of capabilitarian well-being. It may be that a better-informed decision could have been made if the relevant deliberative body used the 'incomplete' picture to inform their decision.

I grant that if the purpose involves potential consequences that are less harmful, for example those that might follow as part of an international conversation about incorporating further kinds of values into welfare metrics, a performance standard could be more than roughly-useful even if incoherent and in danger of being off-target. Here it is useful to consider a quote from UNDP Director Mahbub ul Haq, who helped to develop the HDI: "For any useful policy index, some compromises must be made. But such compromises must not sacrifice the professional integrity of the broad picture that the composite index intends to convey" (ul Haq 1995: 58-59). The question for the HDI is what exactly is the 'professional integrity' of the broad picture, and what purposes motivate the way in which the picture is understood?[47] If the ranking is used to foster a series of political conversations, perhaps to provide further information for the general process by which international aid is granted among nations, then there might be some justification for treating the HDI or a similar multidimensional well-being index as roughly useful even if incoherent. This, in fact, is what the creators of the HDI did. According to Mary Morgan and Maria Back (2018) in "Measuring Development: from the UN's Perspective," the HDI played an influential role in a paradigm shift of measuring welfare from single grade GDP indicators to multi-dimensional measures such as the Millenium Development Goals (MDGs) and Sustainable Development Goals (SDGs), both of which have been dominant measures for the previous two decades.

Note, however, that if the index were used to provide a more localized picture, say to inform different kinds of policies such as whether to fund further levels of education, infrastructure, or healthcare, the dangers that follow from the possibility that the measure is off-target are at least sufficient to warrant further investigation. As I demonstrate in the next chapter, there are significant dangers associated with transgressing the commitments to either value-ladenness or non-commensurability, and these dangers should be recognized and considered by the relevant deliberative body. From a pragmatic perspective, it is prudent to try to make judgments in terms one kind of value at a time, rather than to try to reduce different kinds of values into some sort of evaluative abstracta, the kind that can be used to

[47] I'm thankful to José Manuel Roche of OPHI for this point.

always provide total orderings. Anderson notes that "[c]omparative value judgments *using scales limited to specific dimensions of welfare* are considerably more plausible, justifiable, and easier to make than judgments of global welfare (Anderson 1993: 61, *emphasis added*)." As we saw in the previous section, Nussbaum is similarly committed to the qualitative integrity of different kinds of central features.

Chapter Four:

The Representational Theory of Measurement
and Two Extended Demonstrations of Incoherence

4.1 Introduction

In chapter one I outlined how Aristotle's conception of well-being provides a philosophical background for Sen and Nussbaum's commitments to the value-ladenness and the qualitative integrity of different kinds of capabilities, which I discussed in chapters two and three. In this chapter I further explain how measures of capabilitarian well-being can be incoherent with respect to either of those commitments such that they may be off-target, meaning there is good reason to believe they may give results that are at odds with what is actually the case in terms of Sen and Nussbaum's concept. Because such measures can be used for political purposes that affect people's livelihoods, practitioners should investigate the significance of the possibility that outputs may be off-target as part of an attempt to make the relevant incoherence transparent.

I also give an outline of the *representational theory of measurement* (RTM) that I use as a framework to establish how some measures are incoherent with respect to the characterization or representations (or both) of Sen or Nussbaum. Following this framework, I present formal claims that establish the relevant kinds of incoherence both for conceptual clarification and then distil some guidelines for practitioners working with Sen and Nussbaum's concept.[1] These guidelines generally follow the commitments given above, indicating that practitioners should be wary of procedures that treat different kinds of capabilities as value-free as well as those that always give total orderings of capabilitarian well-being. I then give two extended examples of how incoherence with respect to either commitment can occur. These examples help to illustrate the significance of the ways in which the related measure may be off-target, and also preface the analysis that I give in the next chapter of some contemporary measures of capabilitarian well-being.

[1] See, especially, the final sections of this chapter and chapter five.

The first kind of incoherence manifests when practitioners treat capabilities as 'value-free' or 'value-neutral' at any stage of the measurement process, contradicting Sen and Nussbaum's commitment that capabilities are value-laden. I borrow the terms 'value-free' and 'value-neutral' from Ingrid Robeyns, who argues in *Wellbeing, Freedom, and Social Justice* that, at their most general level, capabilities should first be recognized as doings and beings that are not value-laden, after which they can be evaluated through a secondary process (Robeyns 2017: 44). I show that Robeyns' commitment to separating the 'value-free' descriptive content of capabilities from any subsequent process of evaluation mirrors a similar separation in Sen's (1985a) representation of capabilities, in which Sen treats descriptive 'facts' about sets of functionings as separate from procedures that can be used to evaluate those sets. [2]

Yet because real-world deliberation is always value-laden, any procedure of selection that represents different kinds of capabilities as 'value-free' or 'value-neutral' is contradictory. In his reply to Sen (1985b), Bernard Williams provides examples that challenge this kind of separation, suggesting that it provides opportunities for important kinds of capabilities to be overlooked. I argue, following not only Williams (1985b) but also Sabina Alkire (2006), Morten Byskov (2016), and Chad Kleist (2017) that operational practices or representations that treat capabilities as value-free or value-neutral may be off-target in the sense that they allow for omission of influential values and perspectives within deliberation. Call this '*Value Omission Incoherence*.'

A second kind of incoherence manifests when practitioners implement procedures that can always provide total orderings of capabilitarian well-being, contradicting Nussbaum and Sen's commitment to the qualitative integrity of different kinds of capabilities. Call this '*Total Ordering Incoherence*.' Recall from chapter two that practitioners apply Sen's Vector View representation incoherently when they collapse the Vector View according to what Sen calls a 'secondary' process of evaluation commonly used in decision-making. During that process of evaluation, different elements of the vector are assigned numerical values that are then weighted against one another, after which weighted values can be

[2] See next section for further clarification regarding 'objectively-observable.'

combined (typically summated) by further procedures that assure the measure always provides complete rankings of items. I have shown how such procedures produce rankings for cases for which, according to Sen's representation of capabilities as sets of co-possible sets, there may be no fact of the matter in terms of capabilitarian well-being. Recall also that even if practitioners decide, perhaps as a practical matter, that weights are to be used, Sen argues such weights tend to be given ranged intervals in deliberation. On Sen's view, rankings provided by ranged weights are likely to be incomplete, especially when individual capability sets feature disparate levels of different kinds of functionings.

Procedures that always provide total orderings are furthermore incoherent - especially with respect to Nussbaum's characterization - when they treat different kinds of capabilities as if they can be traded off against one another.[3] Consider again the weight-setting procedures mentioned just above. In a survey of such practices, Decancq and Lugo (2013) follow authors such as Chowdhury and Squire (2005) when they appear to shrug their shoulders regarding which – if any – of the weight-setting methods they analyze accords with Sen and Nussbaum's concept.[4] Yet in their conclusion Decancq and Lugo offer a practical means for determining whether such methods provide even roughly-useful outputs: "the definite test for any weighting scheme should be in terms of its reasonability *in terms of implied trade-offs* between the dimensions" (Decancq and Lugo 2013: 30, *emphasis added*). In his policy research paper "Mashup Indices of Development," Martin Ravallion makes a similar claim for the "need to know the tradeoffs … built into a composite index if it is to be properly assessed and used" (Ravallion 2010: 11). I apply this practical test to the second extended example to further illustrate why, for cases that exhibit Total Ordering Incoherence, practitioners should be concerned how the measure supports unintuitive tradeoffs, especially when – as I have explained in the previous chapter – indicators for those dimensions are below the relevant threshold.

[3] See, for a similar argument, Eric Schokkaert's discussion of the problem of tragic trade-offs related to what he calls the 'indexing problem' within the capabilities approach (Schokkaert in Anand et al. 2009: 552 – 563)

[4] Chowdhury and Squire (2005) offer a similar survey of methods and, like Decancq and Lugo, shrug off the problem of arbitrary weight-setting. See chapter one section two.

In the final section of this chapter, I return to the extended example of Value Omission Incoherence to argue - in keeping with discussion from the previous chapter - that there are other practical reasons to maintain the commitment to qualitative integrity of different kinds of capabilities during evaluative processes of public deliberation. Following Kleist (2019) and also Byskov (2016), as well as Williams (1985b) and Alkire (2006) mentioned above, I argue that practices of deliberation that attempt to give voice to pluralistic perspectives should avoid procedures that always give complete rankings of capabilitarian well-being. Such procedures may prevent recognition of relevant valuations, especially when they prevent deliberators from recognizing the limitations of their own experience and perspectives. I draw from Putnam's (2017) process of 'Pragmatic Inquiry' to further consider how evaluative processes of public deliberation should often provide incomplete or open-ended measurement outputs. Subsequent decisions among options might justifiably be made according to assessments that collapse those open-ended outputs, but such subsequent decisions are no longer processes that measure capabilities as characterized by Nussbaum and Sen.

4.2 The Representational Theory of Measurement, Incoherence, and How We Should Understand Likely Off-Target Capabilities Measures

As a means of analyzing different capabilitarian measures of well-being, I apply the so-called '*representational theory of measurement*' (RTM) that has become the standard in philosophy of science following the significant work by philosopher Patrick Suppes and social scientist Duncan Luce, psychologist S.S. Stevens, and other philosophers of science. As Vadim Keyser says in a recent paper on measurement and experiment,

> Early accounts of measurement focus on the systematic assignment of quantitative values (e.g. numbers or vectors) to objects in the world… Philosophers of science refer to this tradition as the 'Representation Theory of Measurement' (RTM). The general focus of RTM has been to analyze the conditions for representation in measurement by looking at the relation between numbers, instruments, scales, and magnitudes.
> (Keyser 2017: 1)

There is a vast literature explaining the RTM, illustrating it, and using it to reconstruct specific measures in both the natural and social sciences, including three thick volumes by Suppes and Luce themselves. I shall primarily make use of a recent, succinct account from philosopher of science Nancy Cartwright and social psychologist Norman Bradburn, who pioneered work on social indicators and quality of life measurements in the tradition in which sits the capabilities approach to measuring well-being.

Cartwright and Bradburn claim that there are three necessary ingredients of a measure: its characterization, representation, and operational procedures. 'Characterization' refers to the set of features taken to be constitutive of the concept to be measured. 'Representation' refers to the formal apparatus that gives mathematical or diagrammatic structure (or both) to the measure. 'Operational procedures' refers to practices and methods for capturing what I've called the 'objectively-observable' features of the concept. By 'objectively-observable' I mean such features can be captured by some set of physical measurement practices, for example those that record preference orderings on surveys or, in the natural sciences, those that measure temperatures on thermometers. In a good measure, all three ingredients must mesh in the right kind of way such that the characterization is appropriate to the

representation and the operational procedures are appropriate to both the characterization and representation. As I explain below, what I mean by 'right kind of coherence' admits of some variation, but generally entails that practitioners can defensibly claim that the measure, taken holistically, is measuring what practitioners claim to be measuring, in the way they claim to be measuring it.

Relationships of coherence can be established by arguments. Just below I consider one kind of argument – a representation theorem – that can be used to substantiate that the characterization implies that certain features hold for the representation, or, in other words, that the characterization and representation match appropriately. Similarly, arguments can be given to show that the value given by the relevant operational procedures corresponds to the value for the concept as that concept is characterized. Many such arguments that justify the appropriateness of operational procedures are more internal to the system, given that they establish how certain features relate to one another in systematic ways. However, as I explain below, justificatory arguments can also rely on external principles such as natural laws, mathematical assumptions, and even practicalities relevant to local context. Also, as I have noted numerous times in previous chapters, justificatory arguments must regard the purpose to which the measure is lent.

I offer two examples to clarify how arguments can be given to justify that ingredients mesh 'appropriately,' after which I discuss how such arguments often must be sensitive to combinations of context and purpose. First, let's consider a case of measuring units of length, say with a ruler. The concept of 'length' can be characterized by describing facts about it that are taken to hold in the objectively-observable world. For instance, if an object of length l, like a ruler, is laid n times from end to end along a straight line, then the overall length of the interval covered can be expressed mathematically as nl. In this case, repetitions of procedure can be counted as units of length that, in turn, can be represented according to ratios of intervals along a numerical scale. For example, six repetitions of the procedure of laying down a ruler measure a length with three times the magnitude as that measured by

two repetitions. Similarly, a difference of length counted as ten repetitions of the procedure will stand in a ratio of 2:1 with a difference of length counted as five.

Arguments can be given to show that the objectively-observable properties cited in the characterization imply the numerical properties of the appropriate scale, in this case those of a ratio-interval scale. For example, with respect to the procedure of laying down a ruler as a measure of length, certain assumptions about the unchanging nature of three-dimensional space have to hold, along with the assumption that the ruler is exactly the right kind of length, that it holds that exact length over time and in different contexts, and so on. On the basis of such arguments, the three ingredients of the measure can be said to coherently mesh.

As a further illustration, let's consider how the charge on an electron can be measured according to how its trajectory changes as it moves through an electromagnetic field. In this case, the empirical properties of 'charge' - cited in the usual ways of characterizing it – imply the numerical properties of the representation, in this case as units (coulombs) on a numerical scale. Using already accepted laws of physics regarding relationships among mass, acceleration, and force, the arguments show that a change in the electron's trajectory is equal to a certain amount of force, and that amount of force corresponds to the measure of the particle's charge. In this more complex case, a deeper pool of arguments is needed to justify the 'appropriateness' of operational procedures than in the case of the ruler. For example, arguments regarding the consistency and strength of the magnetic field, the credibility of the original path of the electron, and other physical properties of the measurement apparatus would need to be accepted in order to justify that those procedures work together appropriately to determine units of charge as coulombs.

In many cases, arguments that justify measurement system coherence – especially those that justify the appropriateness of operational procedures – must be sensitive to combinations of context and purpose. Consider, for example, early attempts to measure what would later be called 'temperature' as related by Hasok Chang (2004) in *Inventing Temperature.* Early scientists (often called 'natural

philosophers') proceeded by establishing 'fixed points' along a scale of comparison, points that include temperatures for an array of phenomena including 'first night frost,' 'highest summer heat,' 'congealing oil of aniseed,' and more well-known points such as 'blood temperature' and the freezing or boiling points of water (Chang 2004: 10). Each of these earlier attempts at a fixed point included a set of procedures for finding the relevant value that could then be represented, according to the right kind of formal arguments, along an ordered scale by comparing it against other well-established fixed points.

Note that arguments justifying the appropriateness of procedures for setting a fixed point according to 'blood temperature' differ significantly from those justifying procedures for setting a fixed point according to the 'boiling point of water.' Regarding the first case, there would need to be, inter alia, well-defended arguments that different persons are similarly homeothermic, that the context is not one that would significantly alter a person's standard body temperature, etc. Regarding the latter, it turns out that procedures for determining the exact boiling point of water were surprisingly challenging for Royal Society scientists to justify, given that the boiling point of water changes depending on purity, the elevation, the kind of vessel used to boil the water, how the vessel is heated, and so on. Louis Dufour (1832 – 1892), for example, was able to bring "drops of water floating in other liquids up to 178 °C, without even purging the air out of the water," a piece of evidence that made it difficult for Royal Society Scientists to standardize the meaning of 100 °C (Chang 2004: 22). Arguments had to be given to establish the stability of each of the boiling point parameters relative to the purpose of establishing a fixed point within the acceptable margin of error. From these two cases we can recognize that different justifications become relevant depending on how those procedures are taken to fit with the relevant context.

Let's consider another set of formal arguments that can be used to justify that the representation is appropriate to the characterization. As explained by Luce and Suppes in their (2002) essay on "Representational Measurement Theory," practitioners can give a 'representation theorem' that logically connects structural properties that both characterize the concept in question and are collected by the

measure – what Luce and Suppes call the concept's 'empirical structure' – to a set of logical propositions that mathematically represent those structural properties. In such cases, the empirical structure functions as a characterization of the target phenomenon, and the representation theorem proves that the characterization implies that a certain set of features hold for the representation, for example that the empirical structure of degrees of temperature can be represented along a cardinal scale in which the intervals between degrees carry meaning.[5] In many cases, a representation theorem can establish an 'isomorphism' between the empirical structure and a numerical structure such that some function maps the domain of the set theoretic representation of the empirical structure as a 'bijective' – or mirrored one-to-one - mapping of the co-domain of the numerical structure (Luce and Suppes 2002: 3).[6] By showing the representation of the empirical structure (as supposed by the characterization) to be logically connected to a numerical structure, practitioners can substantiate that the representation of the concept is appropriate to the concept as they have characterized it.

Representation theorems are widely implemented for various physical measures such as length, charge, or temperature, as well as contemporary economic measures like those of people's preference orderings among different choices in a set. Though a formal representation theorem might lend credibility to claims from practitioners regarding the meshing of the characterization and the formal representation, the examples above indicate there are other forms of evidence and justification that suffice to establish a measurement system as coherently appropriate. Cartwright and Bradburn point to a further kind of coherence for which principles and theories surrounding the given measurement context can be used to establish whether the system – and in particular the relevant set of operational procedures - really is tracking what practitioners claim it to be tracking. Cartwright and Bradburn explain that

> coherence along a *variety* of dimensions seems to provide the best solution in practice in both the natural and the social sciences: Does the quantity as measured by the proposed

[5] See chapter one section one footnote two for clarification regarding cardinal versus ordinal scales.
[6] In essence, the mapping of the set theoretic representation matches, as a one-to-one relationship, the mapping of the numerical structure. The domain can then be said to 'mirror' the co-domain. I note here that representation theorems need not be limited to isomorphisms among set theoretic structures (one to one mappings). As Alexandrova (2017: 124) explains, mappings can also be homomorphisms (many to one), among others.

> method behave as it is expected to? Do the results cohere with those of other reasonably defended methods? Do the empirical principles needed to support the method cohere with other reasonably justified empirical principles and theories?
> (Cartwright and Bradburn, 2011: 10, *emphasis added*)

Cartwright and Bradburn suggest that arguments regarding how phenomena 'behave' or how well they accord with surrounding principles or accepted practices may not neatly fit together into a formal theorem.

Anna Alexandrova (2017) makes a similar point when she claims that logical connections established by representation theorems do not serve well as justifications for the reliability of various psychological measures of well-being. Alexandrova cites earlier work by Lee Cronbach and Paul Meehl to claim that to "measure *x*, we need to know how *x* behaves in relation to other properties and processes that are systematically connected with *x* by law-like regularities" (ibid: 124).[7] The Coulomb case considered above illustrates what Cronbach and Meehl allude to by 'systematically connected' processes and properties, though in the coulomb case a more formal set of arguments – using laws of physics – is used to show that the representation is appropriate to the characterization. Similarly, as Chang (2004) explains, Royal Society scientists drew from well-established processes and properties to establish that operational procedures for taking a certain amount of a specific kind of water, putting it in a certain kind of vessel, and heating it in just such a way will consistently produce a fixed point of 100 °C. For example, physical laws relating to vapor pressures of boiling liquids and the accepted procedures for measuring them can be used to justify the appropriateness of the procedures for measuring a fixed boiling point. Further practices in the surrounding 'neighborhood' become more or less relevant depending on practical considerations relevant to the given measurement context.

In cases for measuring psychological phenomena such as life satisfaction or intensity of emotion, arguments showing systematic connection of related processes and properties may not be expressed with the same formality as laws of physics, yet may still be accepted as sufficient in context. For example, if a

[7] Alexandrova also argues that a psychological well-being measure might still be called 'valid' if it behaves well in accord with supporting 'background knowledge' (Alexandrova 2017: 124).

life satisfaction survey asks individuals to rank current 'life satisfaction' on a 1 – 5 Likert scale, arguments can be given to establish that individuals read and understand the survey in the same way, and that their marks on the survey credibly establish the right kind of introspective practice such that marks on the same kind of survey can be measured across individuals. In such cases, the relevant set of operational procedures may be given further credibility by showing that related sets of procedures – for example, other kinds of life satisfaction surveys or correlations between survey results and dopamine concentrations in people's brain stems – tend to produce similar measurement outputs.

The preceding arguments emphasize a practical way of understanding the coherence of a measurement system, one that follows philosopher of science Otto Neurath's famous boat metaphor: "We are like sailors who have to rebuild their ship on the open sea, without ever being able to dismantle it in dry-dock and reconstruct it from our best components" (Neurath 1983: 92).[8] Useful pieces are meshed with neighboring but still functional pieces into an assemblage that may be deemed sufficient even when some features do not fit together according to the kinds of standards that would typically be applied. Following this analogy, Hasok Chang makes the important practical point that some incoherence among ingredients may be admissible when ingredients of the measure interact "with respect to a very particular purpose, namely keeping the boat afloat" (Chang 2017: 11). If there is incoherence yet the boat sails well enough, then it may not be the kind of incoherence worthy of further concern in that context.

As an illustration, let's consider another example from Chang's (2004) history of measuring temperature. Nineteenth century the scientist Henri Victor Regnault managed to defend as appropriate the practice of using air thermometers to measure temperature, even when there was no conclusive theoretical justification for the claim that expansion of units of air is proportional to the increase in units of temperature (Chang 2004: 74-76). Regnault's exhaustive calibration work of measuring temperatures and comparing the corresponding expansions of other substances was deemed to be sufficient justification of the 'law-like regularity' with which expansion of air can be taken to accord with increase in

[8] See Neurath *Philosophical Papers* [1932/1933] 1983: 92.

temperature. On the basis of Regnault's work, the air thermometer became the accepted apparatus for measuring then-theoretical reaches of temperature during that period.

The example illustrates that even if some features of different ingredients of a measurement system don't mesh – i.e. are incoherent -, such features may, when validated by outputs given by other sets of related procedures, be given further credibility for that combination of purpose and context. Throughout this thesis I've considered a similar kind of argument that can be given by the supporter of the Justification of Total Orderings on the Basis of Practical Comparison (JTOBPC) for incorporating procedures that always provide total orderings of capabilitarian well-being. On the supporter's view, the intuitive readiness with which people accept procedures for making practical comparisons between different kinds of capabilities provides sufficient warrant for incorporating incoherent procedures. Yet the examples of Regnault's calibration work and Neurath's boat indicate that practitioners still need to question whether such practices really manage to 'keep the boat afloat' relative to the combination of purpose and context to which the measure is lent.

A further consideration applies to whether a measure of capabilitarian well-being manages to 'keep the boat afloat.' In social science it is not uncommon for an intuitively appealing, theoretically valid concept to have indefinite features that are difficult to measure directly. 'Well-being' is arguably one such indefinite concept. Indefinite social concepts like well-being are what Cartwright and Bradburn, following philosopher Otto Neurath, call *ballung* concepts, translated as 'congestations.' An essential quality of a *ballung* concept is that though there is often a 'loose' sense of the term understood according to ordinary language, different features of the concept matter in different contexts. This quality is similar to the Wittgensteinian phenomenon of 'family resemblance' in which a person knows that another is a family member but cannot pinpoint, in general, any single difference-making feature or set of features sufficient to make that distinction across all members of her family.[9] Alexandrova (2012, 2017)

[9] See Wittgenstein in *The Blue and Brown Books* (1958: 17-18) and *Philosophical Investigations* (1953: sections 66-67). For further information on Ballungen in social sciences, see Cartwright and Bradburn (2011: 4-5).

convincingly argues that dispersed sets of features matter for well-being in different contexts.[10] For example, what 'well-being' is for a gerontologist differs significantly from what 'well-being' is for a policy maker working in development, though intuitive resemblance of the ordinary sense of 'well-being' is still shared between the two domains.[11] Similarly, Sophia Esftathiou (2016) argues that well-being is a 'found' concept sensitive to different combinations of purpose and context. Following these authors, I claim that 'capabilitarian well-being' is also a *ballung* concept. Consider that Nussbaum claims that capabilities are 'multiply realizable' depending on culture and context, and Sen characterizes capabilities such that they are evaluated in diverse kinds of ways in a process of pluralistic deliberation.[12]

Cartwright and Bradburn explain that when constructing the measure of a *ballung* concept, practitioners can respond to the indefiniteness of the concept by either (a) leaving contextually-relevant ambiguities in the representation in order to maintain coherence with a characterization closer to the ordinary, loose sense of the concept, or (b) collapsing different features of the concept in order to provide a formal structure that is more useful for predictive or comparative purposes.[13] However, the possibility of incoherence caused by reductively collapsing represented features presents a dilemma: either the complex but descriptively accurate characterization of the *ballung* concept remains too vague to support certain purposes such as comparisons or predictions, or the representation of the measure is narrowed such that it may no longer measure the target concept.

For example, in her (2015) thesis *Casual Inquiry in the Social Sciences: The Promise of Process Tracing* Rosa Runhardt explains that when political scientists attempt to define a standard of 'civil war,' a related feature of 'ethnicity' that should strongly correlate with civil wars has a problematically low or washed-out correlation if the characterization of civil war is too broad. On the descriptively accurate side of the dilemma, the attempt to give the characterization a wide contextual scope can result in low

[10] See Alexandrova (2012: 682-688) for discussion of and support for a variantist theory of well-being, which indicates that different substantive theories may supply the conceptual referents for different practices of well-being, such as well-being for a caregiver vs national well-being.
[11] Recognition that measures operationally interact with or quantify a different set of features depending on context is supported by the operationalism of Percy Bridgman (1927) and more recently by the pragmatist theory of measurement defended by Hasok Chang (2004).
[12] See chapter one regarding the definition of 'capabilitarian well-being' for the purpose of this thesis.
[13] See Cartwright and Bradburn (2011) p. 10-12.

correlation with important features, in this case that of ethnicity. Runhardt explains that one researcher has redefined civil war as 'ethnic civil war' in order to remedy the inconsistency, yet this reductive characterization of civil war no longer measures the original civil war concept across a wide set of contexts, thereby illustrating the second side of the dilemma.[14]

As mentioned above, responses following either (a) or (b) are often aligned with different social scientific purposes such as descriptive understanding, prediction, or explanation. Following Cartwright and Bradburn, I link some of those purposes with the following definitions of responses to measuring *ballung* concepts, considered at the representational level:

> **Type-(a) response** incorporates tables or similar disaggregated structures to represent sets of features that accurately apply in different contexts, but often provide incomplete bases for comparison, prediction, or explanation and generally support only partial orderings of individual features to be measured.
>
> **Type-(b) response** collapses the *ballung*-type concept into a particular class of features, often dependent on the relevant combination of purpose and context, allowing for the representation to be used for statistical correlation, causation pathways in models, cost-benefit-type comparisons, etc.

I would emphasize, following discussion from this and earlier chapters, that both the purpose of the measure (e.g. prediction or comparison) and the scope of the measurement context play a role in justifying whether, in the case of *ballung* concepts, response (a) or (b) is appropriate.[15] Response (b), while useful for more specific purposes like prediction or - in the case of measuring well-being - comparison, often does so at the expense of descriptive validity.

In chapter one I gave an analogy that illustrates the same kind of response distinction: while the Goode-Homolosine projection of the three-dimensional globe is in many ways more descriptively

[14] See Runhardt (2015) chapter two. See also Runhardt and Cartwright in Philosophy of Science: A New Introduction, (2014: 269-270).
[15] See chapter two section five.

accurate, the overall 'orange peel' picture it provides is an incomplete portrayal of the globe. The Mercator Projection provides a more complete but flawed picture. Each projection is more suitable for different purposes. For example, if the goal is to teach global geography to students, the picture provided by the Mercator Projection is more felicitous.[16] Similarly, when practitioners claim to measure capabilitarian well-being according to Sen or Nussbaum's characterization, an approach that favors (a) is more coherent than one that favors (b). However, following Hasok Chang's point above that some incoherence may be acceptable relative to different contexts and purposes, I recognize practitioners may in some cases still have sufficient practical justification for implementing often-incoherent (b)-type procedures that collapse the *ballung* concept.

I now turn to consider, following RTM, the kind of claim that can be used to establish incoherence for a capabilities measurement system:

> **General Incoherence Claim:** If a capabilities measure, following the concept given by Martha Nussbaum or Amartya Sen, is characterized or represented according to certain features yet represents or operationalizes capabilitarian well-being in a manner that contradicts those features, that measurement system is incoherent and may provide results that are off-target, meaning there is good reason to believe the measure may not be measuring what it is supposed to, unless substantive argument can be given to defend that the possibility of being off-target is not significant for that combination of context and purpose.

This claim provides a warrant for further investigation, meaning incoherence is merely the starting point of analysis, not the end of it. To use a common automotive metaphor, incoherence is like a flashing 'check engine' light in an automobile dashboard. We do not look at the light and consider ourselves satisfied that we understand the problem; a further set of principles and procedures must be invoked to

[16] It unfortunately turns out that if one seeks to accurately lob projectiles across large distances, the Goode-Homolosine is better suited. See chapter one section three.

extend analysis into the given context. To extend the metaphor, consider that further signs pointing to a slow oil leak or a faulty spark plug give less cause for concern than, say, those pointing to a damaged transmission or a mis-firing piston. We use incoherence in the way the mechanic uses the check engine light; further analysis of relevant features of that incoherence provides the grounds for understanding how the system may be off-target, and how significant are the possible harms associated.

In the case of an automobile with a check engine light on, various tools and indicators – along with relevant background assumptions - become appropriate depending on what area of the automobile seems to be affected, as do background theoretical principles. The mechanic will use one set of tools and principles for an electrical problem, another for a problem with the carburetor or the pistons, etc. The extended examples I give later in this chapter similarly invoke further principles and procedures that allow us to recognize and offer solutions for problems related to the incoherence of capabilities measurement systems with respect to Nussbaum and Sen's commitments to value-ladenness and qualitative distinctiveness.

In the broadest sense, a measure's being 'off-target' is a possible correlate of incoherence of measurement system ingredients, as stated above. In a more mediate sense it can indicate important omissions (as in the value-laden extended example in section three) or distortions of comparison (as illustrated in section four). Yet this mediate sense does not directly address the significance of off-target measurement outcomes; again, that concern must also regard the purpose of the measure as well as the context.[17] Ingrid Robeyns' (2006) article on *The Capability Approach in Practice* provides a thorough list of contemporary uses of the capabilities approach, including "general assessments of the human development of a country; the assessment of small scale development projects; identification of the poor in developing countries; poverty and well-being assessments in advanced economies; an analysis of the deprivation of disabled people; the assessment of gender inequalities; theoretical and empirical analyses

[17] See the final section of this chapter for further discussion of the relevance of purpose. The thought here parallels an influential paper from Richard Rudner (1953) who argues that values are inescapably a part of the scientific process, even measurement. Those values reflect, for example, the purpose of that scientific inquiry or research program.

of policies; critiques on social norms, practices and discourses; and finally, the use of functionings and capabilities as concepts in non-normative research" (ibid: 360-361). If the measure is utilized in a way in which its being slightly off-target can make a profound difference for human lives (and capabilities measures used to guide policy decisions do so), then the possibility of being even slightly off-target can be significant cause for concern.

4.3 Argument and Extended Example for Value Omission Incoherence

I give the following claim for *Value Omission Incoherence*, which acts as a sub-argument for the general incoherence claim given at the end of the previous section:

> ***Value Omission Incoherence:*** If, following the concept given by Sen or Nussbaum, a capabilities measure characterizes or represents (or both) capabilities such that they are intrinsically value-laden, then when the same measure represents or operationalizes capabilities as value-free or value-neutral at any stage of the process of measurement, that measure is incoherent and may provide results that are off-target, meaning the results are at odds with what actually is the case in terms of capabilitarian well-being.

Recall from chapter one that values, in Ruth Chang's 'ordinary' sense of the term, can be recognized according to a diverse set of action-guiding reasons, expression of deep emotions, surrounding cultural norms, particular perspectives, etc.[18] Recall from chapter two the view presupposed in Sen's characterization that evaluative processes of deliberation should allow for expression both as valuation simpliciter and as impartial sets of reasons. Lastly, recall that capabilities are value-laden in two notable ways: different kinds of capabilities and capabilitarian well-being values can be placed into at least partial orderings by some process of evaluation, and different kinds of capabilities are valuable when they are recognized as relevant – what has also been called their process of 'selection.' As the extended example in this section will demonstrate, Value Omission Incoherence can be manifested as a myopic selection of capabilities functionings, which can be of significant concern when the omitted valuation is that of a stakeholder in related policy decisions.

Though capabilities have been characterized by Sen and Nussbaum as value-laden, recent work in the capabilities approach challenges this characterization. In *Wellbeing, Freedom and Social Justice* (2017), Ingrid Robeyns argues that the shorthand characterization of capabilities – given in chapter two as

[18] The definition of value-ladenness also accords with what Flavio Comim calls the 'Valuational Foundation' of the capabilities approach. See Comim 2008: 162-166. See also Sen and Nussbaum 1987 *WIDER Paper 30: Internal Criticism and Indian Rationalist Traditions*.

'doings and beings people have reason to value' - is problematic as it implies people should only consider capabilities for which they would express a positive valuation. Robeyns argues that "some functionings have no value or even have a negative value" and so "functionings should, therefore, be value-neutral" to allow for those with no value or with a negative value (Robeyns 2017: 41). Robeyns gives examples of 'ambiguous' functionings like caretaking or sex work, which can be valued positively or negatively depending on context. On the basis of these and similar examples, Robeyns argues that sets of doings and beings must be represented and treated as value-free (or value-neutral) during the process of selecting capabilities, after which they can then be evaluated as positive or negative depending on context. For example, if a municipality is selecting which indicators to use for a capabilitarian well-being measure – perhaps, as Robeyns suggests, to be used for explanatory research (ibid: 142) -, the deliberative group can 'select' all possible functionings that can be imagined by the deliberative group, irrespective of how they might be valued, and then evaluate those functionings relative to one another as a different stage of the process.

Let's take a moment to address Robeyns' treatment of the shorthand characterization. Though the phrase "have reason to value" can be interpreted such that only positively-valued capabilities should be selected and evaluated, Sen and Nussbaum incorporate both negative and positive valuations into their respective characterizations. For example, though Sen's Human Development Index includes proxies for income, longevity, or education that are generally recognized to have a positive valence, when those proxies are evaluated according to Sen's representation as 'sets of co-possible sets,' such evaluation can include negative valuation. For example, if individual y_1 attains high levels of income at the expense of lower than usual levels of health or education, that life may be valued negatively, or at least have tragic components that would be expressed as negative valuations within deliberation.[19] Negative values

[19] To further extend the point within the context of Sen's characterization and representation, let's note that the way an individual's capability set is evaluated depends on the background resources available to that individual. Thus, in agreement with Robeyns' earlier point, even sex work might have a positive value if there are few or no other wage-earning options, if the sex-worker is a single parent, etc. However, at least on the Sennian view, it is the holistic set of co-possible sets that are considered and so there might be other limitations associated with sex work (or caretaking work, or low-paying work with long shifts, etc) that make it such that the holistic set has significant negative value compared with other sets people have reason to value, as some complex combination of positive and negative valuations.

similarly feature in Nussbaum's characterization and representation, albeit in a rougher way. For example, if individual y_2 has the three indicators for different dimensions of health, education, and affiliation, and y_2, similar to y_1, attains a high level of income but ends up below the relevant minimum thresholds for health and affiliation, then those deprivations are valued negatively. Moreover, even though indicators might represent capability deprivation as a positive number, say as a longevity of 55 years instead of 72, that deprivation can have a negative valuation when considered within deliberation.

In a recent article, Morten Byskov follows Robeyns in claiming that capabilities should be characterized neutrally as "the real freedom to achieve certain doings and beings" rather than those that 'individuals have reason to value' (Byskov 2019: 11). Unlike Robeyns, Byskov claims that evaluative processes of deliberation should at all stages include the commitment to value-ladenness, and so his value-neutral characterization is limited to non-deliberational purposes "that either are not concerned with neither reasonableness nor values, such as some empirical and conceptual purposes, … are not concerned with whether the values in question are reasonable, … and/or are concerned with unreasonable values, such as in the context of identifying adaptive preferences" (ibid: 8). I respond that Byskov, like Robeyns, makes a conceptual clarification that may be useful for a narrow range of application, but disregards the importance of recognizing that capabilities are at all times value-laden for most practical purposes of measurement.

Regarding Byskov's more theoretical gesture to purposes that 'are not concerned with reasonableness or values,' actual work in the capabilities approach done according to these kinds of 'empirical and conceptual' purposes is, as Robeyns notes, marginal.[20] Regarding the applicability of Byskov's neutral characterization to identification of adaptive preferences, Byskov over-emphasizes Serene Khader's (2011) call to use processes of deeper reflective scrutiny to investigate possibly-adaptive values given within first person narratives. Possibly-adaptive preferences must be recognized as value-

[20] Robeyns notes that "to the best of my knowledge, few scholars use the capability approach in this [conceptual/explanative] way" (Robeyns 2017: 142).

laden, even if adaptive values may not hold up under certain forms of critical scrutiny.[21] Following Khader (2011), I recognize it might indeed be useful to treat expressions of possibly adaptive preferences as value-neutral in an effort to determine if the relevant valuations are 'deep' valuations that hold even after consciousness-raising methods have been implemented. For example, following Chad Kleist's (2017) example, if a group of indigenous peoples downplay the capability to read in favor of more practical forms of education, it might be helpful, following Khader, to promote further dialogue in an effort to determine if the values shift once people are 'more aware' of the practical value of reading.

However, even in this kind of situation, it may be more authentically value-tracking for practitioners to accept the valuation given as indicative of what indigenous people consider to be more pressing in terms of increasing their own well-being. Here there is a careful balance between progressive 'consciousness raising' and the muting of the voice and influence of valuations given by stakeholders who will be affected by relevant policy decisions.[22] I return to this point later in the section and in the final section of this chapter. For now, I argue that though treatment of capabilities as if they are value-free or value neutral may be useful with respect to the more marginal purposes and contexts indicated by Robeyns and Byskov, even those marginal purposes may be made more effective – and certainly coherent - by retaining the commitment to treating capabilities as value-laden. There is not sufficient justification to argue that Sen and Nussbaum's characterizations and representations support a general commitment to treating capabilities as value-free or value-neutral.

However, there is some precedent in Sen's earlier writing allowing that capabilities can at least be represented as value-free. A misinterpretation of this precedent could be used to support the claim that capabilities can be treated similarly during some processes of measurement. In his monograph *Commodities and Capabilities* (1985a), Sen makes an analytic distinction similar to that of Robeyns and Byskov, suggesting that the objectively-observable content of capabilities can be separated from the

[21] See Serene Khader (2011): 'Deliberative Perfectionist Approach to Adaptive Preference Intervention.'
[22] Contemporary research in this area is recently summarized by Clark, Biggeri and Fredani: *The Capability Approach, Empowerment and Participation* (2019).

process of evaluation.[23] Sen gives a nested set of functions that together represent how a capabilities set - given as the set of co-possible sets of achievable functionings - can be evaluated:

"x_i	=	the vector of commodities possessed by person *i*
$c(\cdot)$	=	the function (not necessarily linear) converting a commodity vector into a vector of characteristics of those commodities
$f_i(\cdot)$	=	a personal 'utilization function' of *i* reflecting one pattern of use of commodities that *i* can actually make (in generating a functioning vector out of a characteristic vector of commodities possessed)
F_i	=	the set of 'utilization functions' f_i, any one of which person *i* can in fact choose …

...If the person chooses the utilization function $f_i()$, then with his or her commodity vector x_i, the *achieved functions* will be given by the vector b_i,

$$b_i = f_i(c(x_i))$$" (Sen 1985a: 11).

Here Sen defines b_i as one or another set - or 'bundle' - of achievable functionings that an individual can attain given a certain set of resources (commodities). The *entire set* of all co-possible b_i bundles that an individual is free to choose - what Sen (1985a) posits as Q_i - is a representation of the individual's capabilities:

[23] See previous section regarding objectively-observable.

"If the person's choice of commodity vectors is restricted to set X_i, then the person's feasible functioning vectors are given by the set $Q_i(X_i)$,

$$Q_i(X_i) = [b_i \mid b_i = f_i(c(x_i), \text{ for some } f_i(\cdot) \in F_i \text{ and for some } x_i \in X_i]...$$

$Q_i(X_i)$ represents the freedom that a person has in terms of the choice of functionings … Q_i can be called the 'capabilities' of person *i*" (ibid: 13)

Note that $f_i(c(x_i))$ represents an objectively-observable utilization pathway for an individual who has commodity vector $c(x_i)$. Similarly, Q_i is an objectively-observable economic 'fact' about what the individual can do or be with their possible commodity vectors given their resources and personal characteristics. Though the true 'freedom' of Q_i cannot be directly measured given its latent nature (how could practitioners measure all the possible 'what-ifs' that a person could choose in terms of attainable bundles of functionings), it *is* possible to measure many of the objectively-observable functionings that people manage to achieve, arguably in the same sort of value-free way that we can use a ruler to measure the fact of some object's length. Yet such an argument fails to consider how values influence the selection of objectively-observable functionings to be measured. I return to this point just below.

Sen (1985a) then provides a nuanced representation of both valuation and evaluation of capabilities, acknowledging - as he repeatedly does in later writing - that there is a variety of ways that a person might express the positive or negative value for some vector of functionings. Sen first explains that an individual can give a valuation according to the following definition:

"If $v_i()$ is the valuation function of person *i,* then the value of that vector of functionings b_i is given by

$$v_i = v_i(f_i(c(x_i))) \ldots$$

Given the valuation function $v_i(\cdot)$, it is of course possible to characterize the values of well-being that [one] can possibly achieve, given by the set V_i,

$$V_i = [v_i \mid v_i = v_i(b_i), \text{ for some } b_i \text{ in } Q_i]$$" (ibid: 13-14).

Note that v_i is a specific valuation, one of many possible positive or negative valuations that can be expressed by an individual regarding some i^{th} bundle of attainable functionings.[24] Recall from chapter two that Sen allows that representations of valuation may operate on different features of the relevant capabilities set – for example the number of achievable functionings, the most highly-valued or lowest-valued option in the set, etc.[25] During the process of evaluation, such features are considered holistically with further features of other co-possible sets as a representation of how the individual values their overall set (Q_i) of possible 'lives worth living.' Though the actual process of evaluation holistically allows that different features of co-possible sets become more or less salient – often, as Sen explains, relative to the expressions that survive critical scrutiny within a forum of pluralist, public deliberation -, we can at least think of V_i as the set of relevant valuations that feature in the evaluation of the capabilities set. Some further process is then used to determine which v_i within V_i are relevant for purposes of evaluating the capabilitarian well-being of individuals or entire populations. In many cases, as Sen notes in this and later writings, evaluation can provide orderings of capabilities sets, though Sen argues such orderings are often partial (ibid: 31).

Let's return to the question of value-neutrality in the representation. Ostensibly, there seems to be no reason to claim incoherence for Sen's (1985a) representation with respect to the commitment to

[24] See chapter two section two regarding the difference between 'valuation' and 'evaluation' in Sen's characterization.
[25] See chapter two section three.

value-ladenness, as V_i defensibly stands in for valuations that inform some evaluative process of deliberation. It would appear, then, that valuation and evaluation can conceptually be allocated to functions that operate at some later stage of measurement than do the procedures that together capture the objectively-observable 'facts' about people's capabilities, including what resources (X_i) people have, how they can use them (F_i), and what bundles of functionings they can choose given those commodities and utilization pathways (Q_i). This kind of representational move is important to economists as it allows that 'valuation' can be separated from procedures used to capture objectively-observable 'facts' that depend on what the world is like and what are its circumstances - facts like length or, arguably, achieved functionings.

However, there is a significant problem when Sen's (1985a) representation is operationalized as the set of an individual's utilization pathways. Not every actually possible set of utilization pathways can be considered – there are in principle very many possible things one can do with a given bundle of commodity characteristics, and many of these will seem clearly ridiculous or irrelevant in circumstances that can include severe resource and tactical ability constraints on what can and should actually be measured. Practitioners using the measurement system either need to select some set of possible functionings that can be measured - say, for example, whether the individual has access to running water or is able to attend some quantifiable form of public school - or will need to select one or more broader indicators meant to stand as proxies for possible functionings or even for sets of them - for example broader indicators such as income level or years of education. Either set of operational procedures is inescapably value-laden at the moment of selection.

As Bernard Williams notes in his (1985b) reply to Sen, "we shall want to say that certain capabilities are important in a given society because in that society the cultural understandings legitimate or emphasize certain possibilities" (Williams in Sen 1985: 101). Let's explore the implication of Williams' claim. Firstly, deliberators consider different possible functionings or proxies because they know those proxies are valuable. For example, access to running water may be quite valuable when

villagers have to walk a mile to the nearest well. By contrast, a functioning such as one's having 'completed six years of schooling' may not have much value for deliberators, and, accordingly, might not be selected – in that exact form - as either a functioning or an indicator.[26] Secondly, practitioners – often working as or with deliberators – choose proxies based on practical limitations such as available data, resources, etc. It might be that years of education is the best approximate indicator that can be gathered, even if an indicator of reading or linguistic ability would better indicate the kind of functioning that is valued by the community. The choice to use the more obtuse indicator carries with it the risk that some kinds of functionings covered by the indicator will be over-emphasized, and others ignored. When practitioners select representative indicators, they are not applying values at a later stage of the measurement process.

Williams considers the problematic example of a deliberative body operating in an Islamic state, where a potential functioning for selection is whether people attend a Catholic Church. In this example, dominant cultural values would likely preclude even the recognition of such a functioning by the deliberative body, though that functioning might be valuable for some sub-section of the population.[27] The deliberative body could even claim to be 'measuring' only functionings that are value-free or value-neutral, yet still manage to 'miss' the opportunity to consider such a functioning. Or, to further the example, let's allow that the functioning to attend some kind of church is recognized by the deliberative body as possibly important for measuring capabilities. Even so, that functioning might not be captured by the proxy indicator that is later selected, perhaps because the proxy serves a political function of ignoring the valuations expressed by a minority population. For example, the panel might choose a 'value-neutral' proxy indicator that measures 'frequency of church attendance' but does not consider which kind of church. In either case, the end result is the same: the deliberative body omits the functioning, which cannot be evaluated.

[26] The 2018 MPI, for example, counts at a 1/6 weight the capability deprivation described as "No household member aged 10 years or older has completed six years of schooling." See Alkire and Jahan (2018: 11): "The New Global MPI 2018: Aligning with the Sustainable Development Goals."
[27] The extended example given later in this section mirrors Williams' example.

Let's again consider whether Sen's 1985 measure incoherently represents the selection of functionings. Even though functionings or indicators can be represented as value-neutral at the point of selection, they are – as Williams' example illustrates - intrinsically value-laden before any valuation function such as v_i operates on them. If operational procedures honor the analytic distinction between descriptive 'fact' and later processes of 'evaluation,' the measure becomes incoherent with respect to Sen and Nussbaum's commitment to the value-ladenness of capabilities. It is perhaps in recognition of this problem that later in the same monograph, Sen recognizes that his representation of functionings "divides up the problem of evaluation of well-being into two distinct (though *not independent*) parts, viz., (i) specification of the functioning achievements, and (ii) valuation of the functioning achievements" (Sen 1985: 30, *emphasis added*). Though a conceptual distinction can be made between selection and evaluation of functionings, Sen is careful to note that those two 'parts' are *not* independent; they occur simultaneously in practice.

Let's consider a similar potential ambiguity with respect to Nussbaum's characterization and representation. Nussbaum's objective list representation is constructed according to a reflective inquiry regarding the potential achievements we characterize as constitutive of a good life, and so appears to be value-laden. Yet it could be argued that Nussbaum's more basic capabilities – those that act as a general foundation for other important doings and beings - are 'value-neutral' if they can be represented as objectively-observable 'facts,' unencumbered with cultural values. However, as explained in the previous chapter, even Nussbaum's basic capabilities must incorporate the values associated with defining a good life according to a sense of flourishing or dignity. Moreover, following Nussbaum's feature of multiple realizability, the expression of one or another capability becomes more heavily value-laden the more localized the population of focus. In order to operationalize the system as a concrete set of measurable functionings, values inevitably play a role right from the start.

Thus far I have provided a claim for Value Omission Incoherence and have argued against more theoretical claims made by two contemporary practitioners – Robeyns and Byskov - who challenge the

claim that capabilities are value-laden. Following Robeyns and Byskov, I accept that it may be useful to represent capabilities as value-free or value-neutral in limited contexts, as these representations may allow for better-informed deliberation regarding functionings that have ambiguous or negative values, or that may be evaluated according to adaptive attitudes. Yet given the common application of capabilities to the creation of population-level measures, it is practical to represent and operationalize capabilities as value-laden at all stages of measurement. In general, the operational procedures that follow this commitment are those that attempt to provide voice and influence to minorities or oppressed individuals whose perspectives are often dismissed.[28]

Byskov (2016) makes a similar point when he explains that procedural methods used to evaluate capabilities must derive capabilities and functionings "from an open-ended inquiry into people's actual values" (Byskov 2016: 206). Following other capabilities scholars like Sabina Alkire, Serene Khader, and Rutger Claassen, Byskov argues that the right kind of deliberation procedures should "form a critical input that aims to enable local stakeholders," or, similarly, to "engage in a critical dialogue with stakeholders" (ibid: 219). Because there might be miscommunication between local stakeholders and practitioners, Byskov argues for anthropologically-informed intervention by professional mediators to give any possibly oppressed stakeholders further voice and influence.

Other authors follow Byskov in recognizing that during processes of deliberation, stakeholders need to be given voice and influence. Mario Biggeri et al. (2006), for example, argue that children should be treated as participants in the process of delineating their own 'core capabilities' and propose a methodology honoring this proposal. Similarly, in her essay "Public Debate and Value Construction in Sen's Approach," Alkire recognizes that "when judgments as to which capabilities are valuable must be made, and when these judgments affect wider groups of people, procedural considerations enter" (Alkire 2006: 135). Alkire considers participatory planning techniques meant to address procedural considerations, a preferred methodology of which is Participation, Reflection and Action (PRA) in which

[28] See also Konyndyk and Rose Worden's essay "People-Driven Response: Power and Participation in Humanitarian Action" (2019).

the facilitator "hands over the stick" and implements various techniques to avoid "often unconscious … ways of dominating others" (ibid: 143).

Yet even if the deliberative process includes PRA-type procedures, local stakeholders may still feel intimidated or express adaptive preferences other than their actual values. In his 2016 thesis *Developing Capabilities: A Feminist Discourse Ethics Approach*, Chad Kleist mirrors Alkire (2006) in arguing that "social power relations are ubiquitous and anyone purporting to develop a global moral theory must attend to how these power dynamics may be influencing the process of theory construction and defense" (Kleist 2016: 107). As indicated by the Williams example above, social power relations are present in the form of dominant norms or values that can affect the selection of functionings or proxy indicators. If deliberation is to occur such that the right kinds of valuations are not only recognized but also given the right levels of influence, then the operational procedures must address ubiquitous power imbalances that, unfortunately, can occur in unconscious ways even for well-meaning practitioners.

Let's consider an extended example of these concerns. A deliberative body - call it B_1 - is following Sen and Nussbaum's concept to create a multidimensional measure of educational capabilities that will be used to assess different development initiatives within a community development plan. Such a community development plan would be similar in nature to a Village Development Plan, for example such as that established by the Sungi NGO in Hazara, Pakistan (Alkire 2006: 147). Most deliberative bodies that create capabilities measures are technocratic, but many use surveys, dialogues, or similar procedures to try to engage local stakeholders. Accordingly, let B_1 be a small panel of social workers, academics, and some representative sampling of the population. The panel provides access to the same interview and statistical data to all members and follows accepted methodology to assure that local cultural values are actively considered in deliberation. For example, panel members can place talking points on agendas, communicate with other members, interject, make motions, and share their thoughts before and after presentations. Importantly, B_1 operates under the assumption that evaluation is a *separate* procedure from the selection of functionings: at the point of selection, functionings are

considered to be value-free or value-neutral 'facts' similar to physical facts regarding temperature or color, and they are treated as such before they are 'evaluated' at a later stage in the process.

In this example, the framing of 'selection' as a process of deliberation regarding objectively-observable 'fact' rather than as a value-laden participatory process structured by social power dynamics increases the likelihood of myopic selection. Social power dynamics are further leveraged when the procedure of B_1 indicates that 'proper evaluation' of capabilities – meaning deliberation regarding relative weights for different indicators - is to occur after the selection. Consider, for example, that during selection, B_1 determines that being able to read at an 8^{th} grade level and to perform a certain level of computational math useful for running most small businesses are important functionings. The two functionings selected might seem appropriate enough and are the result of a well-constructed process of democratic deliberation, one that reflects the panel's general emphasis on development and reflects best available data.

However, consider that the functionings selected do not include forms of culturally-sensitive education, say in the oral history or the language of more indigenous sub-groups of the population, either because that sub-group was simply not represented by B_1 or because arguments for inclusion of such functionings made by representatives of the sub-group were misunderstood, consciously or unconsciously. Or consider that functionings selected do not include educational focus on aesthetics, music, or literature, again perhaps because representatives that would voice support for these functionings are not represented within B_1 or because dominant definitions of education deny those functionings as relevant for a socially-guaranteed minimum for a 'flourishing' or 'dignified' existence. Or consider that functionings do not include practical forms of education such as agricultural training, basic first aid, negotiating skills, collective governance training, etc., even though local stakeholders highly value such skills.

The existence of these kinds of omissions are accepted as fact by feminist standpoint theorists, anthropologists, and ethnographers. Though examples are myriad, Susan Moller Okin's Justice, Gender

and the Family presents exactly this kind of case in which the Rawlsian conception of the family unit within the original position elides important claims for justice regarding shared workload, gender identities, and family roles. Similarly, Iris Marion Young's Justice and the Politics of Difference points to oppressive objectifications of gender or minority status that can be presented as if they are natural, social facts.[29] A basic tenant of feminist standpoint theory is that privileged perspectives take as objective what is clearly not from other perspectives, especially those of the disadvantaged. As Jonathan Wolff and Avner de-Shalit explain in Disadvantage, the standpoints of minorities are, if they can be effectively sought out and represented, the practical source of our perspective on what functionings matter the most to us and why.

Given the growing number of authors within the capabilities approach who support participatory methods, it might be a stretch to claim that B_1 uses a value-neutral or value-free deliberative democratic procedure rather than participatory methods for the selection. Let's alter the example and consider the same deliberative panel instead implementing participatory methods during selection. Local stakeholders are encouraged to share what functionings they might value – either positively or negatively - such that the committee can then debate whether to include those functionings in the measure. In such a context, the alternative functionings mentioned above stand a greater chance of getting selected, even if alternative functionings challenge dominant social norms or standards.

However, if social power dynamics are not *explicitly* recognized and challenged – especially at the procedural level -, such perspectives are easily dismissed or made tacit, even when space for deliberation has been provided. Moreover, even if some participatory procedure can be implemented to encourage local stakeholders to express values, stakeholders may still face significant psychological barriers that prevent their expressing the kind of valuation that would be shared in a more private setting. For example, if asked "what kind of educational initiative do you think would be best for our

[29] See also: The Second Sex by Simon de Beauvoir, Yellow by Frank Wu, Life Support by Kalindi Vora, Resisting Reality by Sally Haslanger, Tony Lawson (1999) *Feminism, Realism, and Universalism*, and Donna Haraway (1988): *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.*

community," stakeholders might provide the response that seems most likely to be implemented or that best respects surrounding social norms, even if the stakeholder privately attaches deeper meaning to alternative functionings. This can especially be the case when stakeholders fear negative consequences for speaking up, for example against traditionally-accepted, paternalistic separations of gender roles, methods of policing, exploitation in the form of labor practices, and so on. Moreover, even if the stakeholder speaks up, there is no guarantee that valuation will be interpreted and received by other panel members who have different, perhaps dominant, perspectives and values.

The example indicates that procedures for evaluation must transparently push against barriers created by social power relations. This same point is mirrored by Alkire's (2006) critique of PRA: even well-trained methodologies of participatory planning run the risk of 'ventriloquism' in which minority or suppressed perspectives echo the values they believe the facilitator wants them to express. Alkire explains that in an impact assessment of the $110 million Madhya Pradesh District Poverty Investment Project (MPDPIP) in India for which PRA was implemented, it was found that "the projects facilitated by irrigation engineers tended to be lift irrigation projects; veterinarians' communities tended to "demand" livestock projects – and so on. Of course there were some sterling exceptions, but the 'ventriloquist' trend was noticeable to those managing and monitoring the project" (Alkire 2006: 143). Alkire also notes that the threat of violence in between episodic rounds of deliberation is also a very real concern (ibid: 149). It should be no surprise to facilitators that sub-groups will appear to accept dominant standards and not give voice to functionings that are nevertheless important to them.[30]

Its relative simplicity notwithstanding, what the extended example makes clear is that selection of capabilities is a politicized procedure in which deliberation is framed by surrounding social values. What might be represented as a value-free or value-neutral selection of functionings or indicators is actually, at all times, influenced by social norms and standards. Procedures such as democratic votes or presentations from technocratic specialists do provide avenues for consideration and selection of important

[30] See Sen 1999: 62-63. See also Teschl & Comim, 2005: 245 where they argue that not adaptation but resignation is likely inescapable in the capabilities selectioning space.

functionings. Yet without ex-ante recognition that valuation is inescapably present and that, relatedly, social power relations influence the recognition and interpretation of valuation within deliberation, even well-meaning practitioners are at the risk of over-selecting functionings deemed valuable from a dominant perspective.

4.4 Argument and Extended Example for Total Ordering Incoherence

I give the following claim for *Total Ordering Incoherence*, which acts as a sub-argument for the general incoherence claim given at the end of section two:

> ***Total Ordering Incoherence:*** If, following the concept given by Nussbaum or Sen, a capabilities measure characterizes or represents capabilities or capabilitarian well-being values such that they are of different, qualitatively distinct kinds, then when the same measure incorporates procedures that always provide total orderings of capabilitarian well-being, that measure is incoherent and may provide results that are off-target, especially if the measure treats different kinds of capabilities as if they can be traded off against one another.

In this section I give an extended example of Total Ordering Incoherence to clarify why practitioners should be concerned regarding the possibility that the measure may be off-target. I note here that the significance of possible off-target-ness depends on further combinations of context and purpose, which I consider in section five. I also note, following discussion in previous chapters, that the extended example holds for a class of capabilities measures that incorporate procedures often used to make decisions among options featuring different kinds of values.[31] The typical members of this class are measures that use what Sen (1980) calls a 'secondary' process of evaluation to render a Vector View of capabilities into total orderings of capabilitarian well-being, often by assigning weights to the values given to different indicators.[32]

Recall from the introduction in this chapter Decancq and Lugo's (2013) claim that the aptness of one or another method for assigning weights to indicators can be determined by checking the reasonableness of the tradeoffs implied by such weights. I apply Decancq and Lugo's (2013) practical test to the extended example from this section to highlight how the measure may be off-target. However,

[31] See especially chapters one and two.
[32] See chapter two section three.

in recognition of the practitioner who argues according to something like the Justification for Total Orderings on the Basis of Practical Comparisons (JTOBPC), I allow that incoherent procedures that always provide total orderings of capabilitarian well-being may still be justified as 'roughly useful' when deliberators agree with the implied trade-offs among different kinds of indicators. I also show, following an earlier discussion of Sen's (2009) nuanced treatment of the subject, that though implied trade-offs might be more intuitively acceptable for dyadic comparisons, trade-offs among three or more indicators are less intuitively acceptable.[33]

Recall that incoherence can be diagnosed in different ways, much in the manner of a flashing check engine light, as discussed in section two. Later in this section I consider how practitioners may 'misdiagnose' Total Ordering Incoherence, claiming the incoherence corresponds to procedures that combine different classes of data rather than different kinds of indicators standing for capabilities or capabilitarian well-being values. Authors such as Alkire and Foster (2011) and Alkire et al. (2015a), for example, tend to focus on problems related to combining ordinal and cardinal data and ignore the deeper philosophical problem of incoherence with respect to Sen and Nussbaum's commitment to qualitative distinctiveness.

Consider, then, as an extended example of Total Ordering Incoherence, an index that combines some individual or population's four different capabilities indicators in order to always provide complete rankings of that individual or population against others. According to Sen's characterization and representation, these indicators stand for basic achievements, or functionings, that in turn enable achievement of other sets of co-possible functionings. According to Nussbaum's characterization, each indicator stands for a different, qualitatively distinct dimension of capabilities – what I've also called a 'central feature' of Nussbaum's objective list representation.[34] For the first version of the extended

[33] See chapter two section four.

[34] Except for income, which does not reliably indicate one or another dimension, and so is at best a proxy for achievement of functionings in multiple dimensions. According to Nussbaum's definition, use of income as an indicator would provide a very rough, off-target indicator of basic capabilities as opposed to an approach that keeps capabilities dimensions, and their relative indicators, separate.

example, let the available data be a fairly typical combination of ordinal and cardinal indicators defined as the following dimensional variables:

a: Years of School, representing attainment of Education

b: QALYS (Quality-Adjusted-Life-Years), representing attainment of Health

c: Exclusion Survey Average (1-5 Likert Scale), representing attainment of Affiliation

d: Income Level in 10^3 USD, representing ability to attain other meaningful functionings

Values given by these indicators can be aggregated by the function V(x), which will assign a scalar value from indicator measurements for each of individuals or populations x_1, x_2 and x_3:

$$V(x) = (1/4)\,[\alpha_1\,(a/12) + \alpha_2\,(b/75) + \alpha_3\,(c/3.5) + \alpha_4\,(d/40)],$$

where α_i is the relevant weighting value given to the indicator such that setting α to 1 gives each dimensional variable equal weighting by means of the (1/4) multiplier.

For the sake of simplicity (and because most multidimensional capabilities indices do so as a baseline setting), let $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$. Denominator values corresponding to each dimensional variable then represent, on the Nussbaumian view, thresholds for each dimensional variable such that if individual x_i has an indicator value greater than that threshold, that dimension will take a scaled value greater than one.[35] For example, if an individual shows 16 years of school, 16/12 = 1.25 as a scaled value for the relevance of that dimensional indicator, where a scaled value of 1 would represent that the individual is on par with the 12 year threshold. Consider then x_1, x_2 and x_3 to be aggregated and ranked by V(), again with $\alpha_i = 1$, given that indicator values for dimensional variables for each individual are the following:

[35] In this extended example, denominator values are meaningless as they are absorbed by the coefficient value weights α_i which could take on values other than 1, yet they are included here to help illustrate how weight-setting works in practice. This treatment of denominator values as thresholds follows the Foster-Greere-Thorbeck (FGT) Method used by the Multidimensional Poverty Index (MPI). I return to an analysis of this method in the next chapter.

x_1: {a, b, c, d}* = {8, 50, 3, 10}, so $V(x_1) = .61$

x_2: {a, b, c, d} = {4, 60, 5, 8}, so $V(x_2) = .69$

x_3: {a, b, c, d} = {10, 70, 4, 15}, so $V(x_3) = .82$

*{a, b, c, d} are {education, health, social, income} indicators.

Thus x_3, with dimensional variables representing 10 years of education, 70 quality-adjusted life years, a Likert scale value of 4, and an income of $15,000/annum has a higher ranking than x_2, and x_2 has a higher ranking than x_1. Note also that V() could operate on any number of individuals x_i and completely rank their aggregated and averaged results.

If we follow Sen's characterization and representation, the ranking of individuals or populations is given along an ordinal scale at best. That x_3 is .13 scaled units higher than x_2 does not mean the interval separating x_3 and x_2 is greater than the numerically-smaller interval of .07 separating x_2 and x_1. This recognition is important because some practitioners treat the ordinal ranking as if the scale carries cardinal meaning.[36] I return to this point later in the section. All that we know from our ordinal index is that x_3 is more highly ranked than x_2 and x_1, but we cannot say anything about differences of the intervals of capabilitarian well-being among different individuals.

Let's also note that if we stay closer to Sen's representation, α_i values would in most cases exhibit *ranges of value*, not constants. For example, α_1 might have a ranged weighting of .7 to .9 while α_3 might have a ranged value of .5 to .8. In this case, rankings could shift: given those weighting ranges, x_2 and x_1 do not have a clear ranking against one another, and only x_3 maintains ranking dominance. In this case, the ranking is then incomplete, offering only a partial ordering in terms of capabilitarian well-being. The index, insofar as it incorporates procedures that always provide a complete ranking, provides an output

[36] See Figure 2 at the end of the section.

for situations for which there should be no fact of the matter in terms of either Sen's or Nussbaum's characterization and representation.[37]

Like many contemporary measures, our example index assigns α_i specific values such that a complete ranking can always be created. It is prudent, then, for practitioners to consider whether these kinds of results can be significantly off-target. That worry can be further investigated by applying Decancq and Lugo's practical test: the denominator values given in our index force us to accept, by means of some simple arithmetic, the following implied trade-offs:[38]

1 Year of School = 5.8 QALYs = .42 Likert Difference = $3,333 Yearly Income

This conclusion, or any similar conclusion that equates intervals among different types of capabilities indicators, will tend to break against people's common intuitions. Moreover, even if we allow for the rare case in which these averaged tradeoffs were accepted by deliberators as intuitive, it is difficult to find credible the further implication that either the second or the eleventh year of school is worth the same interval of income, Quality-Adjusted Life Years, or a certain interval of Likert-scaled social exclusion.[39]

Corresponding to the two concerns just given above, we can note that the extended example displays two notable senses of intransitivity, both of which highlight the significance of the possibility that the measure may be off-target. In the first sense, the same ratio that may justifiably hold for some interval of one indicator against that of another may not justifiably hold for intervals of further indicators. For example, deliberation regarding intersubstitutability of intervals according to established weights may include the credible belief that, on average, a year of school is worth $3,333.[40] The same deliberative

[37] Procedures that force point weights in order to always provide total orderings rather than incorporating ranged weighting are a substantial form of incoherence with respect to qualitative distinctiveness, but one that I will not pursue other than to claim that were α_i values given ranges of value instead of constant values, the relevant measures would provide a less distorted measurement outcome, though many rankings would be incomplete. Fuzzy sets do provide a representation that is closer to ranged weighting values. They are also more closely aligned with Sen's characterization of incompleteness and partial ranking, though for fuzzy sets there is a cardinal notion of *distance* that is again used to establish ranking. Distance is yet another artificial cardinality that functions as a common currency, the implied trade-offs of which would be worthy of further investigation. I'm thankful to José Manuel Roche of OPHI for this point.

[38] I note that these weightings and trade-offs are a toy example. Nonetheless, some set of weights is given for most indices according to similar procedures.

[39] This second point indicates the need for practitioners to check that indicators are sensitive to differences of valuation for different intervals, which is not a common procedure.

[40] To pump our intuitions, let's assume that in that context average yearly salary is in the hundreds of dollars.

body might also grant that a year of school is, on average, worth 5.8 QALYS, perhaps giving arguments testifying to the power of opportunities granted by learning to read, write, or do basic arithmetic. However, the same deliberative body might just as easily agree that 5.8 QALYS is never worth $3,333 (or even $33,333 and in many cases not even $333,333). I argue, following Aristotle's analysis from chapter one of the *asummetria* of different kinds of economic objects, that such 'common currency' equivalences among three or more different dimensions of value may function 'sufficiently' for some practical purposes, yet the measure often cannot capture all relevant values.

A second notable sense of intransitivity is that sub-intervals for different dimensions won't hold as common ratios. For example, the second year of education might intuitively be worth 5.8 QALYS (basic reading and writing are, after all, pretty important for most people), but the deliberative body may not express a similar tradeoff for the 7^{th} or 14^{th} year of education, and so on. This kind of intransitivity might be remedied by providing a more complex equation that provides α_i-type weightings for sub-intervals of one dimensional variable relative to another, perhaps in this case according to a quadratic pattern representing an increased value of subsequent years of education up to an optimum, followed by diminishing returns on extra years of education in terms of QALYs.

Yet such a remedy would only partially address a problem that compounds once we attempt to combine more than two variables. Imagine, for example, assigning different sub-interval weightings for the 2^{nd} year of education versus eight different QALY sub-intervals (say one for each decade), each of which have sub-interval weights for another *n* income sub-levels, etc. The practical implications for the second kind of intransitivity make it a significant problem, one that is perhaps avoided by the practitioner claiming that, *on average*, the values given by the weighted index will provide a measurement outcome that reliably indicates the patterned nature of the measured phenomenon. Yet faith in averages rests on the implausible assumption that variances of evaluation among different intervals of separate indicators

will magically cancel one another out in the aggregate such that the resulting pattern reliably indicates the target feature.[41]

Furthermore, consider the unintuitive implications for trade-offs when an excess in one dimension can be weighed against the lack in another, contradicting Nussbaum's characterization.[42] Is it credible, following the extended example above, to claim that an individual with half the sufficiency threshold of years of education (6/12) but sitting well above the threshold for income (50/40) is, on the whole, on par with an individual who is well over the threshold for years of education (15/12) but makes half as much per year (20/40)?[43] Nussbaum and Anderson would claim, as discussed in the previous chapter, that though a numerical trade-off has been established, it merely externalizes one dimension in terms of another. I take Anderson and Nussbaum to be indicating that our measurement of capabilities should capture this tragic component; we should recognize that procedures that always provide total orderings by treating deprivations of different kinds of capabilities as if they are intersubstitutable do not, in fact, capture what we hold to be important when we measure different types of capabilities. Our intuitions regarding those trade-offs help to indicate the significance of the concern that the measure may be off-target in terms of capabilitarian well-being.

Let me now return to the above issue that the incoherence may be 'misdiagnosed' such that practitioners do not fully address the significance of how the measure may be off-target. Though it should already seem like intransitivity holds for a range of cases greater than those that combine different classes of data such as cardinal with ordinal data, it is important to provide a response to the interlocutor who – perhaps following Alkire et al. (2015a) - looks at the extended example and notes that Likert Scales and QALYs are both arguably ordinal indicators, which means that comparisons of intervals

[41] For an excellent discussion of averages, balances, and their role in economics, See Harro Maas (2001): *An Instrument Can Make a Science: Jevons' Balancing Acts in Economics* as an example of why we might worry about the assumptions of systematic balance that underly practices of averaging.
[42] As I discuss in the next chapter, the Alkire and Foster Deprivation Depth measure engages in this kind of practice.
[43] Granted, this example uses the pseudo-dimension of 'income,' but we can imagine some similar example, say one that trades off, on average, longevity for better educational opportunities, etc.

among those and other cardinal indicators are already meaningless.[44] Consider, as an illustration, the problem of combining an ordinal Likert scale social exclusion indicator (1-5) with a cardinal measure of yearly income. Intervals between numbers on the former – say from 1 to 3 versus from 1 to 5 – are unlikely to represent the same degree of psychological magnitude, while intervals between numbers on the latter represent differences of currency that can be compared as ratios.[45] If the measure incorporates procedures that treat both ordinal and cardinal indicators as if they are intersubstitutable, the measure is certainly incoherent and may produce results that are significantly off-target.[46]

But does this problem with combining different classes of data capture *all* our concerns regarding how a measure that features Total Ordering Incoherence may provide results that are off-target? Let's consider an amended case in which the indicator values for dimensional variables are all cardinal, say for example the following:

a: Years of School, representing attainment of Education

b: Longevity (life span), representing attainment of Health

c: Number of Times Voting in National Elections, representing attainment of Civic Functioning[47]

d: Income Level in 10^3 USD, representing ability to attain other functionings

We don't need to add in all the math; it should be clear even at the outset that whatever ratios for trade-offs among of intervals are established for indicators a through d, those trade-offs will often break against common intuitions just as in the previous iteration of the extended example. Although years of school, years of life, participation in elections, and income level are all cardinal in nature, they cannot readily be

[44] See chapter one section one footnote two for clarification regarding cardinal (in which intervals carry meaning) versus ordinal (in which magnitudes of intervals don't matter, but the ordering does) data.
[45] The point is debatable; some claim that Likert scales are actually cardinal. For the purpose of this example, I assume ordinality.
[46] I would like to point out that some popular capabilities measures combine both ordinal and cardinal data in order to create rankings. The OECD Better Life Index (BLI), for example, includes ordinal indicators of self-reported health and life-satisfaction with numerous cardinal indicators.
[47] I recognize this is a tendentious indicator; I include it merely because it is something cardinal that relates to a functioning similar to social exclusion. The indicator would have to be tempered by the age of each citizen, so that a citizen who had the chance to vote in 3 elections but only voted in 2 would get a score of 2/3, while a citizen who had the chance to vote in 7 elections and voted in 2 would get a score of 2/7. The measure would already be flawed because younger citizens would be likelier to have higher scores. This kind of problem could be 'cleaned' from the data in order to provide a generally-reliable representation of the target phenomenon. In reality, it is challenging to find cardinally meaningful indicators for individually-valued social experiences.

treated as intersubstitutable if they are to accord with Nussbaum and Sen's commitment to qualitative integrity. Total Ordering Incoherence does not reduce to clashes of ordinal and cardinal data.

Yet, in their 2015 OPHI Working Paper 83 on Multidimensional Poverty Measurement, Alkire et al. downplay Total Ordering Incoherence in just this way. They recognize, in concurrence with Sen, that comparisons of two (or more) indicators in terms of their capabilitarian well-being may at best be expressed on an ordinal scale, though for Alkire et al. this recognition is best understood as a symptom of attempting to combine different classes of nominal, ordinal, or cardinal data.[48] Alkire et al. give little consideration to the deeper problem of qualitative distinctiveness, other than to recognize that some empirical measures may lack 'common units' such that their combination would not carry significant 'meaning.'[49] Alkire et al.'s threadbare discussion of 'meaning' trivializes the significant problem of incoherence with respect to qualitative distinctiveness in favor of extensive guidelines for combining different types of ordinal, cardinal, or quasi-ordinal (what they call 'ordered categorical') data. In fact, as the extended example in this section has demonstrated, there are further reasons for practitioners to be concerned regarding when different kinds of indicators are combined according to procedures that can always provide total orderings.

I would add a further set of practical concerns: procedures that always provide total orderings of capabilitarian well-being can be misinterpreted by practitioners *and* especially by layperson deliberators who treat intervals as though they are intersubstitutable, for example when they are used to provide measurements of how much one individual or population has progressed along the scale (see Figure 2 below).

[48] See Alkire, Foster, Seth, Santos and Roche (2015a) *OPHI Working Paper 83*, especially pages 27-33.
[49] Here 'meaning' is defined such that once transformations have been applied to measures with different scales, statements about the transformed data remain unchanged. See OPHI 83 pages 8, 30-33, 40, and 55 for relevant definitions of 'meaningfulness.' See the next chapter for an illustration of the kinds of capabilities are Alkire and Foster seem to assume have a 'common unit of account.'

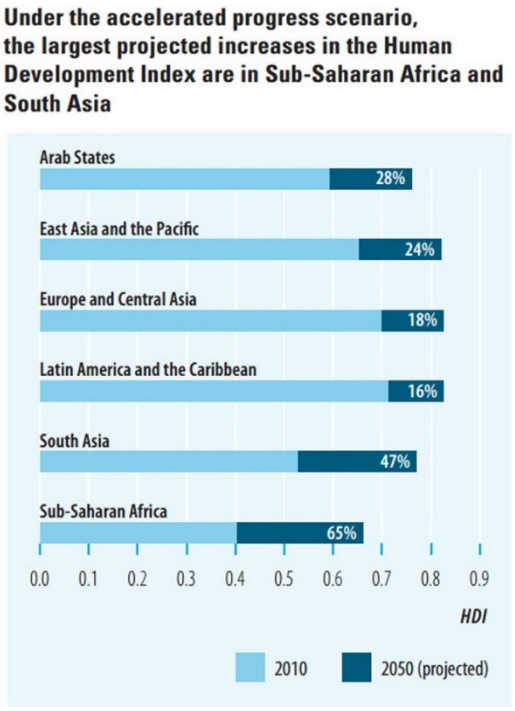

Figure 2: Projected Percent Increases in HDI 'Capabilitarian Well-being' from the 2013 UNDP Report p. 118

If the scale is at best ordinal, any measure of the magnitude of improvement is meaningless, especially as a means of comparing improvements across different populations. Figure 2 shows percentage improvement, a magnitude. The figure can be misinterpreted by readers who would assume, for example, that if East Asia and the Pacific had a 48% projected increase instead of a 24% percent projected increase, the magnitude of increase would be double. In fact, the percentage is based on weighted indicators that have been combined on an ordinal scale. Putting the significant problem of unintuitive intransitivity aside, all that can be known about a 48% increase on an ordinal scale is that it is ranked above a 24% increase. 'Percent increase' of capabilitarian well-being is misleading as it suggests the increase can be treated as if it carries cardinal meaning, when it is in fact only ordinal.[50]

[50] I would add, however, following the discussion in the previous chapter of practical uses of such measures as 'performance standards,' that figure two does provide a useful metric if it turns out that all participant states have agreed to the particular weightings of values that have been given by the standard, in this case the HDI. The rankings serve a competitive function among populations, which is perhaps not unhelpful, as Mahbub Ul Haq has noted, for furthering a certain kind of political conversation. These points aside, practitioners must still weigh the concern that the metric provided allows for serious misinterpretation, and can lead to significant harms if applied in more local contexts, as indicated by the extended example in this section.

Even more concerning in Figure 2 is the implied comparison of changes in magnitude among different populations.[51] The existence of such applications of total orderings of capabilities indicates a lack of awareness of the significance of Sen and Nussbaum's commitment to qualitative distinctiveness. As Sen notes in his 1992 Inequality Reexamined,

> Both well-being and inequality are broad and partly opaque concepts. Trying to reflect them in the form of totally complete and clear-cut orderings can do less than justice to the nature of these concepts. There is a real danger of overprecision here… This relates to a methodological point… that if an underlying idea has an essential ambiguity, a *precise* formulation of that idea must try to *capture* that ambiguity rather than lose it (Sen 1992: 48-49, original emphasis).

A multidimensional capabilities index that incorporates procedures that always provide complete rankings is 'overprecise' to the point of collapsing the ambiguity that characterizes the original concept. When we check the flashing check engine light that is Total Ordering Incoherence and take a closer look at how unacceptable are the kinds of tradeoffs implied by weightings used to create complete rankings, there is cause to be concerned regarding the combination of purpose and context to which the measure is lent. Not only should practitioners be concerned regarding the intuitiveness of implied trade-offs, but practitioners should be concerned regarding how the measure is interpreted, both by further practitioners and also, as I discuss further in the next section, by deliberators themselves.

[51] Though the UNDP has been careful not to engage in such a practice, a November 4th, 2010 article in the Economist titled *Developing Humans* does provide a table that compares percentage increases in HDI for different countries from 1980 – 2010. The UNDP does, however, track percentage increase per annum for each country measured by the HDI and groups different countries according to 'high,' 'medium,' or 'low' improvement (see hdr.undp/org/en/composite/trends).

4.5 Qualitative Distinctiveness, Value-Ladenness and Deliberation as Pragmatic Inquiry

The examples from the previous two sections provide particular demonstrations of how measures may provide results that are at odds with what is actually the case in terms of capabilitarian well-being. Though representing different kinds of capabilities indicators as if they are intersubstitutable is simple enough in the abstract, results can be significantly off-target as evidenced by how difficult it can be for people to accept implied trade-offs. The representation of capabilities as value-neutral or value-free also seems innocuous enough if evaluation is to be included at a later stage, yet the tendency to separate objectively-observable from value-laden content has significant implications at the level of operational procedures. Social power relations must be recognized and challenged, both during the selection of functionings and during any deliberational process of evaluation in order to address concerns related to Value Omission Incoherence.

Interestingly, social power relations can also be recognized and challenged in keeping with the commitment to treating different kinds of capabilities as qualitatively distinct. Let's consider further arguments given by Chad Kleist (2016) according to Feminist Discourse Ethics. Kleist argues that the right kind of deliberation regarding capabilitarian well-being must include a process of 'bearing witness' such that fellow deliberators do their best to inhabit the perspective of those giving expressions of valuation. Kleist also explains that "[Feminist Discourse Ethics] accepts the fact that there will likely always be incommensurability between discursive participants" (Kleist 2016: 196). By 'incommensurability' Kleist refers to what Ofelia Shchutte calls "residue of meaning that will not be reached in cultural endeavors," meaning some cultural and emotional valuation will be misinterpreted or lost among deliberators (ibid: 197). Kleist gives an example of a group of local stakeholders who repeatedly maintain that achievement of literacy is not important to them, even when deliberators argue that once the right kind of 'consciousness-raising' is applied, local stakeholders then appear to greatly value literacy and the kinds of functionings that become attainable along with the development of that

ability. Kleist maintains that in order to best allow deliberators to bear witness, procedures need to be those that allow for 'incommensurability' in deliberation.

I take Schute's culturally-oriented sense of 'incommensurability' to overlap in an important way with Sen and Nussbaum's qualitative distinctiveness. Schutte's incommensurability refers to an issue of translation: the same valuation can have different kinds of meanings depending on the complex set of cultural values people use to recognize those meanings. Recall from previous chapters that people who have had the right kinds of experiences are also able to recognize – and in some cases, reasonably articulate - different kinds of capabilitarian well-being values. I've also argued that in order to prudently capture the relevant values, deliberators should be committed to recognizing values separately.[52] Following this recognition, we can see the overlap with Schutte's 'incommensurability' in the context of this thesis: when deliberators accept that their abilities to recognize values are limited to their social and cultural perspective, they must also recognize that operational procedures must be those that treat different kinds of values separately, in order to best allow multiple perspectives to be recognized and considered.

Kleist argues that deliberators "should value incommensurability of meanings" because that commitment guides deliberation in the right kind of way: deliberators are "committed to dialogues as an ongoing process," they take an epistemic posture of "open-ness," and their recognition of incommensurability "creates a space for further discussion" (ibid: 197-198). These commitments are generally in line with those of Putnam's pragmatic inquiry (2017), in which Putnam claims that facts about social values can be discovered in deliberation – even across different frameworks of value -, though the facts created are influenced by the kinds of operational procedures that are incorporated. Recall from the previous chapter that the commitments of pragmatic inquiry include - inter alia - avoiding "relations of hierarchy and dependence," making experience-based evaluations that "respect new

[52] This claim follows from Nussbaum's discussion of *prohairesis*, commitments of Deweyan pragmatism, and Anderson's pragmatic theory of expressive value, each considered in the previous chapter.

descriptions," and maintaining a "consistent commitment to fallibilism."[53] Taken together, these commitments imply that, as Kleist maintains, deliberators should be open to and seriously consider new meanings when deliberating across different schemas of value. The picture provided by a process of pragmatic inquiry will likely be incomplete due to the commitment that some phenomena or values are understood according to diverse perspectives.

As I've explained in the previous chapter, the process of making prudent decisions in deliberation can be likened to one of unraveling a Gordian knot rather than using a reductive method to slice through different kinds of values. Sometimes a tangled picture of numerous different strands is the more accurate - and the more useful - representation. Recall also that this incomplete but on-target 'unraveling' process is the kind of process that Nussbaum describes as *phronêsis* according to *prohairesis* and that Anderson defends according to her pragmatic theory of expressive valuation. At the operational level of measuring capabilities, the commitments to value-ladenness and qualitative distinctiveness combine to frame deliberation such that – with inclusion of the right kinds of participatory methods – valuations given by local stakeholders are actual values that accord with the target concept of Sen and Nussbaum's capabilitarian well-being.

Yet a familiar set of JTOBPC-type practical concerns is, again, worth noting in this context. In the general incoherence claim given at the end of section two, I included the clause that 'off-target' means "there is good reason to believe the measure may not be measuring what it is supposed to, unless substantive argument is given to defend that the possibility of being off-target is not significant for that combination of context and purpose." Even when a measurement practice is incoherent, there might be sufficient practical reasons to proceed *as if* the measure is generally on-target - that it is roughly useful. In many cases, for example those in which there is some urgency to make a decision, or for which the significance of harms associated with being off-target are not unduly pressing, the need for a practical comparison may provide sufficient justification to proceed in spite of incoherence. However, as

[53] See chapter three section five.

evidenced by the demonstrations given above, practitioners should carefully investigate potential harms associated with how off-target the measure might really be.

Consider how arguments from the previous chapter might apply for such an investigation in the context of crafting a population-level measure of capabilitarian well-being. I have shown that, according to Nussbaum's characterization (and following Sen's commitment to the 'heterogeneity' of capabilities), measures should avoid procedures that compare different kinds of capabilities deprivations, especially as if deprivations can be traded off against one another or against excesses of other kinds of capabilities. However, I acknowledged that rough comparisons of capabilities above such thresholds are more defensible because less tragic. Recall also the discussion of simplified stand-ins from chapter two: it might be justifiable to represent the capabilitarian well-being of different items according to a simpler concept that functions as a common currency, especially if limited to more localized contexts– what I've called a 'Contextual Covering Consideration' following Ruth Chang's (2017) use of a similar term.[54] Yet, given that most population-level measures of capabilitarian well-being attempt to compare below-threshold levels of different kinds of capabilities indicators, we have a strong practical justification for keeping different kinds of population-level indicators distinct by avoiding procedures that weight indicators in order to always provide complete rankings.[55] I take this recognition to hold for most population-level measures of capabilitarian well-being that render different kinds of indicators into total orderings of items, unless, perhaps, the goal of such measures is to provide very rough assessments that satisfy a practical 'standard-setting' purpose such as that related in the previous chapter according to what Anderson calls Goodness-of-a-Kind judgments.[56]

While incorporation of procedures that use a simplified stand-in as a common currency - also often called a '*commensurans*' - may be justifiable for some limited contexts, such procedures challenge

[54] See chapter two section five.
[55] One notable exception to this general, practical claim, is the Dasgupta and Weale's Borda Count Method, in which weights are assigned for a reason that does not intuitively translate into treatment as if different kinds of indicators are inter-substitutable. For this reason, the Dasgupta-Weale method may have more justification than the typical decision-making process that assigns weights for each dimension as if those dimensions can be traded off against one another. But this is a question that is beyond my scope here. See Dasgupta and Weale (1992) "On Measuring the Quality of Life."
[56] See chapter three section five.

the commitment to a truly pluralistic process of deliberation as discussed in previous chapters and illustrated by the extended examples in this section.[57] Application of the commensurans in deliberation entails that there may be unintuitive trade-offs of different kinds of indicators, suggesting that deliberation may not be tracking what people really value in terms of capabilitarian well-being. Moreover, in more local contexts of deliberation, application of the commensurans might trample pragmatic commitments to open-ended inquiry. At the operational level, the concern is that social power relations can be leveraged such that the valuations expressed by those who have the most relevant experience are misconstrued, ignored, or rendered tacit. If the context is one of creating a population-level measure of capabilitarian well-being, it is not practical to treat different kinds of capabilities and capabilitarian well-being values as if they can be rendered in terms of one another.

Even if justification *can* be given to include incoherent procedures such as those outlined in this chapter, the possibly off-target nature of the measure should be made transparent to practitioners and deliberators, especially to communities who stand to be affected by concomitant policies. If we recognize that deliberation is a value-laden, political process, the commitment to a pluralistic, pragmatic inquiry requires ceteris paribus that all parties be given all possible information in an effort to prevent the implementation of procedures that can omit or otherwise mute the valuations that are most relevant for that context. Consider that, for example, what I would call a kind of 'technocratic leveraging' can be used to convince people in deliberation that once weights have been established, they should hold across different kinds of dimensions. And of course, in many instances such indicators *really can* be traded off, especially in pairs, such that deliberators look to such pairings and become convinced that the measure *really is* generally on-target. Yet in such cases, the tragic aspect of trade-offs is given less influence, especially when the measurement output is a single number that ranks individuals or populations against one another. If all parties are aware of the danger of this kind of leveraging from the start, deliberation can proceed in a different manner, such that if total orderings must be given, they are given according to

[57] Regarding *commensurans* see Claassen (2017) "An Agency-Based Capability Theory of Justice."

well-defined purposes, by deliberators who have considered the significance of the possibility that the measure may be off-target.

Chapter Five:

Analysis of Contemporary Measures:
Krishnakumar's Improved Human Development Index
and the Alkire and Foster Measure of Deprivation Depth

5.1 Introduction

In this chapter I analyze two capabilities measures that are incoherent with respect to Sen and Nussbaum's concept, including their commitments that capabilities are value-laden and are resistant to procedures that can always provide total orderings. In the previous chapter I gave an extended example to illustrate how *Value Omission Incoherence* can be manifested when measures treat capabilities as value-free or value-neutral. In such cases, relevant valuations may not be included or given influence during processes of evaluation, for example in forums of pluralistic deliberation. In this chapter I consider how a set of statistical procedures used to set weights among different kinds of capabilities indicators can be incoherent with respect to how Sen and Nussbaum characterize value-ladenness.

In the previous chapter I also gave an extended example of how capabilities measures that can always provide total orderings - also called 'complete rankings' - of capabilitarian well-being manifest *Total Ordering Incoherence*. I demonstrated, using a simplified index, that there can be cause for concern when the measure provides rankings for cases for which there should be no fact of the matter in terms of Sen's or Nussbaum's concept. I also explained that when procedures treat different kinds of capabilities as if they can be traded off against one another, there is further cause for concern regarding the significance of the possibility that the measure may be off-target. As Decancq and Lugo (2013) argue, this concern can be further investigated by considering the intuitiveness of trade-offs among different indicators according to the weights assigned to them.

The first measure I analyze, Jaya Krishnakumar's improved Human Development Index, not only includes procedures that exhibit *Total Ordering Incoherence*, but sets weights according to statistical methods that treat different kinds of indicators as if they are not value-laden in the sense that is relevant

for Sen and Nussbaum's concept. Sabina Alkire and James Foster's 'AF Method' more clearly retains - in various stages of the measure - Sen and Nussbaum's commitments to both the value-ladenness of different kinds of capabilities and the resistance to procedures that can always render them into total orderings of items. However, depending on the operational procedures incorporated, the 'AF Method' measure can be incoherent with respect to the commitment to the qualitative distinctiveness of different kinds of capabilities. In the following sections I discuss the ways in which both the Improved Human Development Index and AF Method measures can be incoherent, and I conclude the thesis with a set of guidelines for practitioners working to measure Nussbaum and Sen's concept of capabilitarian well-being.

5.2: Krishnakumar's Improved Human Development Index (Ĥ)

Let's consider the improved human development index (Ĥ) proposed by Jaya Krishnakumar (2007) in her article "Going Beyond Functionings to Capabilities: An Econometric Model to Explain and Estimate Capabilities." As the title indicates, Ĥ includes an explanatory model that measures statistical connections between unobservable dimensions of 'capabilities' and various observable indicators that can be used to represent those dimensions. For example, Ĥ includes observable indicators of 'adult literacy rate' and 'combined educational enrolment ratio' that together represent an unobservable – also called 'latent' - dimension of capabilities called 'knowledge.' If an individual has more or less 'knowledge,' then the model shows a correlated increase or decrease in the relevant observable indicators. Like the UNDP Human Development Index (HDI) that precedes it, Ĥ includes procedures that combine multiple indicators – together standing for three different 'dimensions' of capabilities – a way that provides a total ordering of capabilitarian well-being. For example, indicators like 'adult literacy rate' and 'combined educational enrolment ratio' are assigned weights and combined with others into a result that provides a total ordering of improved 'Human Development' for different individuals or populations.

It could be argued that because Krishnakumar tends to call the target concept 'human development,' the Ĥ measure may not actually measure 'capabilitarian well-being,' and so the measure may not be incoherent with respect to Sen and Nussbaum's concept. However, Krishnakumar notes that human development "is given by the enhancement of the *set of choices or capabilities* of individuals," and so the measure targets capabilities as characterized within Sen and Nussbaum's approach (ibid: 39, *emphasis added*). Furthermore, Krishnakumar refers to Ĥ as an "aggregate capability index" (ibid: 53), citing both Sen's *Development as Freedom* and *Commodities and Capabilities*. For these reasons, Ĥ is a measure of Sen and Nussbaum's 'capabilitarian well-being' as defined in this thesis.[1]

[1] See chapter one regarding 'capabilitarian well-being.'

By citing Sen and not Nussbaum, Krishnakumar allegedly attempts to measure Sen's characterization and representation of the concept. And to be fair, Ĥ does at least partially follow Sen's commitment to the holistic evaluation of sets of possible sets: the explanatory model that informs Ĥ allows that all three dimensions of health, knowledge, and political freedom can interdependently influence individual indicators. However, each dimension in Krishnakumar's structural equations model is postulated as having a separate kind of influence. For example, the model gives the correlation between the latent 'knowledge' dimension and the 'combined enrolment ratio' indicator (ibid: 45). Recall from previous chapters that the representation of capabilitarian well-being according to *dimensions* of capabilities more closely follows Nussbaum's characterization of an objective list of 'central features' or 'spheres' than it does Sen's characterization of lives worth living. Inasmuch as the model measures degree of correlation between different individual dimensions and relevant indicators, it follows Nussbaum's commitment that different dimensions of capabilitarian well-being are qualitatively distinct. In fact, Krishnakumar and Chávez-Juárez (2014) appear to reconcile this issue when they update the model by treating 'capabilities' as a single dimension. I return to this point later in the section.

If Ĥ more closely follows Nussbaum's characterization, then the process by which the measure can always provide total orderings is incoherent not only with respect to Nussbaum's more stringent commitment to the qualitative distinctiveness of different dimensions, but also her stronger moral commitment to avoiding tragic trade-offs among such dimensions.[2] I return to this issue of tragic trade-offs just below. Even if we were to charitably allow that Ĥ better follows Sen's characterization and representation - perhaps because some set of arguments can be given to justify that indicators in Ĥ might function better as proxies for Sen's sets of co-possible sets rather than for Nussbaum's 'dimensions' -, Ĥ still exhibits *Total Ordering Incoherence*. The measure provides rankings in terms of capabilitarian well-being for cases where there may be no fact of the matter in terms of Sen's - and Nussbaum's -

[2] See chapter three regarding the philosophical justification for treating different dimensions of capabilities as qualitatively distinct, following the Aristotelian view of 'separate spheres' of human flourishing.

capabilitarian well-being.[3] Further analysis is warranted depending on the significance of the possibility that the measure may be off-target.

One way we can further the analysis is by considering the process for assigning weights to different dimensions in Ĥ. Krishnakumar's model uses factor analysis to measure the 'explanatory power' of each of the three different dimensions of capabilities, and then uses the related 'factor score' to assign weights to those dimensions. The factor analysis procedure compares how well different linear combinations of latent dimensions - each represented as a variable in a structural equations model - are able to account for the variance of related observable indicators, producing a 'factor score' for each variable. Krishnakumar takes the inverse variance of each factor score against the weighted average of all factor scores as a measure of the 'precision' of each latent variable, and then scales this measure of precision into the weight assigned to each latent variable. Thus, instead of the 1/3, 1/3, 1/3 so-called 'neutral' weighting scheme applied to the three dimensions of the original HDI, the explanatory model informing the Ĥ index assigns weights of .124, .436, .440 for dimensions of education, health, and political freedom respectively. These weights – in conjunction with inclusion of political indicators rather than income indicators – are then used to provide a set of complete rankings for Ĥ, one that arguably improves on those provided by the earlier HDI.

The Ĥ index is certainly a statistical refinement of the original HDI, but the use of factor scores to assign weights is based on the tenuous assumption that how well latent variables account for variance of indicators is a sufficient basis for weighting different dimensions of capabilities relative to one another. Though the rankings in Ĥ do reflect some normative valuation given that certain indicators have been selected as representative of capabilities according to how well they fit within the explanatory model, the set of procedures for deriving Ĥ provides little or no opportunity for deliberation regarding how different populations might value and evaluate those functionings.[4] The practice of assigning weights according to

[3] See chapter two.
[4] See chapters two and three regarding Sen's commitment to pluralist deliberation and Nussbaum's commitment to *phronêsis* according to *prohairesis.*

factor scores treats indicators as something other than the sense of 'value-laden' supported by Sen and Nussbaum's characterizations and representations. People who might be stakeholders for related policy decisions are not given opportunity to express or recognize valuations (or both), ideally in a setting of public deliberation.

There are further reasons to be concerned regarding the use of factor scores to assign weights for different dimensions of capabilities. Decancq and Lugo (2013) state a related concern that "derivation of weights through principal component analysis [including factor analysis] or explanatory models is not straightforward and lacks transparency." Because there is some disagreement among statisticians regarding when and how to implement different methods of factor analysis or similar forms of component analysis, models that incorporate such methods can exhibit what Henk Elffers calls "indeterminacy of factors and factor scores" (Elffers 1980: 318). So-called 'precision' weights of .124, .436, .440 in $\hat{H}$ might shift significantly depending on the statistical method incorporated into the model, and so the econometric process lacks transparency. Moreover, as I have discussed in the previous chapter, even if the particular method for creating a factor score is assumed to be sufficiently defensible as a statistical practice, Decanq and Lugo are concerned that "statistical approaches can lead to counter-intuitive weights" (Decanq and Lugo 2013: 13). Krishnakumar provides no method for checking what Decanq and Lugo call the "reasonability [of the weighting scheme] in terms of implied trade-offs between the dimensions" (Decancq and Lugo 2013: 20), adding to the concern that the measure fails to fit with the value-ladenness of capabilities in the manner intended by either Sen or Nussbaum.[5]

Let's return to the issue of tragic trade-offs, especially because $\hat{H}$ appears to follow Nussbaum's commitment to treating different dimensions of capabilities as distinct. Recall the extended example from the previous chapter where I illustrated, using a simplified capabilities index, that implied trade-offs among different kinds of capabilities indicators can be intransitive with respect to people's deep intuitions. For example, let's assume that in accord with the weights given for different dimensions

[5] See the previous chapter for a demonstration of Decancq and Lugo's 'practical test' of common intuitions regarding implied trade-offs indicated by assigned weights.

within $\hat{H}$, people in the target population express - as a practical matter - the common intuition that some unit of life expectancy is substitutable with some unit of adult literacy, say in a X:Y ratio. Let's similarly assume that people express a common intuition that some unit of 'political rights' as scored on a 0-6 scale can be roughly traded for some unit of adult literacy, say in a Z:Y ratio. Yet the same population may not express the common intuition that life expectancy and adult literacy can even be roughly traded off in an X:Z ratio. Moreover, even pairwise trade-offs suggested by weights assigned according to factor scores (e.g. X:Y or Y:Z) may not track people's common intuitions. For example, it might not be intuitive to the relevant population that differences in indicators standing for the dimension of health are roughly intersubstitutable with differences in indicators standing for the dimension of political freedom (according to weights of .436 and .440, respectively). Factor scores set the weights $\hat{H}$ by the fiat of explained variance, with little chance of deliberation.

A further concern is worth noting here: inasmuch as the $\hat{H}$ measure fits with Nussbaum's characterization and representation of capabilitarian well-being, practitioners should recognize there are further practical and moral reasons for avoiding procedures that render different kinds of capabilities as if they can be traded off among one another. Though I claim that a check against common intuitions would likely find significant mismatch between trade-offs suggested by factor score weightings and how such trade-offs are valued by the relevant deliberative body, even if there were a general match with common intuitions, practitioners would need to be wary of a measure that suggests an increase in one dimension of capability is a sufficient substitute for a lack of another. It is true that policy-makers, as a practical matter, must often choose how to allocate resources, yet the use of a such a metric might provide a harmful shortcut through the process by which relevant stakeholders would otherwise express influential valuations in terms of Nussbaum's capabilitarian well-being.[6]

In a more recent (2014) paper, Krishnakumar and Florian Chávez-Juárez move away from using factor analysis to assign weights that are then used to provide total orderings of capabilitarian well-being.

[6] See especially chapters three and four regarding problematic reductions of relevant processes of deliberation and decision-making.

They instead further the kind of analysis that can be done with an explanatory model. This line of investigation is promising and highlights a unique application of structural equations modeling for investigating both the existence of capabilities and the effects of different kinds of observable indicators on one another within the model. Unlike the 2007 version, Krishnakumar and Chávez-Juárez's 2014 explanatory model investigates only one dimension of capabilitarian well-being at a time. Because of this feature, it is unclear how the 2014 model can capture ways in which different Nussabaumian 'dimensions' of capability might simultaneously influence observable indicators. Krishnakumar and Chávez-Juárez (2014) sidestep this question by calling the single latent dimension 'capabilities.' Here the use of a plural name for a single dimension begs the question whether there really are - following Nussbaum - different 'dimensions' of capabilities that exhibit separate correlations with different observable indicators within a structural equations model, or whether - following Sen - there are just bundles of indicators standing for sets of co-possible sets of attainable functionings, which themselves can be called 'capabilities.' I bracket further discussion as it wanders from our present focus. The (2014) explanatory model of 'capabilities' at least remains useful for analysis of one 'dimension' at a time, using one or more indicators for each dimension.

Furthermore, the explanatory model provides statistical evidence that can be useful for informing policy decisions. In the 2007 paper, for example, Krishnakumar claims, on the basis of the results indicated by the model, that when the right kind of "support system is provided in an adequate manner we see that not only does it enhance people's capabilities, but leads the system to the path of a 'virtuous' development cycle due to the positive interactions among different dimensions" (Krishnakumar 2007: 57). Though the interpretation of this macroeconomic claim depends in a large part on the assumptions built into the model, it is certainly informative. What Krishnakumar means by a 'virtuous cycle' is that a certain amount of political freedom and general infrastructure "can promote development, but it is also true that development [of capabilities] in turn encourages favourable political and social arrangements by making people more and more aware, involved and demanding, and enforces the participatory element of

progress" (ibid: 41). The model provides evidence supporting the claim that when policy favors certain forms of political education that enable and empower citizen participation, such policy also establishes a feedback mechanism that further increases freedom in terms of capabilitarian well-being, which then allows for increased political education, and so on. This particular claim does illustrate the kind of practical potential that an explanatory model can have for informing policy, even if the model is incoherent with respect to Sen and Nussbaum's concept.

5.3 The Alkire and Foster Measure of Deprivation Depth

In this section I discuss various forms of incoherence that feature in Sabina Alkire and James Foster's 'AF Method,' a matrix-based measure of capabilities that has been incorporated into the OPHI (Oxford Poverty and Human Development Initiative) Multi-dimensional Poverty Index, the MPI. As in the case of Krishnakumar's $\hat{H}$, even though Alkire and Foster mostly cite the characterization work of Sen, their AF Method more closely follows Nussbaum's characterization of different 'dimensions' of capabilities. Each dimension in the AF Method is represented as a different column in a matrix, while different rows on the same matrix represent different individuals. The AF Method incorporates 'deprivation cutoffs' that function as minimal sufficiency thresholds - or social minima - for each column. Further operational procedures can then be implemented to provide a 'headcount' of the number of individuals below a threshold for a given column - what Alkire and Foster call a measure of 'capability deprivation' for that dimension.

Because it separates 'dimensions' of capabilities and because it measures sufficiency thresholds for each dimension, I argue that the AF Method measure follows Nussbaum's characterization of 'central features' of capabilities, each of which can be given a sufficientarian threshold. Even so, the analysis that I give here is in accord with Sen's characterization. At various stages of the measure, the AF Method does follow Sen and Nussbaum's commitments to the value-ladenness and the resistance to procedures that can always provide total orderings. Yet, depending on the procedures incorporated, some stages of the AF Method measure contradict both Nussbaum and Sen's commitment to the qualitative distinctiveness of different kinds of capabilities, supporting a form of *Total Ordering Incoherence.* The AF Method can be used to provide not only the aforementioned headcounts of capability deprivations, but also a measure of 'deprivation depth' that compares the relative magnitude of deprivation across different kinds of capabilities indicators. Both the headcount and deprivation depth stages are incoherent inasmuch

as they can be used to always provide total orderings across different kinds of capabilities, especially when there simply may be no fact of the matter in terms of Sen or Nussbaum's concept.

Alkire and Foster offer their headcount measure as a way to meaningfully combine - and in some cases compare - different classes of indicators. Some indicators are cardinal such that intervals between units on the same scale are meaningful. For example, the indicator for years of education is cardinal because the difference between two and four years of education is at least temporally the same as the difference between five and seven years.[7] Other indicators are ordinal in that the numbers on the relevant scale indicate only rankings, not common meanings across similar differences of interval. For example, many Likert survey results are ordinal in that they indicate ranking only; the difference between an ordinal Likert score of 3 and a score of 5 does not carry the same meaning as the difference between a score of 1 and 3 on the same scale. Still other indicators are nominal in that they indicate a categorical distinction rather than a ranking - for example whether one has a concrete floor or is a member of some social category, etc. Many nominal indicators can be represented as a binary - either a 1 or a 0 - to indicate whether the category holds for the individual or not.

The g^0 stage of the AF Method reduces all indicators to a binary value, allowing nominal indicators to be combined with cardinal and ordinal indicators. Ordinal and cardinal indicators are reduced depending on whether an individual is below the relevant threshold for the given dimension of capabilitarian well-being. To illustrate, let's say the indicators selected are cardinal years of education, an ordinal Likert score of social exclusion, and a nominal indicator of sufficient access to running water. Let's also posit the relevant thresholds, respectively, at *n* years, a score of *p*, and sufficient access or not depending on responses given to a qualitative survey. Individuals who attain below *n* years or below *p* or who lack access to water are represented in the appropriate column of the g^0 dimensional matrix with a 1, while individuals above are represented with a 0. Through this kind of reduction, the AF Method enables comparison and even methods for aggregating different classes of indicators along single dimensions, for

[7] Of course, different years of education might connect with different kinds of values, and so there is a sense in which even this measure is not cardinal in terms of how it might be evaluated. See the discussion surrounding the simplified index example in the previous chapter.

example by providing a population headcount of how many individuals within the population have fewer than *n* years of education, fall below *p*, lack access to water, some combination of the three, etc.

The g^0 stage of the AF Method allows for two basic forms of aggregation: it can be used to provide a headcount of the total number of individuals below the threshold for each dimension by adding the reduced elements of each column in the matrix, or it can provide a count of the total number of different deprivations that each individual has by adding the reduced elements of each row (or both). Because it does *not* support processes that render different kinds of indicators as if they can all be traded off against one another, the AF Method headcount of individuals for a single dimension is coherent with respect to Sen and Nussbaum's commitment to treating different kinds of capabilities as qualitatively distinct. The latter count of an individual or population's total deprivations, though perhaps useful as a rough indication that some individuals or populations face more total deprivations than others, is incoherent with respect to Sen and Nussbaum's commitment to qualitative distinctiveness in that it gives different kinds of deprivations equal weighting. For example, if individual g_1 is below the threshold in two dimensions, say education and political freedom, while another individual g_2 is also below the threshold in two dimensions, for example education and health, then both individuals exhibit the same count of total capability deprivations, yet the count problematically represents a deprivation in political freedom as equal to a deprivation in health. The count suggests an equivalence of value, as if one can be traded for the other.

This problem is reminiscent of Annapurna's tragic decision discussed in previous chapters. Recall that because different kinds of values are relevant for the distinct kinds of deprivations faced by Dinu, Bishanno, and Rogini, Annapurna may not be able to provide a complete ranking of their different situations in terms of capabilitarian well-being, even if she can, as a practical matter, make a decision to award the gardening job to one of the laborers. As has been argued in previous chapters, it may be more prudent for Annapurna to consider the situation of each laborer as a particular rather than for her to make

an assessment based an abstract representation of all three situations.[8] A representation that treats one kind of deprivation as if it can be substituted in terms of another, for example in order to derive a total ordering of capabilitarian well-being, may be a representation that is problematically reductive if it is to be used to inform Annapurna's difficult decision.

To be fair, let's note that there are certainly cases for which headcounts across different dimensions might still remain coherent with respect to Sen and Nussbaum's commitment to qualitative distinctiveness. For example, if individual g_3 is below the threshold in three dimensions, two of which are the same as those of individual g_1, then there is a defensible claim - similar to that for dominance rankings discussed in chapter two - that g_3 exhibits a greater ranking of capability deprivation than does g_1. This special case of complete dominance aside, the g^0 representation that best coheres with the commitment to the qualitative distinctiveness of different kinds of capabilities is a headcount for a single dimension at a time rather than a headcount across dimensions. As Sen has noted many times, it should not be an embarrassment if there are some cases for which the measure provides an output that is 'incomplete.'

Alkire and Foster incorporate further operational procedures into their measure that are incoherent with Sen and Nussbaum's commitment to qualitative distinctiveness. In the context of the AF Method g^0 measure, different dimensions can be assigned weights according to their relative importance, after which they can be aggregated and then divided by the total number of individuals in a population in order to create a measure of the 'intensity' of capability for that population. This set of procedures treats different kinds of indicators as if they are intersubstitutable, assigning weights to indicators as a representation of the evaluation of those indicators in terms of their relative importance. I return to this evaluational issue just below, for now claiming that the AF Method measure of capability deprivation intensity is incoherent in the same way that the 'headcount' of total count of deprivations for each individual is incoherent.[9] Such procedures can always allow for total orderings, in this case in terms of intensity of deprivation of capabilitarian well-being, and so exhibit *Total Ordering Incoherence*. Further

[8] See chapter four section three.
[9] See Alkire et al. OPHI Working Paper 86: 3 for a definition of the intensity measure.

investigation may be warranted depending on the significance of the possibility that the measure may be off-target with respect to Nussbaum and Sen's concept.

The AF Method can also be used to measure 'normalized deprivation gaps' that indicate how far some individual is below the relevant threshold. The value is 'normalized' with respect to the numerical value of the threshold, allegedly allowing for better comparison of deprivation gaps across different kinds of indicators. For example, if the threshold for years of education is 12 years and g_4 is only able to attain 8 years while g_5 is able to attain 10 years, then g_4 has a normalized deprivation value of (12 - 8)/12 = .25 while g_5 has a normalized deprivation value of (12-10)/12 = .17. If, by comparison, the threshold for life expectancy is 60 years and g_4 is able to attain only 45 years while g_5 is able to attain 50 years, normalized deprivation values would be the same for either individual in terms of both years of education and years of life (.25 and.17, respectively), providing a basis for weighted comparison across indicators.

Before we turn to issues of comparison across indicators, let's note that a deprivation gap measure for a *single* indicator is not incoherent with respect to Nussbaum and Sen's commitment to the qualitative integrity of different kinds of capabilities. The measure can provide a meaningful basis for comparison along a single indicator or, in keeping with Nussbaum's characterization, for multiple indicators within a single dimension. For example, a comparison of the normalized deprivation gap value for g_4 versus that of g_5 does at least provide a meaningful ordinal ranking for years of education or for years of life, much in the same way that we can say, according to Sen's representation, that the capability set for g_5 dominates that of g_4, at least in terms of that single indicator. However, the AF Method can then incorporate further procedures that aggregate normalized deprivation gaps *across* different kinds of indicators in order to provide an overall measure of 'depth of multidimensional poverty,' at which point the measure once again exhibits *Total Ordering Incoherence*.

Aggregation of deprivation gaps across different kinds of indicators is incoherent with respect to Sen and Nussbaum's commitment to qualitative distinctiveness because it treats different kinds of indicators as if they are intersubstitutable in ways that can easily defy common intuitions, or - again

following Nussbaum's characterization - that can allow for tragic tradeoffs among deprivations. For example, if individual g_6 has a normalized deprivation gap for years of education of .25 according to a 12 year threshold while individual g_7 has a normalized deprivation gap for life expectancy of .25 according to an 80 year threshold, then the measure that aggregates both deprivation gaps in order to provide total orderings of deprivation depth treats four years of education as equivalent to twenty years of life. Such a tradeoff implicitly carries a value judgment, one that may be a harmful reduction of the way either indicator would actually be valued or evaluated by a deliberative body. Of course the normalized deprivation gaps could be assigned weights such that they better mesh with common intuitions regarding pairwise tradeoffs, and the AF Method includes a variety of procedures for assigning such weights. However, as illustrated in the previous chapter, problems of intransitivity become likely as weights are assigned to more indicators. And regardless, the underlying concern remains: the measure provides total orderings for situations for which there simply may be no fact of the matter in terms of Sen and Nussbaum's capabilitarian well-being.

The point of this analysis is not to claim that AF Method measures of either the intensity or, similarly, the depth of multidimensional poverty are meaningless, nor that they lack practical value. In fact, the intensity measure is a component of the Multidimensional Poverty Index (MPI), which, as Alkire and Santos note in their article on "Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index," provides highly robust pairwise comparisons of country rankings when weights ranging from 25% to 50% are applied to each dimension of deprivation (Alkire and Santos 2014: 260). Part of the reason that MPI rankings remain robust is because, as Jonathan Wolf and Avner De-Shalit have argued in their (2007) *Disadvantage*, capability deprivations tend to cluster such that individuals that exhibit deprivations for a few indicators have an increased likelihood to exhibit deprivations for other indicators. Accordingly, populations that have lower rankings according to one indicator or dimension of capabilities are likely to exhibit similar rankings according to other indicators.[10]

[10] I'm thankful to Nicolai Suppa of OPHI for clarifying this point.

However, even though rankings of the MPI may be robust, it must be recognized that the measurement output produced remains incoherent with respect to the characterization of capabilities because it includes a measure of intensity of deprivation across different kinds of indicators. I follow Decancq and Lugo (2013) and also Ravallion (2010) in arguing that the MPI measure could be better justified - in this case with respect to Sen and Nussbaum's commitment to qualitative distinctiveness - by establishing that trade-offs among indicators suggested by assigned weights consistently follow common-sense intuitions of individuals whose deprivations are being measured, especially when individuals are stakeholders in policies informed by the measure. *Total Ordering Incoherence* warrants this further investigation, especially if the measure is used to inform policies that can have significant effects on the well-being of stakeholders.

5.4 Guidelines for Practitioners

In this section I provide a set of guidelines for practitioners seeking to *measure* capabilitarian well-being in a manner that is coherent with the concept as given by Sen or Nussbaum (or both). Recall from chapter one the Aristotelian discussion regarding different kinds of economic goods that feature different kinds of values - including capabilitarian well-being values. Such goods are resistant to being *measured* as if they can all be traded off against one another in terms of a common currency. Though their common 'exchange value' supports and is reinforced by certain market practices, the exchange value is not a measure because it may not capture the relevant well-being values. Recall also that well-being, of which capabilities is a species, is a *ballung* concept such that it exhibits different features - including different kinds of values - depending on the relevant combination of context and purpose. If the measure is to capture different kinds of capabilitarian well-being values, it may be important to avoid operational procedures that can always provide total orderings, especially those that treat different kinds of indicators as if they can all be traded off against one another. I note, following Sen, that measures of capabilitarian well-being can provide partial rankings and, similarly, might be amenable to special forms of aggregation, for example as an AF Method headcount of individuals exhibiting deprivation for a *single* dimension of capabilitarian well-being, as discussed in the previous section.

However, even when measures include delimited aggregation procedures or otherwise remain committed to providing partial rankings, it remains important for practitioners to make transparent any representational features or procedures that transgress Sen and Nussbaum's commitments to either the value-ladenness or the qualitative distinctiveness of different kinds of capabilities indicators and capabilitarian well-being values. For example, practitioners and deliberators should be aware that procedures for selecting 'value-free' functionings or that assign weights according to statistical properties such as factor scores rather than valuations given in deliberation may not capture the kinds of values that people would find relevant for measuring capabilitarian well-being. Similarly, practices that render

different kinds of indicators intersubstitutable may be problematically reductive representations of tragic trade-offs. Here a guiding concern is that expressions of value that should be given significant influence in public deliberation - for example those related to the experiences of local stakeholders - may be rendered uninfluential by one or another set of procedures, or by features of the representation. The practical justification for maintaining transparency regarding incoherence is that in many contexts - for example those in which the measure is influential for making policy decisions, awarding resources, or informing the public -, even a slightly off-target output can have significant effects concerning people's well-being.

In summary, population-level measures of Sen and Nussbaum's 'capabilitarian well-being' should, when possible, follow or at least make transparent the commitments that different kinds of capabilities indicators are both value-laden and qualitatively distinct, that capabilitarian well-being values are qualitatively distinct, and that capabilities are generally resistant to procedures that can always provide total orderings of capabilitarian well-being. Let's consider a set of more specific guidelines that follow different interventions defended throughout this thesis.

With respect to Sen and Nussbaum's commitment to value-ladenness:

1) Avoid representing different kinds of capabilities indicators as 'value-free' or 'value-neutral,' especially during the process of selecting important functionings to be evaluated.[11]

2) Operational procedures for selecting and evaluating capabilities should include participatory methods that address social power dynamics and provide both recognition and influence to local stakeholders or others whose experiences are relevant in context.[12]

3) Operational procedures for selecting and evaluating capabilities should be ongoing dialogues – what can also be called arenas of 'pluralist deliberation' - that follow

[11] See chapter four sections three and five.
[12] See chapter one section four, chapter two section two, chapter three section four, and chapter four section five.

principles of 'pragmatist inquiry' including a commitment to fallibilism, making full use of human reasoning (including expressions of deep emotion), investigating in a cooperative and democratic manner, and avoiding relations of hierarchy and dependence.[13]

With respect to Sen and Nussbaum's commitment to qualitative distinctiveness, also called 'non-commensurability' of different kinds of capabilities:

4) Avoid representing capabilities such that different kinds of indicators can be assigned weights as if they are intersubstitutable.[14]

5) Avoid procedures that can always provide total orderings of capabilitarian well-being by combining different kinds of capabilities dimensions.[15]

6) Promote transparency regarding the claim that capabilitarian well-being values are of different kinds such that deliberations regarding well-being avoid reductive trade-offs of deprivations of central features of capabilities such as education, affiliation, health, political freedom, etc.[16]

7) Avoid procedures that reduce different kinds of capabilities dimensions into a common currency such that different kinds of indicators can be substituted according to that common currency.[17]

8) Promote transparency regarding cultural or experiential 'incommensurability' among members of a deliberating body such that members are made aware that expressions and valuations can be understood differently according to different cultural norms or particular life experiences. Promote revision and continuous dialogue as part of the

[13] See chapter one section four, chapter two section two, chapter three section five, and chapter four section five.
[14] See chapter one section two, chapter two sections three and four, chapter three section three, and chapter four section four.
[15] See chapter one section two, chapter two sections three and four, chapter three section three, and chapter four section four.
[16] See chapter one section four, chapter two section four, chapter three sections two and three.
[17] See chapter one sections two and three, chapter two sections three and five, and chapter four section four.

process of measuring what a group or population would continue to call 'capabilitarian well-being.'[18]

In previous chapters I have made clear, following Anna Alexandrova (2017), that different well-being measures can be justified as more or less appropriate relative to context and purpose. Accordingly, there are combinations of context and purpose that may justify the use of representations or operational procedures that persist in rendering the measurement system incoherent with respect to Sen and Nussbaum's commitments to either value-ladenness or qualitative distinctiveness. For example, as discussed in earlier chapters, a multidimensional index of capabilitarian well-being might be useful as a 'performance standard' that helps to further a political conversation, especially as it offers a richer informational space than that provided by a single macroeconomic measure like GDP per capita. Even when they are incoherent such that they may be off-target with respect to Sen and Nussbaum's concept, such measures can still be useful for limited combinations of context and purpose. Regardless, in many cases their transgression warrants further investigation.

[18] See chapter one section four, chapter two section two, chapter three section four, and chapter four section five.

Bibliography

Abubakar, Ayesah Uy. (2019) *Peacebuilding and Sustainable Human Development: The Pursuit of the Bangsamoro Right to Self-Determination*. Vol. 16. Springer, 2019.

Adcock, Robert, and David Collier. (2001) "Measurement validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95.3 (2001): 529-546.

Agarwal, Bina. (1997) "'Bargaining' and Gender Relations: Within and Beyond the Household." *Feminist Economics* 3.1 (1997): 1-51.

Aguña, Clara García, and Milorad Kovacevic. (2010) "Uncertainty and Sensitivity Analysis of the Human Development Index." *Human Development Research Paper* 47 (2010): 2010.

Alexandrova, Anna. (2008) "First-person Reports and the Measurement of Happiness." *Philosophical Psychology* 21.5 (2008): 571-583.

Alexandrova, Anna. (2012) "Well-Being as an Object of Science." *Philosophy of Science* 79, no. 5: 678-89.

Alexandrova, Anna. (2015) "Well-being and Philosophy of Science." *Philosophy Compass* 10.3 (2015): 219-231.

Alexandrova, Anna. (2017) *A Philosophy for the Science of Well-Being.* Oxford University Press, New York, NY.

Alexandrova, Anna. (2018) "Can the Science of Well-being be Objective?." *The British Journal for the Philosophy of Science* 69.2 (2018): 421-445.

Alkire, Sabina. (2002) "Dimensions of Human Development," World Development, Elsevier, vol. 30(b) (2002): 181-205.

Alkire, Sabina. (2005) "Why the Capability Approach?." *Journal of Human Development* 6.1 (2005): 115-135.

Alkire, Sabina. (2006) "Public Debate and Value Construction in Sen's Approach." In Kauffman, Alexander ed, *Capabilities Equality*. Routledge, 2006. 143-164.

Alkire, Sabina. (2008a) "Choosing Dimensions: The Capability Approach and Multidimensional Poverty," in Kakwani, Nanak and Jaques Silber eds, *The Many Dimensions of Poverty*, p 89-119. Basingstoke: Palgrave-MacMillan, 2008.

Alkire, Sabina. (2008b) "Using the Capability Approach: Prospective and Evaluative Analyses." *The Capability Approach: Concepts, Measures and Applications, Cambridge University Press, Cambridge* (2008): 26-50.

Alkire, Sabina. (2010) "Human Development: Definitions, Critiques, and Related Concepts." *United Nations Development Programme Human Development Reports Research Paper* (June 2010): 2010/01.

Alkire, Sabina. (2011) "Multidimensional Poverty and Its Discontents." (2011). *OPHI Working Paper No. 64* (November 2011), Oxford University.

Alkire, Sabina. (2015) "The Capability Approach and Well-being Measurement for Public Policy." *OPHI Working Paper No. 94* (March 2015), Oxford University.

Alkire, Sabina, and James Foster. (2009) "Counting and Multidimensional Poverty Measurement." *Oxford Poverty & Human Development Initiative (OPHI) Working Paper 32*, Oxford University.

Alkire, Sabina, and James Foster. (2010) "Designing the Inequality-Adjusted Human Development Index." *United Nations Development Programme Human Development Reports Research Paper* (October 2010): 2010/28.

Alkire, Sabina, and James Foster. (2011) "Understandings and Misunderstandings of Multidimensional Poverty Measurement." *OPHI Working Papers* 43, (May 2011): Oxford University.

Alkire, Sabina, and Selim Jahan. (2018) "The New Global MPI 2018: Aligning With the Sustainable Development Goals" *HDRO Occasional Paper, UNDP*

Alkire, Sabina, Usha Kanagaratnam, Ricardo Nogales, and Nicolai Suppa. (2019a) "The First Revision of the Global MPI: Empirical Insights and Robustness." *Paper Prepared for the IARIW-World Bank Conference, Washington DC, November 7-8, 2019.*

Alkire, Sabina, Usha Kanagaratnam, Ricardo Nogales, and Nicolai Suppa. (2019b) "Sensitivity Analysis in Poverty Measurement: The Case of the Global Multidimensional Poverty Index." *Paper Prepared for the IARIW-World Bank Conference, Washington DC, November 7-8, 2019.*

Alkire, Sabina, James Foster, Suman Seth, Maria Emma Santos, Jose Roche, and Paola Ballon. (2015a) "Multidimensional Poverty Measurement and Analysis: Chapter 2 – The Framework." *OPHI Working Paper No. 83* (January 2015), Oxford University.

Alkire, Sabina, James Foster, Suman Seth, Maria Emma Santos, Jose Roche, and Paola Ballon. (2015b) "Multidimensional Poverty Measurement and Analysis: Chapter 3 – Overview of Methods for Multidimensional Poverty Assessment." *OPHI Working Paper No. 84* (January 2015), Oxford University.

Alkire, Sabina, James Foster, Suman Seth, Maria Emma Santos, Jose Roche, and Paola Ballon. (2015c) "Multidimensional Poverty Measurement and Analysis: Chapter 5 – The Alkire-Foster Counting Methodology." *OPHI Working Paper No. 86* (January 2015), Oxford University.

Alkire, Sabina, and Maria Emma Santos. (2010) "Acute Multidimensional Poverty: A New Index for Developing Countries." *United Nations Development Programme Human Development Report Office Background Paper* 2010/11 (2010).

Alkire, Sabina, and Maria Emma Santos. (2013) "A Multidimensional Approach: Poverty Measurement & Beyond." *Social Indicators Research* 112.2 (2013): 239-257.

Alkire, Sabina, and Maria Emma Santos. "Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index." *World Development* 59 (2014): 251-274.

Alkire, Sabina, Maria Emma Santos, Suman Seth, and Gaston Yalonetzky. (2010) "Is the Multidimensional Poverty Index Robust to Different Weights?" *Oxford Poverty and Human Development Initiative* (November 2010).

Alkire, Sabina, and Moizza Binat Sarwar. (2009) "Multidimensional Measures of Poverty & Well-Being." *OPHI Research in Progress* (2009): RP1a.

Alkire, Sabina, and Vollmer, Frank. (2018) "Towards a Global Assets Indicator: Re-assessing the Assets Indicator in the Global Multidimensional Poverty Index." *OPHI Research in Progress Series 53a* (September 2018).

Anagnostopoulos, Georgios, ed. (2013) *A Companion to Aristotle*. John Wiley & Sons, 2013.

Anand, Paul, Graham Hunter, Ian Carter, Keith Dowding, Francesco Guala, and Martin Van Hees. (2009) "The Development of Capability Indicators." *Journal of Human Development and Capabilities* 10, no. 1 (2009): 125-152.

Anand, Paul, Graham Hunter, and Ron Smith. (2005) "Capabilities and Well-being: Evidence Based on the Sen–Nussbaum Approach to Welfare." *Social Indicators Research* 74.1 (2005): 9-55.

Anand, Paul, Cristina Santos, and Ron Smith. (2007) *The Measurement of Capabilities*. No. 67. Open Discussion Papers in Economics, 2007.

Anand, Sudhir, and Amartya Sen. (1994) "Human Development Index: Methodology and Measurement." *Human Development Report Office Occasional Papers* (1994).

Anand, Sudhir, and Amartya Sen. (1995) "Gender Inequality in Human Development: Theories and Measurement." (1995): New York, UNDP.

Anand, Sudhir, and Amartya Sen. (1997) "Concepts of Human Development and Poverty: A Multidimensional Perspective." *United Nations Development Programme, Poverty and Human Development: Human Development Papers* (1997): 1-20.

Anand, Sudhir, and Amartya Sen. (1998) "Consumption & Human Development." *Human Development Index* (1998).

Anderson, Elizabeth (1993), *Value in Ethics and Economics*. (1993) London: Harvard University Press.

Anderson, Elizabeth. (1997), "Practical Reason and Incommensurable Goods," in *Incommensurability, Incomparability, and Practical Reason*, Ruth Chang (ed.), 1997: Cambridge: Harvard University Press.

Anderson, Elizabeth, and Richard H. Pildes. (2000) "Expressive theories of law: A General Restatement." *University of Pennsylvania Law Review* 148.5 (2000): 1503-1575.

Angner, Erik. (2015) "Well-being and Economics." In Guy Fletcher, Ed., *The Routledge Handbook of the Philosophy of Well-Being (London: Routledge, 2015)* (2015): 15-1.

Aristotle. *Nicomachean Ethics*, Translated by David Ross. Oxford: Oxford World Classics, 1980.

Aristotle. *Politics*, Translated by C.D.C. Reeve. Indianapolis: Hacket Publishing Company, 1998.

Aristotle. *Politics, Books i., iii., iv.(vii.).* The text of Bekker, Translated by WE Bolland, together with Introductory Essays by A. Lang. 1877.

Arneson, Richard J. (1999) "Human Flourishing versus Desire Satisfaction." *Social Philosophy and Policy* 16.1 (1999): 113-142.

Arneson, Richard J. (2000) "Perfectionism and Politics." *Ethics* 111.1 (2000): 37-63.

Arneson, Richard J. (2010a) "Book Review: Disadvantage, Capability, Commensurability, and Policy." *Politics, Philosophy & Economics* 9.3 (2010): 339-357.

Arneson, Richard. (2010b) "Two Cheers for Capabilities." *Measuring justice: Primary Goods and Capabilities* 1 (2010): 101-128.

Arrow, Kenneth J. (1977) "Extended Sympathy and the Possibility of Social Choice." *The American Economic Review* 67.1 (1977): 219-225.

Atkinson, Anthony B. "On the Measurement of Poverty." *Econometrica: Journal of the Econometric Society* (1987): 749-764.

Atkinson, Anthony B. (2003) "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches." *The Journal of Economic Inequality* 1.1 (2003): 51-65.

Barrington-Leigh, Christopher, and Alice Escande. (2018) "Measuring Progress and Well-being: A Comparative Review of Indicators." *Social Indicators Research* 135.3 (2018): 893-925.

Bartholomew, David J. (1980) "Factor Analysis for Categorical Data." *Journal of the Royal Statistical Society: Series B (Methodological)* 42.3 (1980): 293-312.

Begon, Jessica. (2015) "What are Adaptive Preferences? Exclusion and Disability in the Capability Approach." *Journal of Applied Philosophy* 32.3 (2015): 241-257.

Berdinesen, Hein. (2018) "Future Generationas and Concepts of Well-being." *Filozofia* 73.4 (2018).

Biggeri, Mario, Renato Libanora, Stefano Mariani, and Leonardo Menchini. (2006) "Children Conceptualizing their Capabilities: Results of a Survey Conducted during the First Children's World Congress on Child Labour." *Journal of Human Development* 7, no. 1 (2006): 59-83.

Binmore, Ken. (2007) "Interpersonal Comparison of Utility." (2007).

Blair, Douglas H. (1988) "The Primary-Goods Indexation Problem in Rawls's Theory of Justice." *Theory and Decision* 24.3 (1988): 239-252.

Bogen, James, and James Woodward. (1988) "Saving the Phenomena." *The Philosophical Review* 97.3 (1988): 303-352.

Borsboom, Denny, Gideon J. Mellenbergh, and Jaap Van Heerden. (2004) "The Concept of Validity." *Psychological Review* 111.4 (2004): 1061.

Boumans, Marcel. (1999) "Representation and Stability in Testing and Measuring Rational Expectations." *Journal of Economic Methodology* 6.3 (1999): 381-402.

Boumans, Marcel. (2009) "Modeling Strategies for Measuring Phenomena in-and Outside the Laboratory." *EPSA Philosophy of Science: Amsterdam 2009*. Springer, Dordrecht, 2012. 1-11.

Brandolini, Andrea. (2008) "On Applying Synthetic Indices of Multidimensional Well-being: Health and Income Inequalities in Selected EU Countries." *Bank of Italy Temi di Discussione (Working Paper) No* 668 (2008).

Bridgman, Percy Williams. (1927) *The Logic of Modern Physics*. New York: Macmillan, 1927.

Broome, John. (1978) "Trying to Value a Life." *Journal of Public Economics* 9.1 (1978): 91-100.

Broome, John. (2000) "Incommensurable Values." In Roger Crisp and Brad Hooker Eds. *Well-Being and Morality: Essays in Honour of James Griffin* (2000): Oxford University Press, Oxford.

Byskov, Morten. (2016) *Between Experts and Locals: Towards an Inclusive Framework for a Development Agenda*. Diss. Utrecht University, 2016.

Byskov, Morten Fibieger. (2017) "Democracy, Philosophy, and the Selection of Capabilities." *Journal of Human Development and capabilities* 18.1 (2017): 1-16.

Fibieger Byskov, Morten. (2019) "Beyond 'Having Reason to Value': Why We Should Adopt a Procedure-Independent and Value-Neutral Definition of Capabilities." *Journal of Economic Methodology* (2019): 1-18.

Cartwright, Nancy. (1994) *Nature's Capacities and their Measurement*. OUP Catalogue (1994).

Cartwright, Nancy (1999) *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press, 1999.

Cartwright, Nancy, and Norman Bradburn. "A Theory of Measurement." *The Importance of Common Metrics for Advancing Social Science Theory and Research: Proceedings of the National Research Council Committee on Common Metrics* (2011): 53-70.

Cartwright, Nancy, Norman Bradburn, and Jonathan Fuller (2016) 'A Theory of Measurement.', *Working Paper. Centre for Humanities Engaging Science and Society* (CHESS), 2016, Durham.

Cartwright, Nancy, Jordi Cat, Lola Fleck, and Thomas E Uebel, (1996) *Otto Neurath: Philosophy between Science and Politics*, New York: Cambridge University Press, Ideas in Context Series No. 38.

Cartwright, Nancy, Stathis Psillos and Hasok Chang. (2008). "Measurement", in Stathis Psillos, and M. Curd, Eds. *The Routledge Companion to Philosophy of Science*. (2008) New York: Routledge.

Cartwright, Nancy, and Rosa Runhardt. (2014) "Measurement". In: Cartwright, Nancy &Montuschi, Eleonora, Eds. *Philosophy of Social Science: A New Introduction*. Oxford: Oxford University Press 2014.

Chang, Hasok (2004). *Inventing Temperature*. New York: Oxford University Press, 2004.

Chang, Hasok. (2009) "Operationalism." In *Stanford Encyclopedia of Philosophy*: ttps://plato.stanford.edu/archives/win2019/entries/operationalism/ (2009; accessed 2017).

Chang, Hasok (2012). *Is Water H_20? Evidence, Realism, and Pluralism.* Springer Science & Business Media, New York, NY.

Chang, Hasok. (2016) "Pragmatic Realism." *Humanities Journal of Valparaiso* 8 (2016): 107-122.

Chang, Hasok (2017) "Pragmatist Coherence as the Source of Truth and Reality." Proceedings of the Aristotelian Society. Vol. 117. No. 2.

Chang, Ruth, (1997) 'Introduction' in Ruth Chang ed. *Incommensurability, Incomparability, and Practical Reason*. Harvard University Press, 1997.

Chang, Ruth. (2015) "Value Incomparability and Incommensurability." *The Oxford Handbook of Value Theory* (2015): 205-224.

Chang, Ruth. (2017) "Hard Choices." *Journal of the American Philosophical Association* 3.1 (2017): 1-21.

Chowdhury, Shyamal, and Lyn Squire. (2006) "Setting Weights for Aggregate Indices: An Application to the Commitment to Development Index and Human Development Index." *The Journal of Development Studies* 42.5 (2006): 761-771.

Claassen, Rutger. (2011) "Making Capability Lists: Philosophy versus Democracy." *Political Studies* 59.3 (2011): 491-508.

Claassen, Rutger. (2017) "An Agency-Based Capability Theory of Justice." *European Journal of Philosophy* 25.4 (2017): 1279-1304.

Claassen, Rutger. (2018) *Capabilities in a Just Society: A Theory of Navigational Agency*. Cambridge University Press, 2018.

Clark, David Alexander, Mario Biggeri, and Alexandre Apsan Frediani, eds. (2019) *The Capability Approach, Empowerment and Participation: Concepts, Methods and Applications*. Springer, 2019.

Comim, Flavio. (2001) "Operationalizing Sen's Capability Approach." *Justice and Poverty: Examining Sen's Capability Approach Conference*. 2001.

Comim, Flavio. (2008) "Measuring Capabilities" in Mozaffar Qizilbash, and Sabina Alkire, eds. (2008) *The Capability Approach: Concepts, Measures and Applications*. Cambridge University Press.

Comim, Flavio, Shailaja Fennell, and Prathivadi Bhayankaram Anand, Eds. (2018) *New Frontiers of the Capability Approach*. Cambridge University Press, 2018.

Comim, Flavio, Mozaffar Qizilbash, and Sabina Alkire, eds. (2008) *The Capability Approach: Concepts, Measures and Applications*. Cambridge University Press.

Copp, David. (1990) "Explanation and Justification in Ethics." *Ethics* 100.2 (1990): 237-258.

Copp, David. (2015) "Explaining Normativity." *Proceedings and Addresses of the American Philosophical Association*. Vol. 89. American Philosophical Association, 2015.

Crasnow, Sharon. (2015) "The Measure of Democracy: Coding in Political Science." *Standardization in Measurement: Philosophical, Historical and Sociological Issues* (2015).

Crespo, Ricardo F. (2008) *On Sen and Aristotle*. IAE Working Paper Series, DT 03/08. Retrieved October 18, 2008, from http://www. iae. edu. ar/pi/Documentos% 20Investigacin/Working% 20Papers/DTIAE03_2008. pdf, 2008.

Crisp, Roger. "Well-being." (2001) From *Stanford Encyclopedia of Philosophy* Accessed 2017: https://plato.stanford.edu/archives/fall2017/entries/well-being/, 2017.

Crocker, David A. (2008) *Ethics of Global Development: Agency, Capability, and Deliberative Democracy*. Cambridge University Press, 2008.

Dasgupta, Partha, and Martin Weale.)(1992) "On Measuring the Quality of Life." World Development 20.1: 119-131.

Decancq, Koen. (2017) "Measuring Multidimensional Inequality in the OECD Member Countries with a Distribution-Sensitive Better Life Index." *Social Indicators Research* 131.3 (2017): 1057-1086.

Decancq, Koen, and María Ana Lugo. (2013) "Weights in Multidimensional Indices of Well-being: An Overview." *Econometric Reviews* 32.1 (2013): 7-34.

Decancq, Koen, and Erik Schokkaert. (2013) "Beyond GDP: Measuring social progress in Europe." *Euroforum Policy Paper 4* (2013): 1-46.

Deutsch, Joseph, and Jacques Silber. (2005) "Measuring Multidimensional Poverty: An Empirical Comparison of Various Approaches." *Review of Income and Wealth* 51.1 (2005): 145-174.

de la Torre, Rodolfo (2018) 'Justice' ' in *Discussion Paper III: Possible Measurement of the Common Good's Dynamic, UPAEP- IPBC*

Deneulin, Séverine. (2011) "Recovering Nussbaum's Aristotelian Roots." in *Revista Cultura Economica No. 81/82* (December 2011: 31-37).

Despotis, D. K. (2005) "A Reassessment of the Human Development Index via Data Envelopment Analysis." *Journal of the operational research society* 56.8 (2005): 969-980.

DiStefano, Christine, Min Zhu, and Diana Mindrila. (2009) "Understanding and Using Factor Scores: Considerations for the Applied Researcher." *Practical Assessment, Research, and Evaluation* 14.1 (2009): 20.

Drost, Ellen A. (2011) "Validity and Reliability in Social Science Research." *Education Research and Perspectives* 38.1 (2011): 105.

Duclos, Jean-Yves, and Luca Tiberti. (2016) "Multidimensional Poverty indices: A Critical Assessment." in Adler, Matthew D., and Marc Fleurbaey, Eds. *The Oxford Handbook of Well-being and Public Policy*. Oxford University Press, (2016).

Efstathiou, Sophia. "How Ordinary Race Concepts Get to Be Usable in Biomedical Science: An Account of Founded Race Concepts." *Philosophy of Science* 79, no. 5 (2012): 701-13.

Efstathiou, Sophia (2016). Is it Possible to Give Scientific Solutions to Grand Challenges? On the Idea of Grand Challenges for Life Science Research. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 56: 46-61.

Elffers, Henk. (1980) "On Uninterpretability of Factor Analysis Results." *Transactions of the Institute of British Geographers* (1980): 318-329.

Elson, Diane. (2016) *Value: The Representation of Labour in Capitalism*. Verso Books, 2016.

Finnis, John. (2011) *Natural Law and Natural Rights*. Oxford University Press, 2011.

Fleurbaey, Marc. (2006) "Capabilities, Functionings and Refined Functionings." *Journal of Human Development* 7.3 (2006): 299-310.

Frede, Michael. (1987) *Essays in Ancient Philosophy*. U of Minnesota Press, 1987.

Frediani, Alexandre Apsan. (2010) "Sen's Capability Approach as a Framework to the Practice of Development." *Development in practice* 20.2 (2010): 173-187.

Fricker, Miranda. (2007) *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.

Friedman, Milton. (1966) "The Methodology of Positive Economics." In *Essays in Positive Economics* Chicago: University of Chicago Press, 1966.

Frigerio, Aldo, Alessandro Giordani, and Luca Mari. (2010) "Outline of a General Model of Measurement." *Synthese* 175.2 (2010): 123-149.

Fukuda-Parr, Sakiko. (2003) "The Duman Development Paradigm: Operationalizing Sen's Ideas on Capabilities." *Feminist Economics* 9.2-3 (2003): 301-317.

Gaertner, Wulf. (2012) "Evaluating Sets of Objects in Characteristics Space." *Social Choice and Welfare* 39.2-3 (2012): 303-321.

Gaertner, Wulf and Yongsheng Xu. (2005) : A New Measure of the Standard of Living Based on Functionings, *CeGE Discussion Paper*, No. 49.

Gaertner, Wulf, and Yongsheng Xu. (2006) "Capability Sets as the Basis of a New Measure of Human Development." *Journal of Human Development* 7.3 (2006): 311-321.

Mónika, Galambosné Tiszberger. (2018) "The Weight of Weighting – an Empirical Study Based on the OECD Better Life Index." *The Business & Management Review* 9.3 (2018): 443-450.

Gasper, Des (1996) Needs and Basic Needs – A Clarification of Foundational Concepts for Development

Ethics and Policy. *ISS Working Paper* 210. The Hague.

Gasper, Des. (2007) "Human Well-being: Concepts and Conceptualizations." *Human Well-Being*. Palgrave Macmillan UK, 2007. 23-64.

Glass, Gene V., and Peter A. Taylor. (1966) "Chapter VI: Factor Analytic Methodology." *Review of Educational Research* 36.5 (1966): 566-587.

Glatzer, Wolfgang, Laura Camfield, Valerie Moller, and Mariano Rojas. (2015) "Global Handbook of Quality of Life." *Exploration of Well-being of Nations and Continents*. Germany: Springer, 2015.

Gough, Ian. (2014) "Lists and Thresholds: Comparing the Doyal-Gough Theory of Human Need with Nussbaum's Capabilities Approach." *Capabilities, Gender, Equality* (2014): 357-82.

Gough, Ian. (2015) "Climate Change and Sustainable Welfare: the Centrality of Human Needs." *Cambridge Journal of Economics* 39.5 (2015): 1191-1214.

Gould, Stephen Jay, and Steven James Gold. (1996) *The Mismeasure of Man*. WW Norton & Company, 1996.

Gray, John. (2000) *Two Faces of Liberalism*. New York: The New Press, 2000.

Griffin, J. (1986), *Well-Being: Its Meaning, Measurement and Importance*, Oxford: Clarendon Press.

Griffin, J. (1991), "Mixing Values", *The Proceedings of the Aristotelian Society*, Supplementary Volume, 65: 101-118.

Griffin, J. (1997), "Incommensurability: What's the Problem?" in *Incommensurability, Incomparability, and Practical Reason*, R. Chang (ed.), Cambridge: Harvard University Press.

Habermas, Jürgen. "Dogmatism, Reason, and Decision: On theory and Praxis in our Scientific Civilization." *Theory and Practice* (1973): 253-282.

Hacking, Ian, and Jan Hacking. (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, 1983.

Hampshire, Stuart. (1977) "On Having a Reason." *Royal Institute of Philosophy Supplements* 11 (1977): 86-98.

Hampshire, Stuart (1982) "Morality and Convention" in Amartya Sen and Bernard Williams Eds., *Utilitarianism and Beyond* New York: Cambridge University Press, 1982.

Harsanyi, John C. (1996) "Utilities, Preferences, and Substantive Goods." *Social Choice and Welfare* 14.1 (1996): 129-145.

Hausman, Daniel M., and Michael McPherson. (2004) "The Philosophical Foundations of Normative Economics." *Development Dilemmas* (2004): 35.

Haybron, Daniel M., and Valerie Tiberius. (2015) "Well-being Policy: What Standard of Well-being?." *Journal of the American Philosophical Association* 1.4 (2015): 712-733.

Heathwood, Chris. (2016) "Desire-Fulfillment Theory." *The Routledge Handbook of Philosophy of Well-Being. New York: Routledge* (2016): 135-147.

Hersch, Gil. *Measuring Well-Being for Public Policy; Doing without Theory*. Diss. UC San Diego, 2016.

Hidalgo, Cesar A. (2010) *Graphical Statistical Methods for the Representation of the Human Development Index and its Components*. UNDP, 2010.

Hsieh, Nien-hê. (2005), "Equality, Clumpiness, and Incomparability," *Utilitas*, 17: 180-204.

Hsieh, Nien-hê. (2007a), "Is Incomparability a Problem for Anyone?" *Economics and Philosophy*, 23: 65-80.

Hsieh, Nien-hê. (2007b) "Incommensurable Values." in *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/archives/spr2016/entries/value-incommensurable/ (Accessed 2017).

Johnson, Monte Ransome. (2015) "Aristotle's Architectonic Sciences." In David Ebrey Ed., *Theory and Practice in Aristotle's Natural Science*, Cambridge: Cambridge University Press, 2015.

Joint Research Centre-European Commission. (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, 2008.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. (1997) "Back to Bentham? Explorations of Experienced Utility." *The Quarterly Journal of Economics* 112.2 (1997): 375-406.

Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur Stone. (2004) "Toward National Well-being Accounts." *American Economic Review* 94, no. 2 (2004): 429-434.

Kellert, Stephen H., Helen E. Longino, and C. Kenneth Waters. (2006) "Introduction: The Pluralist Stance." in Kenneth Waters, Herbert Feigl, Stephen Kellert, Helen Longino Eds., *Scientific Pluralism*. University of Minnesota Press: 2006.

Kelley, Allen C. (1991) "The Human Development Index: Handle with Care." *Population and Development Review* (1991): 315-324.

Keyser, Vadim. (2017) "Experimental Effects and Causal Representations." *Synthese* (2017): 1-32.

Khader, Serene J. (2011) *Adaptive Preferences and Womens Empowerment*. Oxford University Press, 2011.

Kleist, Chad, "Developing Capabilities: A Feminist Discourse Ethics Approach" (2017). *Dissertations (2009)*. 743. http://epublications.marquette.edu/dissertations_mu/743

Konyndyk, Jeremy, and Rose Worden. (2019) "People-Driven Response: Power and Participation in Humanitarian Action." *CGD Policy Paper* 155 (2019).

Krishnakumar, Jaya. (2007) Going Beyond Functionings to Capabilities: An Econometric Model to Explain and Estimate Capabilities, Journal of Human Development, 8:1,39-63, DOI: 10.1080/14649880601101408.

Krishnakumar, Jaya, and Anirudh Lal Nagar. (2008) "On Exact Statistical Properties of Multidimensional Indices Based on Principal Components, Factor Analysis, MIMIC and Structural Equation Models." *Social Indicators Research* 86.3 (2008): 481-496.

Krishnakumar, Jaya, and Florian Chávez-Juárez. (2014) "Estimating Capabilities with Structural Equation Models: How Well are we Doing in a 'Real' World?." *Social Indicators Research* 129.2 (2014): 717-737.

Krishnakumar, Jaya, and Ricardo Nogales. (2015) "Public Policies for Wellbeing with Justice: A Theoretical Discussion Based on Capabilities and Opportunities." *International Journal of Wellbeing* 5.3 (2015).

Kuklys, Wiebke. (2005). *Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications*. Heidelberg: Springer Science & Business Media, 2005.

Lee, Caroline W. (2007) "Is There a Place for Private Conversation in Public Dialogue? Comparing Stakeholder Assessments of Informal Communication in Collaborative Regional Planning." *American Journal of Sociology* 113.1 (2007): 41-96.

List, Christian. "Social Choice Theory." (2013). in *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/archives/win2013/entries/social-choice/ (Accessed 2021).

Lind, Niels. (2004) "Values Reflected in the Human Development Index." *Social Indicators Research* 66.3 (2004): 283-293.

Lopes, Hedibert Freitas, and Mike West. (2004) "Bayesian Model Assessment in Factor Analysis." *Statistica Sinica* (2004): 41-67.

Lorenz, Jan, Christoph Brauer, and Dirk Lorenz. (2017) "Rank-Optimal Weighting or "How to be Best in the OECD Better Life Index?"." *Social Indicators Research* 134.1 (2017): 75-92.

Luce, Robert Duncan, and Patrick Suppes. (1965) "Preference, Utility, and Subjective Probability." ', in R. Luce, R. Bush, and E. Galanter, eds., *Handbook of Mathematical Psychology*, John Wiley and Sons, New York (1965), pp. 249–410.

Luce, R. Duncan, and Patrick Suppes. (2002) "Representational Measurement Theory." In Hal Pashler and John Wixted Eds., *Stevens' Handbook of Experimental Psychology Volume 4: Methodology in Experimental Psychology* (2002).

Maas, Harro. (2001) "An Instrument Can Make a Science: Jevons's Balancing Acts in Economics." *History of Political Economy* 33.5 (2001): 277-302.

Malik, Khalid. (2013) "Human development report 2013. The Rise of the South: Human Progress in a Diverse World." *The Rise of the South: Human Progress in a Diverse World (March 15, 2013). UNDP-HDRO Human Development Reports* (2013).

Mandler, Michael. (2006) "Cardinality versus Ordinality: A Suggested Compromise." *American Economic Review* 96.4 (2006): 1114-1136.

Mari, Luca. (2005) "The Problem of Foundations of Measurement." *Measurement* 38.4 (2005): 259-266.

Martinetti, Enrica Chiappero. (1994) "A New Approach to Evaluation of Well-being and Poverty by Fuzzy Set Theory." *Giornale degli Economisti e Annali di Economia* (1994): 367-388.

Martinetti, Enrica Chiappero. (1996) "Standard of Living Evaluation Based on Sen's Approach: Some Methodological Suggestions." (1996): 37-53.

Martinetti, Enrica Chiappero. (2000) "A Multidimensional Assessment of Well-being Based on Sen's Functioning Approach." *Rivista Internazionale di Scienze Sociali* (2000): 207-239.

Martinetti, Enrica Chiappero. (2006a) "Capability Approach and Fuzzy Set Theory: Description, Aggregation and Inference Issues." *Fuzzy Set Approach to Multidimensional Poverty Measurement*. Springer, Boston, MA, 2006. 93-113.

Martinetti, Enrica Chiappero. (2006b) "Complexity and Vagueness in the Capability Approach: Strengths or Weaknesses." *University of Pavia, Italy,(Mimeo)* 3 (2006): 1-32.

Martinetti, Enrica Chiappero, and José Manuel Roche. (2009) "Operationalization of the Capability Approach, from Theory to Practice: a Review of Techniques and Empirical Applications." *Debating global society: Reach and Limits of the Capability Approach* (2009): 157-203.

McGillivray, Mark, and Matthew Clarke. (2006) "Human Well-being: Concepts and Measures." *Understanding Human Well-being* (2006): 3-16.

McGregor, Allister, Sarah Coulthard, and Laura Camfield. (2015) "Measuring what Matters: The Role of Well-being Methods in Development Policy and Practice." *Development Progress for ODI* (2015).

Meacham, Christopher JG, and Jonathan Weisberg. (2011) "Representation Theorems and the Foundations of Decision Theory." *Australasian Journal of Philosophy* 89.4 (2011): 641-663.

Meikle, Scott. (1995). *Aristotle's Economic Thought*. New York: Oxford University Press.

Monmonier, Mark. (2010) *Rhumb Lines and Map Wars: A Social History of the Mercator Projection*. University of Chicago Press, 2010.

Morgan, Mary S., and Maria Bach. (2018) "Measuring Development—from the UN's Perspective." *History of Political Economy* 50.S1 (2018): 193-210.

National Research Council. (2002) *At What Price?: Conceptualizing and Measuring Cost-of-Living and Price Indexes*. National Academies Press, 2002.

Nebel, Mathias, and Delgadillo, Jorge (2018): *Discussion Paper II: Measuring the Nexus of the Common Good* (forthcoming)

Nelson, Eric. (2008) "From Primary Goods to Capabilities: Distributive Justice and the Problem of Neutrality." *Political Theory* 36.1 (2008): 93-122.

Neurath, Otto. Robert Cohen and Marie Neurath, Eds. *Philosophical Papers 1913–1946: With a Bibliography of Neurath in English*. Vol. 16. D. Reidel Publishing Company, Dordrecht, Holland, 1983.

Nguefack-Tsague, Georges, Stephan Klasen, and Walter Zucchini. (2011) "On Weighting the Components of the Human Development Index: a Statistical Justification." *Journal of Human Development and Capabilities* 12.2 (2011): 183-202.

Noble, Michael, Gemma Wright, George Smith, and Chris Dibben. (2006) "Measuring Multiple Deprivation at the Small-Area Level." *Environment and Planning A* 38, no. 1 (2006): 169-185.

Nussbaum, Martha C. (1986a) *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge University Press, 2001.

Nussbaum, Martha C. (1986b) "The Discernment of Perception: An Aristotelian Conception of Public and Private Rationality." *Proceedings of the Boston Artea colloqium in Ancient Philosophy* 1. University Press of America: Lanham (1986).

Nussbaum, Martha C. (1987a) "Nature, Function and Capability: Aristotle on Political Distribution" in *Working Papers*, 1987/31.

Nussbaum, Martha C. (1987b) "Non-Relative Virtues: An Aristotelian Approach" in *Working Papers*, 1987/32.

Nussbaum, Martha C. (1992) "Human Functioning and Social Justice: In Defense of Aristotelian Essentialism" in *Political Theory*, (1992) 20:2, 202-246.

Nussbaum, Martha. (1990) "Aristotelian Social Democracy." in R. Bruce Douglass, Gerald R. Mara, and Henry S. Richardson, Eds. *Liberalism and the Good*. Routledge, 1990: 203-252.

Nussbaum, Martha C. (1993) "Reply to Papers" in *Philosophical Investigations* (1993) 16:1, 46-86.

Nussbaum, Martha C. (1995a) "Aristotle on Human Nature and the Foundations of Ethics." in J.E.J. Altham and Ross Harrison, Eds. *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (1995).

Nussbaum, Martha C. (1995b) "Emotions and Women's Capabilities." in Martha Nussbaum and Jonathan Glover Eds., *Women, Culture and Development: A Study of Human Capabilities* (1995): 360-395.

Nussbaum, Martha. (1998) "Political Animals: Luck, Love and Dignity." *Metaphilosophy* 29.4 (1998): 273-287.

Nussbaum, Martha C. (2000a) *Women and Human Development: The Capabilities Approach*. Cambridge University Press, 2000.

Nussbaum, Martha C. (2000b) "Aristotle, Politics, and Human Capabilities: A Response to Antony, Arneson, Charlesworth, and Mulgan" in *Ethics* (2000) 111:1, 102-140.

Nussbaum, Martha C. (2000c) "The Costs of Tragedy: Some Moral Limits of Cost-Benefit Analysis." *The Journal of Legal Studies* 29.S2 (2000): 1005-1036.

Nussbaum, Martha C. (2001) "Symposium on Amartya Sen's Philosophy: 5 Adaptive Preferences and Women's Options." *Economics & Philosophy* 17.1 (2001): 67-88.

Nussbaum, Martha C. (2003) "Capabilities as Fundamental Entitlements: Sen and Social Justice." *Feminist Economics* 9.2-3 (2003): 33-59.

Nussbaum, Martha C. (2008) "Who is the Happy Warrior? Philosophy Poses Questions to Psychology." *The Journal of Legal Studies* 37.S2 (2008): S81-S113.

Nussbaum, Martha C. (2009) "Creating Capabilities: The Human Development Approach and its Implementation." *Hypatia* 24.3 (2009): 211-215.

Nussbaum, Martha C. (2011) *Creating Capabilities*. Cambridge: Harvard University Press, 2011.

Nussbaum, Martha, and Rosalind Hursthouse. (1984) "Plato on Commensurability and Desire." *Proceedings of the Aristotelian Society, Supplementary Volumes* 58 (1984): 55-96.

Nussbaum, Martha, and Amartya Sen. (1987) *Internal Criticism and Indian Rationalist Traditions*. No. 30. World Institute for Development Economics Research of the United Nations, 1987.

Nussbaum, Martha, and Amartya Sen. (1993) *The Quality of Life*. Clarendon Press: Oxford.

Ober, Josiah. (2012) "Relevant Expertise Aggregation: An Aristotelian Middle Way for Epistemic Democracy." (2012).

Ober, Josiah. (2013) "Democracy's Wisdom: An Aristotelian Middle Way for Collective Judgment." *American Political Science Review* 107.1 (2013): 104-122.

Organisation for Economic Co-operation and Development (OECD) Better Life Initiative. (2011) "Compendium of OECD Well-being Indicators." *Organization for Economic Cooperation and Development (OECD), Paris. http://www.oecd.org/sdd/47917288.pdf* (2011).

Organisation for Economic Co-operation and Development (OECD). (2011) *How's Life?: Measuring Well-being*. Paris: OECD, 2011.

Page, Scott E. (2007) "Making the Difference: Applying a Logic of Diversity." *Academy of Management Perspectives* 21.4 (2007): 6-20.

Pattanaik, Prasanta K., and Yongsheng Xu. (1990) "On Ranking Opportunity Sets in Terms of Freedom of Choice." *Recherches Économiques de Louvain/Louvain Economic Review* 56.3-4 (1990): 383-390.

Pattanaik, Prasanta K., and Yongsheng Xu. (2013) “Freedom and Its Value” in Andrew Young School of Policy Studies Research Paper Series WP 13-12.

Pearce, David, Giles Atkinson, and Kirk Hamilton. (1998) "The Measurement of Sustainable Development." in Jeroen C.J.M., van Den Bergh and Marjan W. Hofkes, Eds. *Theory and Implementation of Economic Models for Sustainable Development*. Springer, Dordrecht, 1998. 175-193.

Peleg, Noam. (2013) "Reconceptualising the Child’sRright to Development: Children and the Capability Approach." *The International Journal of Children's Rights* 21.3 (2013): 523-542.

Permanyer, Iñaki. (2012) "Uncertainty and Robustness in Composite Indices Rankings." *Oxford Economic Papers* 64.1 (2012): 57-79.

Peruzzi, Agnese. (2014) "Understanding Social Exclusion from a Longitudinal Perspective: A Capability-Based Approach." *Journal of Human Development and Capabilities* 15.4 (2014): 335-354.

Peruzzi, Agnese. (2015) "From Childhood Deprivation to Adult Social Exclusion. Evidence from the (1970) British Cohort Study." *Social Indicators Research,* 120(a): 117–135.

Putnam, Hilary. (2002) *The Collapse of the Fact/Value Dichotomy and Other Essays*. Harvard University Press, 2002.

Putnam, Hilary (2017). "The Fact/Value Dichotomy and the Future of Philosophy," in Giancarlo Marchetti and Sarin Marchetti, eds. *Facts and Values: The Ethics and Metaphysics of Normativity*. Routledge: NY, 2017.

Qizilbash, Mozaffar. (1997) "Pluralism and Well-being Indices." *World Development* 25.12: 2009-2026.

Qizilbash, Mozaffar. (2002) "A Note on the Measurement of Poverty and Vulnerability in the South African Context." *Journal of International Development* 14.6: 757-772.

Qizilbash, Mozaffar (2012) "Incommensurability or Vagueness? A Comment on Rabinowicz and Sugden", *Proceedings of the Aristotelian Society*, (2012) 112 (3): 333-338.

Qizilbash, Mozaffar. (2014) "Incommensurability and Vagueness: Is the Vagueness View Defensible?" *Ethical Theory and Moral Practice*, (2014) 17 (1): 141-153.

Qizilbash, Mozaffar. (2019) "The Market, Utilitarianism and the Corruption Argument." *International Review of Economics* 66.1 (2019): 37-55.

Rahman, Tauhidur, Ron C. Mittelhammer, and Philip Wandscheider. (2005) *Measuring the Quality of Life across Countries: A Sensitivity Analysis of Well-being Indices*. No. 2005/06. WIDER Research Paper, 2005.

Ravallion, Martin. (2010) *Mashup Indices of Development*. The World Bank, 2010.

Reiss, Julian. (2001) "Natural Economic Quantities and their Measurement." *Journal of Economic Methodology* 8.2 (2001): 287-311.

Reiss, Julian. (2013) *Philosophy of Economics: A Contemporary Introduction*. Routledge, 2013.

Richardson, Henry S. (1997) *Practical Reasoning about Final Ends*. Cambridge University Press, 1997: 110-118.

Riddle, Christopher A. (2014) *Disability and Justice: The Capabilities Approach in Practice*. Lexington Books, 2014: 41-59.

Robeyns, Ingrid. (2003a) "The Capability Approach: An Interdisciplinary Introduction." Paper presented at the 3rd International Conference on the Capability Approach, Pavia, 6 September.

Robeyns, Ingrid. (2003b) "Sen's Capability Approach and Gender Inequality: Selecting Relevant Capabilities." *Feminist Economics* 9.2-3 (2003): 61-92.

Robeyns, Ingrid. (2005a) "The Capability Approach: a Theoretical Survey." *Journal of Human Development* 6.1 (2005): 93-117.

Robeyns, Ingrid. (2005b) "Selecting Capabilities for Quality of Life Measurement." *Social Indicators Research* 74.1 (2005): 191-215.

Robeyns, Ingrid. (2006) "The Capability Approach in Practice." *Journal of Political Philosophy* 14.3 (2006): 351-376.

Robeyns, Ingrid. (2011) "The Capability Approach." in *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/archives/win2016/entries/capability-approach/ (Accessed 2020).

Robeyns, Ingrid. (2017) *Wellbeing, Freedom and Social Justice: The Capability Approach Re-Examined*. Open Book Publishers, 2017.

Runhardt, Rosa. (2015) *Causal Inquiry in the Social Sciences: the Promise of Process Tracing*. Diss. The London School of Economics and Political Science (LSE), 2015.

Russell, Daniel C. (2009) *Practical Intelligence and the Virtues*. Oxford University Press, 2009.

Saith, Ruhi. (2001) *Capabilities: the Concept and its Operationalisation, QEH Working Paper **66***, University of Oxford, February 2001.

Sambanis, Nicholas. (2004) "What is Civil War? Conceptual and Empirical Complexities of an Operational Definition." *Journal of Conflict Resolution* (2004) 48.6: 814-858.

Schokkaert, Erik (2009) “The Capabilities Approach” in Paul Anand, Prasanta K. Pattanaik and Clems Puppe eds. *The Handbook of Rational and Social Choice* (2009: 542-566). Oxford University Press, Oxford.

Sen, Amartya. (1970a) "Interpersonal Aggregation and Partial Comparability." *Econometrica: Journal of the Econometric Society*. 393-409.

Sen, Amartya. (1970b) *Collective Choice and Social Welfare*. Vol. 11. Elsevier, 2014.

Sen, Amartya. (1976) "Poverty: An Ordinal Approach to Measurement." *Econometrica: Journal of the Econometric Society* (1976): 219-231.

Sen, Amartya. (1977) "On Weights and Measures: Informational Constraints in Social Welfare Analysis." *Econometrica: Journal of the Econometric Society*. 1539-1572.

Sen, Amartya. (1979a) *Collective Choice and Social Welfare*. Harvard University Press, 2018.

Sen, Amartya. (1979b) "Equality of What." *The Tanner Lecture on Human Values* 1 (1979).

Sen, Amartya. (1979c) "Utilitarianism and Welfarism." *The Journal of Philosophy* 76.9 (1979): 463-489.

Sen, Amartya. (1980a) "Description as Choice." *Oxford Economic Papers* 32.3: 353-369.

Sen, Amartya. (1980b) "Plural Utility." *Proceedings of the Aristotelian Society*. Vol. 81. Aristotelian Society, Wiley, 1980.

Sen, Amartya. (1983) "Poor, Relatively Speaking." *Oxford Economic Papers* 35.2 (1983): 153-169.

Sen, Amartya. (1984) *Resources, Values, and Development*. Massachusetts: Harvard University Press

Sen, Amartya. (1985a) *Commodities and Capabilities*. Evelsier: Amsterdam.

Sen, Amartya. (1985b) "Well-being, Agency and Freedom: The Dewey Lectures 1984." *The Journal of Philosophy* 82.4 (1985): 169-221.

Sen, Amartya. (1987a) "Reply" in Geoffrey Hawthorn, Ed. *The Standard of Living*. Cambridge University Press, 1987.

Sen, Amaryta. (1987b) *On Ethics and Economics*. MA: Blackwell Publishing, 1987.

Sen, Amartya. (1990) ''Gender and Cooperative Conflicts,'' in Irene Tinker (ed.) *Persistent Inequalities*, pp. 123-149. New York: Oxford University Press.

Sen, Amartya. (1992) *Inequality Reexamined*. Clarendon Press: New York.

Sen, Amartya. (1993) "Capability and Well-Being" in Sen, A. and Nussbaum, M. (eds) *Quality of Life*, (pp 34-66). Oxford: Clarendon Press.

Sen, Amartya. (1995) ''Gender Inequality and Theories of Justice,'' in M. Nussbaum and J. Glover (eds.) *Women, Culture and Development*, (pp 259-73). Oxford: Clarendon Press, 1995.

Sen, Amartya. (1999a) "The Possibility of Social Choice." *The American Economic Review* 89.3: 349-378.

Sen, Amartya. (1999b) *Development as Freedom*. New York: First Anchor Books, 2000.

Sen, Amartya. (2004a) "Economic Methodology: Heterogeneity and Relevance." *Social Research: An International Quarterly* 71.3 (2004): 583-614.

Sen, Amartya. (2004b) "Capabilities, Lists, and Public Reason: Continuing the Conversation." *Feminist Economics* 10.3 (2004): 77-80.

Sen, Amartya. (2004c) "Incompleteness and Reasoned Choice." *Synthese* 140.1/2 (2004): 43-59.

Sen, Amartya. (2009a) *The Idea of Justice.* Cambridge, MA: Harvard University Press.

Sen, Amartya. (2009b) "Capability: Reach and Limits" in Chiappero-Martinetti, E. (Ed.) *Debating Global Society: Reach and Limits of the Capabilities Approach*. Milan, Italy: Fondazione Giangiacomo Felbrinelli. (pp 15-31).

Sen, Amartya, and James E. Foster. (1973) *On Economic Inequality*. Oxford University Press, 1997.

Sen, Amartya, and Bernard Williams. (1982) “Introduction: Utilitarianism and beyond” in Amartya Sen and Bernard Williams, Eds. *Utilitarianism and beyond*. Cambridge University Press, 1982.

Seth, Suman, and Mark McGillivray. (2018) "Composite Indices, Alternative Weights, and Comparison Robustness." *Social Choice and Welfare* 51.4 (2018): 657-679.

Srinivasan, Thirukodikaval Nilakanta. (1994) "Human Development: a New Paradigm or Reinvention of the Wheel?." *The American Economic Review* 84.2 (1994): 238-243.

Stanton, Elizabeth A. (2007) "The Human Development Index: A history." *PERI Working Papers* (2007): 85.

Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi. (2009) "Report by the Commission on the Measurement of Economic Performance and Social Progress." (2009).

Sunstein, Cass R. (1994) "Incommensurability and Valuation in Law." *Michigan Law Review* 92.4: 779-861.

Tal, Eran (2013). Old and New Problems in Philosophy of Measurement. *Philosophy Compass* 8 (12): 1159-1173.

Tal, Eran. (2015) “Measurement in Science” *Stanford Encyclopedia Artic*le http://plato.stanford.edu/archives/sum2015/entries/measurement-science/ (Accessed 01-2016).

Tallabs, Valente (2018) ‘Study Case: Applying the Pentagram to the Municipality of Atlixco, State of Puebla in Mexico’ in *Discussion Paper III: Possible Measurement of the Common Good’s Dynamic, UPAEP- IPBC*

Temkin, Larry S. (2014) *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford University Press, 2014.

Teschl, Miriam, and Flavio Comim. (2005) "Adaptive Preferences and Capabilities: Some Preliminary Conceptual Explorations." *Review of Social Economy* 63.2 (2005): 229-247.

Ul Haq, Mahbub (1995) *Reflections on Human Development*. New York, USA: Oxford University Press.

UNDP. (1990) "Human Development Report 1990." *United Nations Development Program* New York: Oxford University Press, 1990.

UNDP. (1995) "Human Development Report 1995." *United Nations Development Program* New York: Oxford University Press, 1995.

UNDP. (1997) "Human Development Report 1997." *United Nations Development Program* New York: Oxford University Press, 1997.

Vallentyne, Peter. (2005) "Capabilities vs. Opportunities for Well-being." *Philosophy Publications* (2005).

Van Ootegem, Luc, and Elsy Verhofstadt. (2012) "Using Capabilities as an Alternative Indicator for Well-being." *Social Indicators Research* 106.1 (2012): 133-152.

Vazquez, Oscar Garza. "From the Idea of Justice to the Idea of Injustice: Mixing the Ideal, Non-ideal and Dynamic Conceptions of Injustice." *Bath Papers in International Development and Wellbeing, No. 27* (2013).

Von Neumann, John, Oskar Morgenstern, and Harold William Kuhn. (1953) *Theory of Games and Economic Behavior (commemorative edition)*. Princeton University Press, 2007.

Walsh, Vivian. (2003) "Sen after Putnam." *Review of Political Economy* 15.3 (2003): 315-394.

Ward, Ann. (2010) "Justice as Economics in Aristotle's Nicomachean Ethics." *Canadian Political Science Review* 4.1 (2010): 1-11.

Whipps, Judy, and Danielle Lake. (2004) "Pragmatist Feminism." in *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/archives/win2017/entries/femapproach-pragmatism/ (Accessed 2016).

Williams, Bernard. (1972) *Morality: An Introduction to Ethics*. Cambridge University Press, 2012.

Williams, Bernard. (1976) *Problems of the Self: Philosophical Papers 1956-1972*. Cambridge University Press, 1976.

Williams, Bernard. (1981) “Conflicts of Values,” in *Moral Luck*, Cambridge: Cambridge University Press.

Williams, Bernard. (1987) "The Standard of Living: Interests and Capabilities." in Geoffrey Hawthorn, Ed. *The Standard of Living* (1987): 94-102.

Wolff, Jonathan, and Avner De-Shalit. (2007) *Disadvantage*. Oxford University Press, 2007.

Wolff, Jonathan, and Dirk Haubrich. (2008) "Economism and its Limits." In Robert Goodin, Michael Moran, and Martin Rein Eds. *The Oxford Handbook of Public Policy* (2008).

Young, Iris Marion. (1990) *Justice and the Politics of Difference*. Princeton University Press, 2011.

Zhou, Peng, B. W. Ang, and D. Q. Zhou. (2010) "Weighting and Aggregation in Composite Indicator Construction: A Multiplicative Optimization Approach." *Social Indicators Research* 96.1 (2010): 169-181.

Zimmerer, Karl S. (2015) "Environmental Governance through ‘Speaking Like an Indigenous State’ and Respatializing Resources: Ethical Livelihood Concepts in Bolivia as Versatility or Verisimilitude?" *Geoforum* 64 (2015): 314-324.