

# Measure Development and the Hermeneutic Task

Laura M. Cupples

## 1 Introduction

I examine the dynamics of measure development using two case studies: temperature, and health-related quality of life. I argue, following Bas van Fraassen (2008) and Leah McClimans (2010a), that in each case these dynamics have a hermeneutic structure. Just as the hermeneutic circle represents an attempt to overcome epistemic circularity in the interpretation of a text, so too must epistemic circularity be overcome in measure development. Namely, we must establish a mathematical relationship between observable (e.g. volume) and unobservable variables (e.g. temperature), while the value of the unobservable variable remains unknown.

I show that Gadamer's ([1960]1989) philosophical hermeneutics are an effective lens through which to examine the development of the temperature standard as described by Hasok Chang (2004). First, the normative force of tradition found in Gadamer's hermeneutics mirrors Chang's "principle of respect". Second, Gadamer argues that in order to interpret the meaning of a text, it must be applied in a concrete context. Similarly, in measurement we must bridge the gap between abstract theory and concrete practice through the operationalization of the measure. Finally, Gadamer's emphasis on coherence between part and whole is congenial to Chang's justificatory philosophy of progressive coherentism.

I pause briefly to discuss what the analogy between measure development and the hermeneutic task is "good for"—what epistemic purpose it serves in these two case studies. I argue first that the analogy shows that measure development is historically

situated; second, that it shows that the epistemic circularity attending measure development is not vicious; and finally, that it shows that the definition of measurement concepts is not wholly subjective or stipulative.

Despite similar grounding in hermeneutics, I note an important difference between measure development for temperature and for health-related quality of life. Namely, while the meaning of temperature can be standardized, the meaning of health-related quality of life cannot (McClimans 2010a). A strategy of progressive coherentism, i.e., Chang's epistemic iteration (2004), ultimately leads to a theory of heat and temperature, as well as to determinate values for temperature. The same cannot be said for health-related quality of life. Health-related quality of life is imperfectly understood, according to McClimans (2010a). Asking genuine questions about its meaning will aid in interpretation, but just as in Gadamer's hermeneutics, our horizons of meaning must remain open. This is because there are always new questions we might ask about health-related quality of life in various contexts.

The standardization of meaning for the temperature concept represents a limit to the analogy with hermeneutics, as Gadamer argues that the meaning of a text should remain open to new interpretations when encountered by new persons in new historical contexts. On the other hand, I argue that the indeterminacy we find in health-related quality of life measurement is a result not only of an analogy with the hermeneutic task, but of full-fledged participation in it. Health-related quality of life measures are texts authored by researchers and interpreted by respondents, each of whom brings his or her own experiential background to the encounter.

## 2 Setting the stage

## 2.1 Van Fraassen, coordination, and the hermeneutic task

In his 2008 monograph, *Scientific Representation: Paradoxes of Perspective*, Bas van Fraassen briefly suggests that solving measurement's problem of coordination is a hermeneutic task. When scientists work to develop new measures, they must seek out principles of coordination that link the quantity to be measured to specific values on a measurement scale. As they do so, they must try to answer two entangled questions: "(1) What counts as a measurement of (physical quantity) X? and (2) What is (that physical quantity) X?" (van Fraassen 2008, p. 116). These questions together form the "problem of coordination". They cannot be answered separately from one another, according to van Fraassen, and approaching them together seems to mire us in epistemic circularity. How do we determine the value of our quantity without an established measurement procedure, and how do we establish a measurement procedure without knowing what function relates the quantity we are able to observe (for instance, volume of mercury) to the quantity we are trying to uncover (for instance, temperature)? Hermeneutics famously addresses itself to a similar form of circularity, and thus the hermeneutic task can shed light on the problem of coordination. We cannot uncover the meaning of the subject matter of a text or text analogue without presupposing some as yet unjustified, and possibly erroneous meaning coming in to the task (Warnke 1987, p. 76). That is, we cannot take "a view from nowhere" (van Fraassen 2008, p. 122). We must presuppose certain other quantities or concepts (such as length, volume, pressure), and make certain choices about how to proceed in measure development (such as assuming certain fixed temperature points). We must use this provisional assumption as a basis for revising our understanding moving forward. How do we address ourselves to this apparent circularity, and what can measure developers learn from hermeneutics?

## 2.2 Laying out the analogy

If measure development is analogous to the hermeneutic task, what elements in each system correspond to what? Van Fraassen (2008, Chapter 5) is vague on this point, leaving the reader to infer the elements of the analogy, and while Chang's (2004) description of measure development is, on my reading, hermeneutic in structure, he does not specifically make reference to hermeneutics in his work. Instead, I use his more detailed description of the development of the temperature measure to expand on van Fraassen's claims. Of the authors I discuss, McClimans (2010a) makes the analogy between measure development and the hermeneutic task most explicit in her discussion of patient-reported outcome measures (PROMs) and health-related quality of life, one target of such measures. While the analogy McClimans discusses is suggestive of the correspondence relationship between temperature measurement and hermeneutics, I also believe it carries over somewhat imperfectly to the realm of thermometry.

Hermeneutics explores the meaning of the subject matter of a text or text analogue as well as the meaning of the text itself. In this analogy, the instrument and its indications stand in for the text—literally in the case of PROMs, which take the form of surveys, and metaphorically in the case of the thermometer. The subject matter of the text or text analogue is the quantity or construct to be measured. In hermeneutics the meaning of the subject matter is jointly determined by tradition and application, through a dialogical process that involves questioning the text and learning from one's encounter with it. Similarly, the meaning of the quantity or construct to be measured—its definition—is (ideally) jointly determined by theory and practice. This determination takes place through an iterative process that involves putting a provisional measurement

system to work in an experimental setting, observing any incoherencies, and adjusting one's definition in response to one's observation.

### 3 Chang's progressive coherentism

I expand upon van Fraassen's suggestion that measure development is hermeneutical by looking to Hasok Chang's extended analysis of the development of the thermodynamic temperature scale. Like van Fraassen, Chang recognizes the epistemic circularity inherent in measure development. Chang describes what he calls the "problem of nomic measurement" as follows:

1. We want to measure a quantity X.
2. Quantity X is not directly observable, so we infer it from another quantity Y, which is directly observable.
3. For this inference, we need a law that expresses X as a function of Y, as follows:  
$$X = f(Y).$$
4. The form of this function f cannot be discovered or tested empirically, because that would involve knowing the values of both Y and X, and X is the unknown variable that we are trying to measure (Chang 2004, p. 59).

The problem of nomic measurement is illustrated by researchers' attempts to progress from ordinal thermoscopes to numeric thermometers. In this case, X is temperature and Y is the volume of the thermometric fluid. Once points at 0° and 100° Celsius had been fixed, researchers needed to find a way to fill in the rest of the temperature scale. The question is, were they justified in assuming that temperature increases linearly with volume of thermometric fluid? It seems impossible to answer this question without knowing temperature values independently (Chang 2004, p. 60-64). And yet, as Chang

shows, by using a progressive, coherentist strategy, researchers were able to develop and justify the thermodynamic temperature scale.

Chang describes researchers as reliant on two methodological commitments: first, “the principle of respect”, and second, “the imperative of progress” (2004, p. 44). The principle of respect involves a commitment to “respect the prior standard as far as it is plausible to do so” (Chang 2004, p. 44). We must recognize that our attempts at measure development are historically situated. We cannot, as van Fraassen observes, take a “view from nowhere”, but must work within the tradition we find ourselves in (2008, p. 122). Chang (2004, p. 84-89) demonstrates this principle by describing the way early researchers relied on coherence with physical sensation to develop their temperature measures. They chose to respect sense data that taught them that thermometric fluids tended to expand with the addition of heat. While they had only imperfect justification for trusting sense data, choosing to treat it with too much skepticism would have made the advancement of thermometry impossible. Thus, sense data was treated with at least provisional trust. This trust was not indefeasible, as researchers left open the possibility that sensation might at times be mistaken. Indeed, the hope was that thermoscopes built on the foundation of sense data would, when complete, be able to challenge that same sense data (Chang 2004, p. 84-89).

Chang’s second methodological commitment is to the imperative of progress. We want each new iteration of measure development to improve upon the epistemic virtues of its predecessors. Thus, we may look for improvements in the consistency, precision, or scope of our measures (Chang 2004, p. 44). Henri Victor Regnault’s painstaking work testing the consistency of measurement values obtained using various thermometric fluids can be seen as an advance in both precision and consistency. Josiah Wedgwood’s

attempt to expand the standard into the high temperature range through pyrometry can be seen as an advance in scope. Progress is achieved through “creative evolution” (Chang 2004, p. 46), with each new step in measure development building on its predecessor, but not straightforwardly derived from it. Chang calls this process of creative evolution “epistemic iteration” (2004, p. 44-46), emphasizing that each step brings us closer to our epistemic goals.

Chang (2004, p. 186) argues that a measurement system is not complete until researchers bridge the gap between the abstract and the concrete, between theoretical concepts and practice. Theoretical standards must be operationalized if they are to have empirical content. Researchers had established a working temperature standard long before Lord Kelvin tied temperature measurement to the ideal gas law (Chang 2004, p. 159). Chang (2004, p. 206) argues that bringing theory and practice together involved taking a theoretical system and creating an image of it—an image that would function as an idealization of a matching physical system. A series of iterative corrections were then made to bring the values generated by the two systems closer together. The end result was a determinate temperature value in each context (Chang 2004, p. 211-212).

## 4 Hermeneutics

### 4.1 The hermeneutic task

If the development of the thermometer is analogous to the hermeneutic task, what elements does such a task comprise? The hermeneutic task is the task of interpreting the meaning of a text or text analogue, as well as the meaning of the subject matter of the text. This task is worked out differently for different hermeneutic scholars. Romantic era scholars such as Friedrich Schleiermacher saw their task as uncovering the creative

intention behind a text. Because he saw misunderstanding as the natural result of our attempts to interpret historical works, Schleiermacher emphasized the need for a rigorous hermeneutic method. This method was both grammatical and psychological: the grammatical method carefully examined dialect, sentence structure, and genre while the psychological method called for the reader to transpose himself into the author's position in an attempt to recreate the conditions and mindset under which the work was originally created. Finally, coherence between grammatical and psychological elements constrained possible interpretations (Warnke 1987, p. 10-15).

Wilhelm Dilthey, like Schleiermacher, sought to uncover the original vantage point under which creative works were made. He sought a methodology that would both separate the human sciences from the natural sciences and place the human sciences on a secure footing. He emphasized the need for a rigorous hermeneutic method that would make objective understanding in the human sciences possible. Dilthey's hermeneutic philosophy distinguished between two types of experience—scientific experience, which was repeatable and thus verifiable, and life experience, which could not be repeated in the same way. For Dilthey, life experience was the basis upon which the human sciences were built. Individually, we learn from life experience and begin to see the world in new ways. Collectively, life experience forms a kind of *Geist* or spirit that infuses a society's understanding of their shared history. Despite his insight that understanding is conditioned by historical experience, Dilthey, like Schleiermacher, believed that it was necessary to transpose ourselves into the position of historical actors if we wished to understand their original intentions. According to Gadamer, Dilthey's concern that human sciences be made objective undermined his more valuable insight that understanding must be historically situated (Warnke 1987, p. 26-34).

Hans Georg Gadamer broke with both Schleiermacher and Dilthey in his 1960 monograph, *Truth and Method*, in important ways. First, he distinguished between two types of understanding: understanding of an author's creative intentions, and substantive understanding of a truth claim. For Gadamer, the purpose of hermeneutics was primarily the second. He eschewed calls for a rigorous or defined hermeneutic methodology, and instead worked to describe the conditions under which understanding was made possible. For this reason, Gadamer did not believe that we should attempt to transpose ourselves into the position of the original author of a text. Instead, he believed that the perspective brought to bear on a text by the interpreter was essential to the meaning of a text (Warnke 1987, p. 7).

Gadamer (Warnke 1987, p. 26-34) adopted Dilthey's insight that as interpreters, our understanding is inevitably historically situated (Gadamer 1989, p. 300-307). He emphasized the essential role of tradition in our coming to understand a text. Tradition, Gadamer argued, strongly influences our prejudices about the meaning of a text. Beginning our encounter with a text with some pre-understanding or prejudice, Gadamer argued, is a necessary condition for coming to understand its meaning more fully. In seeking to understand a text, we place our pre-understandings at risk and allow them to be challenged by the text itself (Gadamer 1989, p. 267-271; Warnke 1987, p. 97). We engage in dialogue with the text, posing questions that lead to a greater understanding of its subject matter and allowing the text to answer those questions (Gadamer 1989, p. 362-370).

In this way, we achieve what Gadamer called a "fusion of horizons" (1989, p. 306-307)—a new, shared understanding of the subject matter (Warnke 1987, p. 101-103). This fusion of horizons, however, is not the final word when it comes to the meaning of a

text. Tradition and the subjectivity of interpreters are ever evolving, and so are the pre-understandings that may be brought to bear on a historical text. When trying to learn the truth about a subject matter, we continue to address new questions to the text that treats it. Our horizons remain open, as does our interpretation of the text. Thus, according to Gadamer (1989, p 374), we must learn to accept a certain degree of indeterminacy with regard to meaning.

#### 4.2 Whose hermeneutics?

I argue that among the hermeneutic philosophies available to us, Gadamer's is best suited to illuminate the process of measure development. It is important that Gadamer saw texts as truth claims, and not merely as aesthetic objects, the meaning of whose subject matter was determined by the author's intent. The meaning of a text was informed by historical tradition, and not by a single moment in time (Warnke 1987, p. 7-8, 100-106). Both van Fraassen and Chang subscribe to this historically situated view of measure development. In contrast, in Schleiermacher's hermeneutics the meaning of a text is fully determined by the author's intention. I argue that Schleiermacher's hermeneutic task mirrors the naïve conventionalist approach to measurement epistemology rejected by both van Fraassen and Chang, wherein the definition of the target of measurement is stipulated and arbitrary. In both cases, the meaning or definition is ahistorical. The meaning of the text is not informed by intervening tradition, and the definition of the quantity to be measured does not depend on previous iterations.

Secondly, Gadamer's hermeneutics are dialogical in nature. Just as Gadamer's hermeneutics seeks to reach agreement about the meaning of a truth claim through dialogue, measurement seeks a nearer approximation, through iterative progression, if not

to truth, then at least to epistemic virtue (i.e., scope, precision, accuracy). Gadamer's (1989, p. 268-271) claim that tradition carries normative, though not indefeasible force in our attempts to make sense of a text is also important. Tradition informs our pre-understandings of a text, constraining subjectivity in interpretation. Those provisional pre-understandings are a condition for the possibility of uncovering meaning (Warnke 1987, p. 98-99). Theory carries a similar normative force in measure development (Chang 2004, p. 211). Both van Fraassen and Chang propose a process of iterative refinement and correction of candidate standards that brings measurement values into agreement with one another and with theory. This process is dialogical, and truth (or epistemic virtue)-seeking, as is Gadamer's hermeneutics. We create a new iteration of our standard, or gain a new understanding of the subject matter of a text, by putting the old understanding at risk and allowing ourselves to learn from the encounter with the text cum instrument. In so doing, we come to an agreement about the content of the text, or the meaning and value of our measurement concept.

## 5 Gadamer and Chang

In Section 4.2, I explained at a general level why Gadamer's approach to hermeneutics is a better analogue to the dynamics of measure development than other historical approaches to hermeneutics. In this section, I explain in greater detail why the development of the temperature standard in particular, as described by Chang, is hermeneutical in Gadamer's sense.

For Gadamer, the subjectivity of our interpretations is constrained in a number of ways. First, it is limited by the "anticipation of completeness". The anticipation that the text has something true to tell us and that it forms a unified and coherent whole informs

and regulates our provisional interpretation of the text. That anticipation is also the basis upon which we test those provisional interpretations against the text itself. We assume that the parts of the text will form a self-consistent whole when interpreted correctly, and if they fail to do so, we have reason to revise our interpretation (Gadamer 1989, p. 293-294; Warnke 1987, p. 82-91). Warnke (1987, p. 83) for instance, gives the example of provisionally taking a book to be a detective story, and then finding that under that interpretation, the elements of the plot fail to form a coherent whole. Our encounter with the text, along with our anticipation of its completeness, lead us to revise our initial interpretation.

Second, the subjectivity of our interpretation is limited by the need for coherence between the historic text and its present-day application. The application of a text to our own lives is an essential part of the interpretive act (Gadamer 1989, p. 307; Warnke 1987, p. 91-99). Gadamer and Warnke illustrate the need for both the normative force of tradition and for present day application through an analogy with Aristotelian ethics. For Aristotle, we must not only possess theoretical knowledge about general ethical norms, but we must also translate those norms into concrete and situationally appropriate action (Gadamer 1989, p. 312-324; Warnke 1987, p. 92-93). If we are to understand the virtue of generosity or justice or courage writ large, we must know how to act it out in a specific situation.

How do these dynamics align with Chang's progressive coherentism? Recall that measure development for Chang involves commitments to both the principle of respect and the imperative of progress. Like Gadamer, Chang recognizes the indispensability of tradition—in his case, scientific tradition. Whether relying on sense data or ordinal thermoscopes, Chang's researchers provisionally accepted the authority of scientific

tradition. Doing so was a precondition of scientific progress (Chang 2004, p. 43-46). Having done so, they tested experimental findings based on those traditional methods against an anticipation of completeness. These constraints guided researchers in establishing more coherent measurement systems by allowing them to refine and correct the systems they had provisionally endorsed (Chang 2004, p. 43-44).

Furthermore, recall that for Chang (2004, p. 197) a measurement system is incomplete until the gap between the abstract and the concrete has been bridged. Abstract concepts, such as temperature and pressure, must ultimately be married to operation, as occurred when Lord Kelvin linked temperature measurement first to Carnot's idealized heat engine and then to the ideal gas law (Chang 2004, p. 173-197). Abstract concepts carry with them the normativity of tradition, or if you'd rather, of a general ethical norm, while their operationalization is the acting out of that norm or tradition in a concrete, empirical context (Gadamer 1989, p. 324; Warnke 1987, p. 92-93). In this way, the norm, or the abstract concept, takes on empirical content (Chang 2004, p. 197; van Fraassen 2008, Chapter 5). Temperature and pressure become more than mere mathematical terms; they become the targets of measurement.

## 6 Hermeneutics and measure development in the human sciences

In this section, I extend the analogy between measure development and the hermeneutic task to encompass measures in the human sciences as well, relying Leah McClimans's philosophical work on the measurement of health-related quality of life and subjective well-being.

### 6.1 McClimans, PROMs, and Gadamer

McClimans addresses the relationship between measure development and Gadamer's philosophical hermeneutics in her (2010a) paper "A Theoretical Framework for Patient-Reported Outcome Measures." She sees patient-reported outcome measures (PROMs) as texts, the meaning of whose subject matter is to be uncovered. PROMs are survey instruments used to measure health-related quality of life, health status, physical function, and a host of other constructs. For the purposes of this paper, I will specifically be focusing on health-related quality of life as a target of PRO measurement. PROMs targeting health-related quality of life pose questions to respondents about pain, mobility, functional status, fatigue, emotions, and social connectedness, among others. When respondents answer these questions, they give researchers access to phenomena that would not otherwise be observable.

Like the development of the thermometer, PROM development is plagued by epistemic circularity. Recall Chang's problem of nomic measurement. Here health-related quality of life is the unobservable phenomenon we wish to measure, and the factors that contribute to it—factors that respondents report on such as self-care, mobility, and emotional well-being—are made observable for researchers through those reports. Presumably health-related quality of life is a function of these factors, but exactly how they come together to make up health-related quality of life is unknown. Most scholars believe there is no gold standard for health-related quality of life—no independent means of measuring it apart from discerning its relationship to these factors (McClimans 2010a).

Hadorn et al. (1995) suggest that a single, global question posed to respondents about their health-related quality of life could serve as a gold standard, while Kowalski et al. (2012) similarly suggest that idiographic measures of quality of life, such as the SEIQOL could serve as a gold standard. In both cases, the feasibility of the gold

standard depends on health-related quality of life being entirely subjective. There is good reason to suppose that respondents' assessments of their own health-related quality of life are not infeasible and also that respondents have only imperfect insight into the factors affecting their health-related quality of life. A person's memory of recent experience is imperfect, for instance, and they may not be consciously aware of all the criteria they value as part of their quality of life. I do not believe that health-related quality of life is entirely subjective. For instance, sociocultural factors affect individuals' conceptualization and assessment of their quality of life, and rightly so. Some cultures value independence, while others value commitment to family. These factors legitimately influence quality of life, and in some sense of the word, are objective. For these reasons, I disagree with Hadorn et al. (1995) and Kowalski et al.'s (2012) claims that quality of life has a gold standard.

At present, different researchers, and different measures, posit different relationships between health-related quality of life and its constituents. Apart from a general consensus that health-related quality of life is multidimensional, encompassing physical, emotional, and social well-being, there is no widely accepted theory of health-related quality of life to act as a norm or constraint on its many measures (Hunt 1997). McClimans's purpose in her (2010a) paper is to provide a framework for greater theoretical development for PROMs.<sup>1</sup>

McClimans (2010a) sees PROM development as dialogical and truth seeking. She looks to Gadamer's (1989, p. 369-370) logic of question and answer to help make sense of the conversation taking place between researchers and respondents. Importantly,

---

<sup>1</sup> Anna Alexandrova (2017) suggests that any scientifically useful theory of well-being would need to be contextual. She posits a theory of child well-being in her 2017 monograph *A Philosophy for the Science of Well-being*, but does not offer or recognize a consensus theory of health-related quality of life.

the subject matter of their conversation, health-related quality of life, is imperfectly understood. This imperfect understanding has consequences for the type of questions it is appropriate for us to ask about the subject matter (McClimans 2010a). McClimans (2010a) argues, following Gadamer (1989, p. 367), that we should be asking genuine questions about health-related quality of life instead of merely apparent questions. We ask genuine questions when we do not fully understand a subject matter and wish to learn more about it. Genuine questions are open to interpretation and do not have pre-determined answers. For instance, we might ask about the meaning of a particular Shakespearean sonnet, or what makes the Mona Lisa beautiful. We ask apparent questions on the other hand when a subject matter is well known to us and when there are definite criteria for a correct response (McClimans 2010a). Many pedagogical questions such as, what is the capital of South Carolina, are merely apparent.

Unfortunately, because PROM questions are typically standardized, they function as apparent questions and close the door for us to learn more about the meaning of health-related quality of life from the respondents with whom we are in dialogue (McClimans 2010a). McClimans (2010a) suggests that researchers currently reopen that door by supplementing standardized PROMs with think aloud studies and qualitative interviews. Initially, when researchers are trying to define the concept to be measured, respondents are queried about what aspects of health-related quality of life are most relevant to them given their particular illness or disability. For instance, breast cancer survivors undergoing reconstructive surgery nominated the following outcomes as important to their surgical experience and recovery: satisfaction with breasts, overall outcome, process of care, as well as physical, psychosocial, and sexual well-being (Pusic et al. 2009, p. 345).

Later, when researchers begin to test a preliminary measure, respondents are asked to speculate about the meaning of the questions they are faced with and to share their reasons for responding to those questions as they do. While typically these interviews are conducted to ensure that respondents are interpreting questions as researchers intend, they need not serve that role (McClimans 2010a). McClimans (2010a) argues that we should take respondents' answers and interpretations seriously, even when they are unexpected and challenge our preconceived ideas about health-related quality of life. For instance, does a disability that limits mobility necessarily affect one's health-related quality of life, and if it does, how does it affect it? Does adaptation to the circumstances of one's illness play a legitimate role in health-related quality of life, or is that role illusory? When researchers are willing to put their own understandings at risk and to learn from respondents, a new and shared understanding—what Gadamer calls a fusion of horizons—becomes possible (Gadamer 1989, p. 267-270; McClimans 2010a; Warnke 1987, p. 103). In this way, researchers can better understand the meaning of health-related quality of life data and the inferences they draw from those data are more likely to be sound (McClimans 2010a; Schwartz & Rapkin 2004).

Importantly, McClimans (2010a) rejects the arguments of some researchers (see e.g., Schwartz & Rapkin 2004) that the meaning of health-related quality of life is inherently subjective, and perhaps even idiosyncratic. Like Gadamer (Warnke 1987, p. 82-99), she sees certain constraints on interpretation at work in the hermeneutics of health-related quality of life. By now, those constraints should be familiar—coherence, the anticipation of completeness, and the indispensability of application as a part of the interpretive act. While respondents' interpretations should be taken seriously, and we should grant them the authority to teach us more about health-related quality of life, they

are not infeasible. We cannot say that anything goes when it comes to health-related quality of life (McClimans 2010b). McClimans (2010b) offers the example of a woman who has been culturally conditioned to accept female circumcision as a normal social practice. While this woman may claim that female circumcision is not qualitatively different from male circumcision and is perfectly compatible with a good health-related quality of life, if we consider the coherence of her claim there are legitimate reasons to be skeptical. Unlike male circumcision, female circumcision tends to be practiced in the context of oppressively patriarchal cultures that limit the well-being of women in a number of significant ways (McClimans 2010b).

But who decides what counts as coherence? Should not respondents' assessments of their own health-related quality of life be given priority over that of a third party? I would agree that, all other things being equal, they should. That is why, whenever possible, we pose questions about health-related quality of life to patients, rather than asking physicians or other proxies to speak for them. There is good reason to suppose that just as in the case of hermeneutics, agreement must be sought about the meaning of the subject matter between author and interpreter. If they are to reach agreement, sharing a common historical or cultural tradition (for instance, a cultural tradition of seeing men and women as properly equal) is helpful. Here the author is the respondent, who has given researchers their claim about what health-related quality of life looks like for them. The interpreter is the researcher or clinician looking over the responses to the instrument's queries. This seeking of agreement is what leads to a fusion of horizons in hermeneutics.

Suppose that a person with a significant disability provides responses that indicate an excellent quality of life. The researcher/interpreter may have doubts at first about the

respondent's claim. Through qualitative interviewing, he or she may come to realize that, when appropriate accommodations and supports are available, persons with disabilities have a health-related quality of life very similar to people who are able-bodied. If we push on their claims to have a good quality of life, citing concerns about coherence, we may find that in cases where those supports and accommodations are not available, their quality of life is genuinely impaired. There is nothing inherently "bad" about having a significant disability, provided there are appropriate social and environmental supports in place. Thus the claim of the respondent is, under those circumstances, perfectly coherent given that the respondent and the interpreter share a certain underlying tradition: namely, that of the social model of disability. Where author/respondent and interpreter/researcher do not share that tradition, it will be very difficult for them to come to agreement about the meaning of health-related quality of life, or about whether that meaning is coherent.

7 What is this analogy good for?

Typically, when one posits an analogy, that analogy serves a particular epistemic purpose. One purpose of the analogy between measure development and the hermeneutic task (as illustrated in van Fraassen's (2008, Chapter 5) work) is to show that measure development must be historically situated. There is no ahistorical, unbiased position from which to proceed. We must start with certain presuppositions about our measure and our quantity to be measured, as well as a working knowledge of adjacent concepts (in the case of temperature: length, volume, and eventually pressure), and make an initial choice about how to proceed. Whether that first, underdetermined choice will be sustainable depends on its coherence.

Additionally, the analogy between measure development and the hermeneutic task demonstrates that the circularity attending measure development is not vicious. Recall, for example, that we cannot design a thermometer based on empirical law without the data supplied by a provisional thermometer. The solution to this dilemma: to develop and justify measurement systems iteratively, through a creative, dialogical process, constrained by coherence and tradition. We put provisional measures to work, observe their outcomes, and revise based on limitations and incoherencies (Chang 2004, Chapter 1).

Finally, the analogy suggests that the meaning of health-related quality of life, and temperature, while initially underdetermined, is not wholly subjective. It cannot merely be stipulated, either by researchers or by respondents. Individual judgments about quality of life are conditioned by experience and personality, but also by sociocultural norms. Those norms might include valuing independence, or the centrality of family relationships, and stand in for tradition in our analogy. In the case of temperature, our analogy shows that naïve conventionalism is an unworkable solution to the problem of coordination. We cannot stipulate an arbitrary definition for a physical quantity—coherence demands an appeal to empirical regularities in the real world.

## 8 Commonalities and differences between temperature and health-related quality of life

Using case studies in temperature and health-related quality of life, I have tried to show that the dynamics of measure development in both the physical sciences and the human sciences are hermeneutical. Furthermore, I have argued that Hans Georg Gadamer's (1989) hermeneutics are a better model for the dynamics of measure development than earlier hermeneutic philosophies. In this section, I will demonstrate an important

difference between measure development in the physical sciences and the human sciences, and how that difference is rooted in Gadamer's hermeneutics. Namely, I will show that the meaning of temperature is standardizable, while the meaning of health-related quality of life is not (McClimans 2010a). As a consequence, temperature outcomes in a given situation can be made to converge around a single, determinate value, while health-related quality of life outcomes cannot. This difference is rooted in an incomplete analogy between temperature measurement and the hermeneutic task.

### 8.1 Standardization of meaning

In order to see measurement outcomes converge around a single value, we must standardize the meaning of the concept being measured. For measurement concepts, that meaning is jointly determined by theory and operation through a process of progressive coherentism. Thus, the meaning of temperature is determined in part by norms set forth by the ideal gas law, and also in part by functioning thermometers that operationalize the temperature standard (Chang 2004, Chapter 4). For a text, meaning is determined hermeneutically through genuine, truth-seeking dialogue resulting in a fusion of horizons. Tradition, subjectivity, and the text itself are brought together for the interpreter, who must make sense of the text before her (Warnke 1987, p. 74-75).

It is common for immature measures to lack grounding in theory. Measurement practices often fall into place before those practices are married with theory (Hacking 1983).<sup>2</sup> Chang's narrative about the development of the temperature standard illustrates

---

<sup>2</sup> Kuhn (1961) seems to argue the contrary—that we are often unable to make a successful measurement until theory tells us what outcomes to expect. Yet he distinguishes the situation where we are making measurements to justify a theory from the situation where we are still learning what quantity we are measuring. For van Fraassen, these two processes, taking a measurement and learning what we are measuring, are inevitably entangled and form the problem of coordination. I claim that solving this problem is analogous to Gadamer's hermeneutic task.

as much, as the numeric thermometer had been established well before Lord Kelvin anchored the temperature scale first to the heat engine and then to the ideal gas law and thereby set abstract norms for its meaning. While the thermodynamic temperature standard has slowly come to maturity over the course of centuries, PROMs are, by contrast, relative newcomers having appeared on the scene in the 1960s and 70s as part of the social indicators movement. Sonja Hunt (1997) has complained about the widespread proliferation of health-related quality of life measures in the absence of solid theoretical grounding, and her complaint remains valid twenty years later.

But the failure to ground health-related quality of life measurement in theory is not merely due to a lack of maturity. The meaning of quality of life varies both with researchers' purposes, and with a matrix of respondents' experiences, personality, and sociocultural traditions, all of which they bring to bear when they encounter the instrument. If we adhere to McClimans's (2010a) commitment to take unexpected interpretations of PROM questions seriously, we open the door to operational variance as different respondents may, in effect, be answering different questions from one another. This variance is not without bound but is still problematic for those who would try to nail down a standardized meaning for health-related quality of life (McClimans 2010b). McClimans (2010a) sees this variance in interpretation and conceptualization as a perennial factor in health-related quality of life measurement. While researchers and certain cohorts of respondents may reach a fusion of horizons regarding the meaning of health-related quality of life, that shared understanding will not be universal, and it will not be finally determinate any more than the meaning of a historical text. Instead, as Anna Alexandrova (2017, Chapter 1) has argued, we should see the meaning of well-being as varying with context. Indeed, this is why PROMs must be newly validated for

different populations of respondents or when they are put to work in different contexts (FDA 2009).

## 8.2 Convergence of measurement values

According to Chang (2004, p. 90), temperature measurement is regulated by the ontological principle of single value. That is, researchers are committed to finding or constructing a single value for temperature each time it is measured. Historically, however, different thermometers often gave different temperature readings from one another when used to measure the same body. These disparate outcomes were brought into accord with each other, or made to converge, by iteratively correcting each outcome and bringing different operationalizations into alignment with theory (Chang 2004, p. 211-217).

Convergence around a single value is a hard-won achievement, and not one that is guaranteed. Without a widely accepted theoretical model of the measurement system, without a determinate meaning for the measurement concept, it is difficult, though not impossible, to make the case for choosing a single outcome over others. Regnault, for instance, was able to endorse the outcomes provided by the air thermometer over those of the mercury and spirit thermometer by testing for consistency. He found that while different air thermometers typically agreed well with one another, different mercury and spirit thermometers did not. However, there was no guarantee that Regnault's strategy would be successful. The choice of thermometric fluid might well have been underdetermined by his findings (Chang 2004, p. 96). For this reason, theory has an important role to play in providing a normative standard.

For instance, consider Eran Tal's (2011) account of the construction of the standard second. The definition is an idealized one whose meaning is specified by a theoretical model of a cesium fountain clock. Because (1) that definition can never be perfectly realized by an actual clock, and (2) an abstract idealization alone does not have empirical content, metrologists bridge the gap between the two by de-idealizing the definition. De-idealization involves identifying sources of error and uncertainty and taking them into account. It allows researchers to correct the outcomes given by real cesium fountain clocks and bring them into alignment with the single value endorsed by the ideal definition (Tal 2011).

Yet as I have argued above, there is no widely accepted theoretical definition for health-related quality of life (Hunt 1997; McClimans 2010a). There is no ideal model to provide a norm for the correction of disparate measurement values. While not infeasible, respondents' subjective conceptualizations of health-related quality of life and their individual interpretations of PROM questions play an important role in shaping the meaning of health-related quality of life, and thus in its measurement (McClimans 2010a). When a measurement concept does not have a determinate meaning, or a theoretical definition, different operationalizations and different measurement values often remain permissible. I argue that this indeterminacy is an unavoidable feature of health-related quality of life measurement due to the nature of the construct. This is because much of the meaning of the construct is open to interpretation and is the proper subject of genuine questioning (McClimans 2010a).<sup>3</sup>

---

<sup>3</sup> Nancy Cartwright and Rosa Runhardt argue that health-related quality of life is a Ballung concept. Ballung concepts—such as poverty, disability, and quality of life—comprise a cluster of meanings, and do not admit of unique measurement values. Instead, they are often best represented by a table of values reflecting their multivocal status (Cartwright and Montuschi 2014).

Insofar as a single theoretical definition, or a single conceptual meaning for target constructs is possible in the physical sciences, measure development in the physical sciences seems, in fact, to diverge from the dynamics of the hermeneutic task. For Gadamer, there are always new questions to ask of a text, and new pre-understandings to bring to bear on that text as tradition evolves. For this reason, new and different understandings are always possible. The perspective brought to bear on a text by the historically and socially situated interpreter is an essential part of the fusion of horizons that uncovers the meaning of a text. Because that perspective is ever changing, that meaning is not determinate (Gadamer 1989, p. 362-370; Warnke 1987, p. 96-99).

While measure development in the physical sciences is in many ways analogous to the hermeneutic task, that analogy has limitations. The most important of these limitations is the determinacy of meaning for the measurement concept and the resultant convergence of measurement outcomes around a single value. PROM development represents a more complete analogy with the hermeneutic task, as the meaning of its target concept remains indeterminate. Indeed, the relationship goes beyond mere analogy. PROMs, according to McClimans (2010a), are texts whose subject matter is imperfectly understood and whose meaning is to be uncovered. These instruments are hermeneutic objects, and the researchers and respondents who interact with them are interpreters of meaning.

## 9 Conclusion

In this paper I have shown that the dynamics of measure development are analogous to the hermeneutic task as described by Hans Georg Gadamer. This is true not only for archetypal measures in the physical sciences, such as temperature measures, but also for

measures in the human sciences, such as PROMs. Chang's work on the development of the temperature standard, and McClimans's work on health-related quality of life measurement, illustrate that measure development is dialogical, is shaped by tradition, and is constrained by coherence. Yet this analogy has certain limitations, at least for physical measures. Insofar as a determinate and standardized meaning for measurement concepts is achievable, the analogy with Gadamer's hermeneutics is broken. In this way, the development of health-related quality of life measures bears a closer resemblance to the hermeneutic task than the development of the temperature standard.

## References

- Alexandrova, A. (2017). *A Philosophy for the Science of Well-being*. Oxford: Oxford University Press.
- Cartwright, N. and Montuschi, E. (2014). *Philosophy of Social Science: A New Introduction*. Oxford: Oxford University Press.
- Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.
- Food and Drug Administration. (2009). Guidance for industry on patient-reported outcome measures: Use in medicinal product development to support labeling claims. *Federal Register*, 74: 1–43.
- Gadamer, H. G. ([1960]1989). *Truth and Method*, 2<sup>nd</sup> Revised Ed. Trans. J. Weinsheimer & D.G. Marshall. New York: Crossroad Publishing.
- Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.
- Hadorn D. C., Sorensen, J., and Holte, J.. (1995). Large-scale health outcomes evaluation: How should quality of life be measured? Part II: Questionnaire validation in a cohort of patients with advanced cancer. *Journal of Clinical Epidemiology*, 48: 619-29.

- Hunt, S. (1997). The problem of health-related quality of life. *Health-related quality of life Research*, 6: 205-212.
- Kowalski C. J., Bernheim, J. L., Birk, N. A. and Theuns, P. (2012). Felicitometric hermeneutics: Interpreting quality of life measurements. *Theoretical Medicine and Bioethics* 33: 207-20.
- Kuhn, T. (1961). The Function of Measurement in Modern Physical Science. *Isis*, 52(2): 161-193.
- McClimans, L. (2010a). A Theoretical framework for patient-reported outcome measures. *Theoretical Medicine and Bioethics*, 31: 225-240.
- McClimans, L. (2010b). Towards self-determination in health-related quality of life research: A Dialogic Approach. *Medicine Health Care and Philosophy*, 13: 67-76.
- Pusic, A., Klassen, A. F., Scott, A. M., Klok, J. A., Cordeiro, P. G., and Cano, S. J.. (2009). "Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q." *Plastic and Reconstructive Surgery* 124(2): 345-353.
- Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of health-related quality of life assessment in light of response shift and appraisal. *Health and Health-related quality of life Outcomes*, 2: 14.
- Tal, E. (2011). How accurate is the standard second? *Philosophy of Science*, 78(5): 1082-1096.
- Van Fraassen, B. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.
- Warnke, G. (1987). *Gadamer: Hermeneutics, Tradition and Reason*. Stanford: Stanford University Press.