Cognition 191 (2019) 103942

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

The career of measurement

Kensy Cooperrider*, Dedre Gentner

Department of Psychology, Northwestern University, 2029 Sheridan Road, Chicago, IL 60637, United States

ARTICLE INFO

Keywords:
Measurement
Abstraction
Units
Comparison
Analogy
Cognitive tools

ABSTRACT

Units as they exist today are highly abstract. *Meters*, *miles*, and other modern measures have no obvious basis in tangible phenomena and can be applied broadly across domains. Historical examples suggest, however, that units have not always been so abstract. Here, we examine this issue systematically. We begin by analyzing linear measures in the Oxford English Dictionary (OED) and in an ethnographic database that spans 114 cultures (HRAF). Our survey of both datasets shows, first, that early length units have mostly come from concrete sources—body parts, artifacts, events, and other tangible phenomena—and, second, that they have often been tied to particular contexts. Measurement units have thus undergone a shift from highly concrete to highly abstract. How did this shift happen? Drawing on historical surveys and case studies—as well as data from the OED and HRAF—we next propose a reconstruction of how abstract units might have evolved gradually through a series of overlapping stages. We also consider the cognitive processes that underpin this evolution—in particular, comparison. Finally, we discuss the cognitive origins of units. Units are not only slow to emerge historically, they are also slow to be acquired developmentally, and mastering them appears to have cognitive consequences. Taken together, these observations suggest that units are not inevitable intuitions, but are best thought of as culturally evolved cognitive tools. By analyzing the career of measurement in detail, we illustrate how such tools—abstract as they are today—can arise from concrete, often bodily origins.

1. Introduction

Measurement permeates and supports much of modern life. Transportation, manufacturing, commerce, science, and other endeavors rely critically on the ability to accurately quantify properties of the world—such as length, weight, temperature, and time—and to communicate those properties to others. Fundamental to modern measurement are our current systems of units, in particular the metric and imperial systems. The units in these systems are precisely defined and broadly shared, owing to more than two centuries of efforts to refine them and spread them globally (Alder, 2002; Astin, 1968; Crease, 2011). They are also highly abstract in at least two senses: (1) most have no obvious tie to concrete phenomena; and (2) they are of broad scope, applicable across contexts and domains. For example, the meter is now officially defined, not with reference to the body or any tangible object, but with reference to the distance that light travels in a fraction of a second. Moreover, along with its derivatives (e.g., centimeters, kilometers), it can be used broadly, for measuring the height of a skyscraper or the width of pencil lead. The ubiquity of such abstract units—and the fluency with which we use them—can make them seem like an inevitable part of human understanding.

But evidence suggests that measurement units have not always been so abstract. Historians have noted that, for centuries in England and across Europe, units of length were often concrete and context-specific (e.g., Kula, 1986; Whitelaw, 2007). Many were based on artifacts such as bows, chain links, and goads for driving oxen, as well as on spans of the body such as finger-widths, hand-breadths, and arm-lengths. (The *foot*, with its concrete etymology, is a vestige of this earlier era.) Such units were often limited in application, with certain ones favored for measuring cloth, others for horses, others for land, and with different measures sometimes used for length, depth, height, and distance. The extent of such concrete grounding remains to be investigated more systematically, particularly outside of British and European contexts, but observations like these suggest a striking shift in the nature of measurement, from highly concrete to highly abstract. Our aim here is to trace this shift and investigate its cognitive underpinnings and consequences.

Many scholars have noted that such a shift has taken place, citing both historical observations (e.g., Crease, 2011; Crosby, 1997; Kula, 1986) and cross-cultural variability within modern times (e.g., Best, 1918; Crump, 1990; Hallowell, 1942). Several of these accounts also highlight important transitions as measurement systems evolve. For

* Corresponding author.
E-mail addresses: kensy.cooperrider@northwestern.edu (K. Cooperrider), gentner@northwestern.edu (D. Gentner).

https://doi.org/10.1016/j.cognition.2019.04.011
Received 12 October 2018; Received in revised form 10 April 2019; Accepted 12 April 2019
Available online 11 July 2019

instance, Kula (1986) states that standardization—the shift from measuring with "*your* foot" to measuring with "*the* foot"—marks an "intellectual turning point" in the "the transition from concrete to abstract concepts" (p. 24). Hallowell (1942, p. 69) remarks on the "inestimable importance from a psychological point of view" of the introduction of measuring tools like rulers. Best (1918, p. 190) notes the importance of systematization—that is, the relation of units to each other—in scientific measurement. Here we synthesize these and other insights to offer an account of the evolution of measurement in its linguistic, practical, and cognitive aspects. We take inspiration from stage-wise accounts of the evolution of other abstract conceptual systems, especially numeral systems (e.g., Epps, 2006; Hurford, 1987; Schmandt-Besserat, 1996; Wiese, 2007).

To preview, we propose that the evolution of measurement involves four overlapping stages. First, people make ad hoc comparisons between one concrete thing—a *target* to be measured (e.g., a log)—and another—a *comparator* (e.g., someone's foot). Such comparisons are often driven by communicative or practical needs. Second, people come to favor certain comparators over others—in short, conventions emerge. Third, people abstract across these conventional comparators (e.g., many examples of people's feet) to develop an idealized standard comparator (e.g., *the* foot). Only at this stage does measurement involve concepts that resemble our modern notion of abstract units. Fourth, people begin to create systems of units—to define units in terms of each other and nest them hierarchically. Ultimately, such systematization extends across scales (e.g., a foot and a mile) and orientations (e.g., distance and depth). Across these stages there are key changes to the linguistic, practical, and cognitive aspects of measurement—that is, changes to how people talk about units and measurement, the tools and practices they use to measure, and, we suggest, how they conceive of units and measurement. Our investigation focuses on length and other forms of linear extent, such as height, depth, and distance. However, many of our proposals should apply to other physical dimensions, such as weight and volume.

A focus of our proposal is the cognitive processes that underpin the evolution of abstract units. In particular, we posit a central role for comparison. The idea that measurement involves a comparison between a to-be-measured object and something else (i.e., a *comparator*, such as a ruler) is widely accepted in current accounts (Crease, 2011; Crump, 1990; Hallowell, 1942). Going beyond these accounts, we suggest that comparison enters into the evolution of units in two ways. First, we show that, over the evolution of measurement, what changes is the nature of the comparators used. Second, we argue that comparison is a critical driver of the transitions from one stage to another. Our proposal thus fits with the broader idea that comparison can foster the emergence of abstract knowledge (Gentner & Hoyos, 2017; Gentner & Medina, 1998). This idea is already well supported at the level of the individual. Prompting learners to compare examples helps them arrive at more abstract, general understandings (e.g., Alfieri, Nokes-Malach, & Schunn, 2013; Gentner, 2010; Gick & Holyoak, 1983; Kurtz, Miao, & Gentner, 2001). For example, being asked to describe the similarities between two stories leads people to recognize a schema common to both stories and promotes the use of this same schema in later problem solving (Gick & Holyoak, 1983). But the importance of comparison in the formation of abstract knowledge is also evident at the level of cultural history. Many of our abstract concepts started out as novel figurative comparisons, which gradually became conventional and then entered the lexicon (Bowdle & Gentner, 2005; Xu, Malt, & Srinivasan, 2017). The word *blockbuster*, for example, originally referred to a type of aerial bomb that could destroy an entire block; but many English speakers today will only recognize its more abstract use to refer to anything of great popularity or importance (Bowdle & Gentner, 2005, p. 209). We hypothesize that the same trajectory is evident in the case of measurement: Novel comparisons get the process started; but over repeated comparisons, abstractions emerge that can be used without knowing where they started.

Our plan is as follows. In Section 2 we present historical and ethnographic evidence that units of length were once concrete in the two senses highlighted earlier: (a) tied to tangible phenomena and (b) used in particular contexts. Historians have offered suggestive examples of such concreteness, but here we systematically examine the evidence, first in English (Section 2.1) and then across a broad range of cultures (Section 2.2). In Section 3 we offer an account of how length units may have evolved from their concrete origins into the highly abstract systems they are today. To reconstruct this evolution, we draw on historical surveys and case studies, in addition to the OED and HRAF data, and we consider the cognitive processes involved. In Section 4 we discuss the cognitive origins of units. Based on our historical and ethnographic survey and on additional developmental evidence, we argue that the idea of unitizing physical dimensions is not a natural intuition but a culturally evolved one; it is an idea that was slow to develop historically and is slow to be learned by children. But it is also an idea that, once mastered, has clear consequences on both individual and cultural-historical levels. In this sense, units may be considered consummate cognitive tools.

2. Concreteness in length measurement

The examples already presented suggest that early measurement units in England and across Europe were often highly concrete—that is, transparently tied to tangible phenomena and restricted in their application. Here, we systematically assess how common this kind of concreteness was, beginning with an analysis of length units in English.

2.1. Length measurement in English

To assess the concreteness of length units in English, we examined all the linear measures in the Oxford English Dictionary (OED) (http://www.oed.com/). To identify these measures, we searched for words within the category of 'Measurement' that had the terms 'length' or 'linear measure' in their definitions. This search returned 124 words, from which 73 words were determined to be units (others were related to measurement but were not length units per se). A further 20 of these units were excluded because they were described as belonging to another language (e.g., *remen*; Egyptian), and five further entries did not offer an etymology for the unit (e.g., *lug*). The data set thus includes 48 English linear measures with known origins (available at: https://osf.io/znqtu/).

The first sense in which units may be concrete is by being based on tangible phenomena. To examine this for the OED measures, we examined the etymological information provided for each entry; we then classified the sources of these words. Concrete sources included the human body, artifacts, and other tangible phenomena such as seeds. We discuss these in turn.

The most frequent concrete source for length units in English is the human body, accounting for 38% of the words (18 of 48). Beyond the familiar 'foot,' these terms include the 'fathom' (the length between the outstretched arms), the 'ell' (the full length of one arm), and the 'cubit' (the span from the elbow to the fingertips). Shorter spans include the 'hand,' 'palm,' and terms based on both finger length and finger breadth (Table 1).

Length terms derived from artifacts account for another 29% (14 of 48). These include 'yard' (originally a type of pole), and other terms derived from elongated objects, such as 'rod,' 'perch,' and 'virgate.' Other artifact-based terms include 'bow,' 'chain,' 'link' (of a chain), and 'goad,' a tool used for driving draft animals.

Terms deriving from other concrete sources account for a further 19% (9 of 48). These include terms from the natural world—e.g., 'poppy seed,' 'barley-corn,' and 'reed'— as well as terms from agricultural contexts—e.g., 'furlong,' a compression of 'furrow' + 'long,' and 'ox-gang,' based on the amount of land a team of oxen could plow in a certain time.

Table 1
Examples of length units in English (OED).

Word	Source	Source type	Scale
poppy seed	poppy seed	other concrete	small
barley-corn	grain of barley	other concrete	small
digit	breadth of finger	human body	small
link	link of chain	artifact	small
palm	palm of the hand	human body	small
prime	from word for ‘first’	abstract (relational)	small
yard	type of spear	artifact	small
fathom	outstretched arms	human body	small
gad	rod for driving oxen	artifact	medium
reed	reed plant	other concrete	medium
perch	pole	artifact	medium
furrow	furrow in a field	other concrete	medium
mile	abbreviation of Latin for ‘thousand steps’	abstract (relational)	large

The remaining 15% (7 of 48) have an abstract origin. A unit was considered to have an abstract origin if its etymology: (1) was inherently relational—that is, derived from a subdivision or multiple of another unit; or (2) had no relation to a tangible phenomenon. Examples of inherently relational units include ‘inch,’ from the Latin for ‘twelfth,’ and ‘mile,’ from the Latin for ‘thousand’ (abbreviated from the phrase ‘mille passus,’ meaning ‘thousand paces’). The only example of an intangible unit is ‘meter,’ which derives from a Greek word for ‘measure.’

The majority of these length measures (33/48, or 69%) are *small* in scale (i.e., with an extent less than or equal to that of the human body). An additional 13 measures are *medium* in scale (i.e., with an extent between that of the human body and approximately 100 m), and the remaining two are *large* in scale (i.e., with an extent larger than approximately 100 m). The specific sources for these length units differ somewhat according to the size of the unit. For instance, artifact-based terms account for only 18% (6/33) of the small-scale units, but account for 62% (8/13) of the medium-scale units. (We expand on this point in the next section.) In sum, the vast majority of measurement terms in English have their origins in tangible phenomena.

The OED also provides evidence for the second sense in which early measurement terms were concrete: several of these units were primarily used in specific contexts, as mentioned in their definitions or inferable from examples of usage. ‘Bow’ was confined to archery; ‘chain’ and ‘prime’ were used chiefly in surveying; ‘furrow’ and ‘land’ were specific to agriculture; ‘nail’ was used primarily in measuring cloth. Context-specificity in English units may be much more pervasive than these few examples suggest. It is likely, for instance, that ‘step’ was primarily used for measurements on the horizontal plane, but the OED does not explicitly note this. Thus, while the evidence for this second sense of concreteness is less complete, there are enough examples to support the contention made by historians that early measurement terms and practices in the English-speaking world were often context-specific (e.g., Kula, 1986).

The etymology of English unit terms provides a valuable window into the history of measurement, but its scope is limited. For a broader understanding of the career of measurement, we need to go beyond world languages like English and beyond Anglo-European cultures (Lupyan & Dale, 2010; Majid & Levinson, 2010). Moreover, we also need to examine practices in addition to language. To address these limitations, we next look across cultures at measurement terms and practices that have been documented by ethnographers.

2.2. Length measurement across cultures

To broaden the scope of our investigation, we analyzed the Human Relations Area Files (HRAF) ‘World Cultures’ database (http://ehrafworldcultures.yale.edu/ehrafe/); this resource compiles and topic-codes ethnographic accounts from 311 cultures, with a focus on non-industrialized, small-scale groups. We searched the topic code ‘Weights and measures’ (HRAF topic code 804), yielding entries from 193 cultures spanning every geographic region. In all, observations about length measurement were available for 114 cultures. We extracted all mentions of length units—that is, linear extents described as conventionally used for measurement.[1] This resulted in a total of 352 length units from 84 different cultures. (We did not include data from 30 cultures in the HRAF, for which length measurement practices were described without identifying any units in particular.) We did not exclude possible borrowings, as this determination was not always possible. We note that many of these ethnographies are from first half of the 1900s or earlier (date range = 1764–2001); we thus refer to all HRAF data as historical, that is, as characterizing measurement as it *was* in different cultures, not necessarily as it still is. The full data set is available at: https://osf.io/znqtu/. We next describe the sources of these measures. There is more data than in the OED, allowing for more a detailed analysis, so we consider small-, medium-, and large-scale units in turn.

Small-scale units. In all, 278 small-scale units are reported in the HRAF (from 69 cultures in the database). They are overwhelmingly based on the body (95%, or 264/278) (Table 2). Of the 69 cultures in which small-scale units are reported, 66 had body-based units and 59 had *only* body-based units. Many groups had rich inventories of such units (range = 1–16). Best (1924) describes 12 small-scale body-based units used by the Maori, from one equivalent to the first joint of the thumb (*konui*) to one equivalent to the full length of the body when lying down with the arms extended above the head (*takato*). On Wolei atoll in the Caroline Islands, body-based units ranged from as short as a finger joint to as long as a fathom (Damm, 1938). The Tzeltal of Mexico had a series of units, from the *nab*, the span between the thumb-tip and end of middle finger when all fingers are extended, to the *yankabal*, the distance between the armpit and the fingertips of the opposite arm when outstretched (Villa Rojas, 1969).

Across these systems, certain spans of the body—in particular, salient divisions of the upper body and forelimbs—occur with especially high frequency: the cubit (reported in 22 cultures), the fathom (40 cultures), and variants of the hand-stretch (i.e., tip of thumb to tip of index, middle, or little finger when hand is stretched) (52 cultures) (Fig. 1). Aside from the common ‘foot’ and ‘pace,’ measures based on the lower extremities are rare. Although most small-scale measures are based on the body, there are also some units based on the natural world, such as a unit based on the ant, used by the Chagga of Africa (Marealle, 1963), or on the sesame seed, used in Burma (Scott, 1910).

Medium-scale units. In all, 24 medium-scale units are reported in the HRAF (from 15 cultures), making them much less widely observed than small-scale measures (see Table 3 for examples). The most common source of these units appears to be events (50%, or 12/24), specifically events that are punctate in nature. (This source type is not attested in the OED analysis.) Some event-based units are based on brief actions, including ‘bow shot’ (e.g., in the Andaman Islands; Man, 1932a) and ‘stone’s throw’ (e.g., in Morocco; Blanco Izaga, 1975). Others are based on sound, with measures derived from the distance at which one could still hear a person calling or a musket sound (in Burma; Scott, 1910). Other medium-scale units are derived from artifacts, such as a tool for cutting banana leaves (in Chagga; Marealle, 1963), or a lasso (in Saami; Itkonen, 1984); a few are derived from multiples of body-spans, such as a measure equivalent to 40 forearm-lengths in Ethiopia (Messing, 1985).

[1] It should be cautioned that not all of these units were units in the modern Western sense. Even when units were reported as such, researchers sometimes noted that they were “more expressive than informative” (Anderson, 1978, p. 543), or qualified them as neither “mathematical” (Richards, 1939, p. 204) nor “precise” (Best, 1918, p. 26).

Table 2
Examples of small-scale units across cultures (HRAF).

Source	Source type	Culture	Reference
Finger (breadth)	Human body	*several* (e.g., Bhil)	Naik (1956)
Finger joint (length)	Human body	*several* (e.g., Amhara)	Young (1972)
Hand (breadth)	Human body	*several* (e.g., Aymara)	La Barre (1948)
Hand span (thumb to little finger)	Human body	*several* (e.g., Zapotec)	González (2001)
Hand span (thumb to index finger)	Human body	*widespread* (e.g., Tlingit)	Emmons and De Laguna (1991)
Foot	Human body	*several* (e.g., Hopi)	Talayesva and Simmons (1942)
Cubit (i.e., elbow to fingertip)	Human body	*widespread* (e.g., Karen)	Marshall (1922)
Half-fathom (i.e., finger tip to sternum)	Human body	*widespread* (e.g., Trobriands)	Senft (1986)
Pace	Human body	*several* (e.g., Bambara)	Paques and Turner (1954)
Elbow to finger tip of opposite arm	Human body	Caroline Islands	Damm (1938)
Stretch of legs apart	Human body	Chagga	Marealle (1963)
Fathom (i.e., span of outstretched arms)	Human body	*widespread* (e.g., Trobriands)	Senft (1986)
Person (with arms extended above head)	Human body	Maori	Best (1918)

Note: In all tables, a unit is marked as occurring in 'several' cultures if it was attested in at least four groups, and as 'widespread' if it was attested in ten or more.

Fig. 1. A marble architectural relief, from the Aegean or Western Turkey, 460–450 BC. The relief depicts several body-based measures, including the fathom (the full extent of which is now missing), the foot (imprint above right upper arm), the fist (see inset in right forearm). The relief is believed to have been displayed above the door to a public office regulating weights and measures. Image © Ashmolean Museum, Oxford University (Image: ANMichaelis.83).

Table 3
Examples of medium-scale units (HRAF).

Source	Source type	Culture	Reference
Bamboo	Other concrete	Bhil	Naik (1956)
Lasso	Artifact	Saami	Itkonen (1984)
Ten fathoms	Abstract (relational)	Maori	Best (1924)
Field	Other concrete	Bhil	Naik (1956)
Stone throw	Event (punctate)	*several* (e.g., Karen)	Marshall (1922)
Bow shot	Event (punctate)	Kogi	Reichel-Dolmatoff (1949–1950)

Large-scale units. In all, 50 large-scale units are reported in the HRAF (from 31 cultures). These are most often based on protracted events (60%, or 30/50, of large-scale measures reported) (Table 4). In 10 cultures in the database, large distances were reckoned in terms of days spent traveling. (The same principle motivates the contemporary astronomical term 'light-year'.) Another common device was to measure the distance of journeys in terms of consumption habits—for example, by enumerating the number of betel nuts chewed (Karen of Southeast Asia; Marshall, 1922), coffee stops required (Saami; Itkonen, 1984), or young coconuts drunk en route (Nicobarese of the Pacific;

Table 4
Examples of large-scale units (HRAF).

Source	Source type	Culture	Reference
Day (distance covered in day of travel)	Event (protracted)	*widespread* (e.g., Saami)	Itkonen (1984)
Young coconut (distance covered while drinking)	Event (protracted)	Nicobarese	Man (1932b)
Pipe bowl (distance covered while smoking)	Event (protracted)	Ojibwe	Jenness (1935)
Coca leaf (distance covered while chewing)	Event (protracted)	Aymara	La Barre (1948)
Post (distance between administrative posts)	Other concrete	Burma	Scott (1910)
Stream-crossing (distance between stream crossings)	Other concrete	Ovimbundu	Ennis (1962)
Wolf day (distance covered by wolf in day)	Event (protracted)	Saami	Itkonen (1984)
Reindeer day (distance covered by reindeer in day)	Event (protracted)	Saami	Itkonen (1984)
Sleeps (number of nights spent on journey)	Other concrete	*several* (e.g., Ojibwe)	Jenness (1935)

Table 5
Examples of context-specific measurement practices (HRAF).

Target	Comparator	Culture	Reference
String money	Tattoos on forearm	Yurok	Kroeber (1925)
Buffalo horns	Forearm	Toraja	Nooy-Palm (1979)
Canoes	Various body measures	Chugach	Birket-Smith (1953)
Trees (girth)	Arms	Maori	Best (1924)
Flutes	Fingers	Kogi	Reichel-Dolmatoff (1949–1950)
Pigs (girth)	Forearm	Siwai	Oliver (1955)

Man, 1932b). Others measured distance by counting salient land features, such as capes (Mi'kmaq of North America; Le Clercq, 1910) or marshes (Bembu of Africa; Richards, 1939). At least one culture, the Ojibwe of North America, found a way to use the body to measure large-scale distances. This was done by superimposing the outstretched hand on the arc of the sun: one 'hand-stretch' was considered one fourth of the arc from sunrise to zenith, and could thus be used to convey how much of a day it would take to travel the target distance (Jenness, 1935).

Many measurement practices in traditional societies are described as confined to particular contexts (Table 5). These practices often center on culturally important activities—such as planting seedlings, building houses, allocating meat, and measuring currency—or manufacturing practices—such as making canoes, nets, coffins, arrows, and instruments. The Toraja of Indonesia, for instance, had a conventional set of points on the arm used for measuring buffalo horns, an important commodity (Nooy-Palm, 1979). The Siwai of Papua New Guinea had a conventional system for measuring the girth of pigs (Oliver, 1955). A practice for measuring string money among the Yurok in California sometimes involved tattooing measurement landmarks on the arm (Kroeber, 1925).

3. Reconstructing the career of measurement

The foregoing survey shows that length units around the world have most often been drawn from concrete sources and that measurement units and practices have often been tied to particular contexts. This contrasts with units currently used in the industrialized world, which mostly have no obvious tie to concrete sources and are quite general in application. Here, we offer an account of this apparently radical shift by reconstructing four key stages in the evolution of measurement. Our reconstruction draws on in-depth studies of measurement within particular cultures (e.g., Alkire, 1970; Best, 1918; Hallowell, 1942; Pankhurst, 1969), on general overviews of the history of measurement (e.g., Alder, 2002; Crease, 2011; Crosby, 1997; Kula, 1986), and on data from the OED and HRAF.

A key feature of our account is that it elaborates on the general observation that the basis of measurement is comparison (Crease, 2011; Crump, 1990). Most basically, the hands-on activity of measuring something—of aligning an object against a ruler or other tool—entails a comparison process. But we argue that the role of comparison in measurement also goes deeper. First, changes to the nature of measurement can be viewed as changes to the kinds of comparators involved. Second, more speculatively, we suggest that comparison processes drive changes to the kinds of comparators used. In short, comparison is both a window into the abstraction process and an important engine of that process.

We propose a series of four stages. First, people make ad hoc comparisons between concrete objects; second, certain concrete comparators become conventionally used; third, some of these conventional comparators become standardized; fourth, standardized comparators become interrelated with each other, forming systems of units. Critically, these stages are overlapping in a few ways. Within a given culture, some units may be abstract, while others remain concrete and context-bound. Also, the processes of standardization and systematization co-occur and interact. Finally, even after a culture has developed systems of abstract units, ad hoc comparisons will continue to be used in some contexts.

3.1. Stage 1: ad hoc comparison

In our account, the comparisons that enter into the early stages of measurement are *ad hoc*. There is not direct historical evidence of such ad hoc comparisons, but their existence can be inferred from considering contexts in which people engage in ad hoc measurement even today. One such context is when people want to communicate the length of a non-present target, such as a fish that got away. To do this, a person often invokes a point of comparison—a *comparator*—whose length is more accessible ("It was the size of a pig").[2] Man (1932a) describes the importance of such comparisons among the Andamanese, in the Bay of Bengal:

> "In referring to the size, shape, or weight of a small object, they would, if possible, liken it to some seed… or fruit, such as mangosteen, jackfruit, or cocoanut; of larger weights they would say, "as much as" or "more than one could carry" or "lift;" for expressing capacity or quantity they would say "a bucketful," "basketful," "handful," "canoe-load," as the case might be." (p. 256)

Such ad hoc comparisons are often made using language alone, but they can also be done by anchoring the comparison to a present comparator—e.g., "It was as big as this table"—or to a concurrent demonstration—e.g., "It was *this* big," accompanied by a size gesture. Gesture regularly enters into such comparisons, and gestural conventions for ad hoc comparisons of size have been widely reported (e.g., in Nuer: Huffman, 1931; in Mesoamerica: Fox Tree, 2009).

Beyond communication, ad hoc comparison may also have been useful when trying to judge or remember length. For example, consider the utility of comparison when trying to determine which of two targets—call them A and B—is longer (see Hallowell, 1942, for discussion). This can be done by eye when the difference is marked, or by directly juxtaposing the targets when this is possible. However, when the difference is more subtle, or when A and B cannot be directly juxtaposed, the judgment requires a new solution (Hallowell, 1942). For example, suppose you want to know which of two spatially separated trees has a thicker trunk. Assuming "eyeballing" is unsatisfactory, a solution is to introduce a comparator—a third object that can be directly juxtaposed with each target. This comparator could be a body part, tool, or something improvised on the spot to match one of the two targets. Such techniques are widely described in the ethnographic literature and have involved banana fibers (Chagga; Marealle, 1963), string (Gikuyi; Davison, 1996), sticks (Kaska; Honigmann, 1949), plant stalks (Semang; Schebesta, 1954), and vines (Fiji; Thompson, 1940).

[2] Length comparisons are often relative and qualitative (e.g., "This log is longer than that one"), but such comparisons are not measurement in a strict sense (Hallowell, 1942). The impulse to measure is an impulse to express, more or less precisely, *exactly* how long a target is.

Critically, all these materials can be tailored to match a target extent. Similar techniques are still used in the Western world. In the game of bocce, for instance, to decide which of two balls is closer to the pallino, a string can be used to first mark the distance of one ball, creating a comparator that matches one target distance; this is then compared to the second target distance.

We hypothesize that ad hoc comparisons most often occur between a target and a comparator of equal length, rather than between a target that is shorter or larger than the comparator. We suggest that equal comparisons may be more cognitively accessible because the target and comparator are an exact (or close) match on the feature of interest. Further, equal comparisons are cognitively easier in that they require no further computation. By contrast, unequal comparisons require either subdividing the comparator (A is one-third of C) or iterating the comparator (A is three C's in length). Consistent with this hypothesis, when judging length, children spontaneously perform equal comparisons before unequal comparisons, as discussed later (Piaget, Inhelder, & Szeminska, 1960).

At this stage, comparators are not units proper, they are simply objects or tangible phenomena recruited on the spot for measurement purposes. People do not think of these objects as having a dedicated measurement function, nor do the words for these objects have any separate, measurement-related meaning.

3.2. Stage 2: conventionalization

Over time, some comparators that were initially used on an ad hoc basis become conventional—that is, people begin to use them on a routine basis within a community. While we know of no deterministic account of which length comparators become conventional, some factors seem likely to be involved. These factors include: *availability* (i.e., how readily available the comparator is—a barley-corn would never catch on outside of a farming community), *juxtaposability* (how readily the comparator can be physically aligned with the target—the human ear is readily available, but is not easy to align with a target), and *aptness of scale* (the foot is available and alignable, but would be an impractical extent for measuring distance between towns) (see Crease, 2011, p. 18–21 for discussion; Kula, 1986). The universal use of body-based spans for measuring small-scale extents is perhaps explainable in terms of these factors: the body is always available and portable, readily alignable (certain spans more than others), and apt for small-scale extents.

We suggest that conventionalization is initially gradual. As a given ad hoc term begins to be more widely used, it is likely to become more general in application (e.g., Bybee, 2003) and more cognitively available (e.g., Segui, Mehler, Frauenfelder, & Morton, 1982). These changes reinforce each other: as terms are used more frequently, they come to mind more often and in a wider set of contexts; in turn, they become used even more frequently. We hypothesize another key change: as a comparator becomes more frequent, people become more likely to use it in *unequal* comparisons. As discussed earlier, unequal comparisons are cognitively taxing, requiring either proportional reasoning (e.g., half the length of foot) or counting (e.g., the length of three feet). However, as conventional comparators become more cognitively accessible, people may find it natural to use them in unequal comparisons, despite these costs. Indeed, some researchers have occasionally noted that, in a given community, only certain comparators may enter felicitously into unequal comparisons. In describing body-based units in the Caroline Islands, Alkire (1970) notes that, while one could speak of "two forearm-lengths" or "two hand-spans," informants rejected the same construction with other spans, such as 'palm-width.' Instead, one had to say something like "two spans of palm-width size." 'Palm-width' could only be used to describe "the exact length of an object" (p. 29). Our interpretation of this is that 'forearm-length' and 'hand-span' were more conventionalized as length comparators, and so could enter into a grammatical construction specialized for unequal comparisons; other spans, such as 'palm-width' still retained an ad hoc character, and so could not.

More generally, changes in conventionalization are often reflected in grammatical constructions. This is has been shown in the case of figurative comparisons, such as metaphors and similes. Bowdle and Gentner (2005) demonstrated what they called a "grammatical concordance principle": people strongly favor the basic comparison construction "X is like a Y" (X = target, Y = base) for novel metaphorical bases (comparators), but generally prefer to use the category membership construction "X is a Y" for conventional bases (see also Gentner & Bowdle, 2001). We suggest a similar change may be at play in the case of units. Specifically, we propose that, for unequal comparisons, speakers may go from favoring the basic property comparison construction "X is as long as two Ys" (X = target, Y = comparator) to favoring the basic unit construction ("X is two Ys long") as conventionalization proceeds.

Given that most common units in English are centuries old, such a pattern may be hard to observe directly. However, the trajectory might be observable in phrases like *football field*, which has recently gained ground as a size comparator (Morris, 2017), and which now seems to be emerging as an informal unit of length. An analysis of this phrase in the Google Ngram corpus[3] shows that it was first used (around 1930) in equal comparisons using the basic property comparison construction ("long as a football field") (see: https://osf.io/znqtu/) (Fig. 2). For example, in a *National Geographic* article from 1947, a natural arch in Utah is described as being as "long as a football field." Only later, around 1938, did football field come to be used in unequal comparisons ("long as two/three/four football fields"). And it was not until over a decade later, around 1951, that "football field" was first used in the unit construction ("two/three/four football fields long"). As an example of this last usage, someone in a *National Geographic* article from 1971 describes an oil tanker as "nearly four football fields long." Since around 1990, "football field" has been more often used in the unit construction than in the basic property comparison construction. This case illustrates the gradual way in which comparators become conventional and exemplifies how this conventionalization process may be reflected in subtle linguistic patterns.

3.3. Stage 3: standardization

Once comparators such as foot and finger have become conventional, people encounter the problem that not all instances of those comparators are the same length. The historical and ethnographic record is full of evidence of people recognizing—and trying to work around—this problem. At one point in China, a distinction was made between units based on the male hand and the female hand (Crease, 2011). Similarly, the Mapuche of South America adjusted the *wima*—a measure based on half of a fathom—according to whether the unit was used by a woman or a man (Hilger, 1957, p. 92). Elsewhere, people have taken advantage of the fundamental imprecision of body-based measures. In Ethiopia, it was common to bring to the market a person with "long arms" to help one measure purchases (Pankhurst, 1969, p. 36).

A solution to the imprecision problem is to develop an idealized version of the conventional comparator, or *standard*. Although we

[3] Some have noted problems with using the Google Ngram database to make inferences about the popularity of terms over time (e.g., Pechenick, Danforth, & Dodds, 2015). A particular issue is that the corpus is increasingly dominated by scientific literature through the 1900s. However, the rise of informal phrases like "long as football field" is unlikely to be due to a rise in scientific texts.

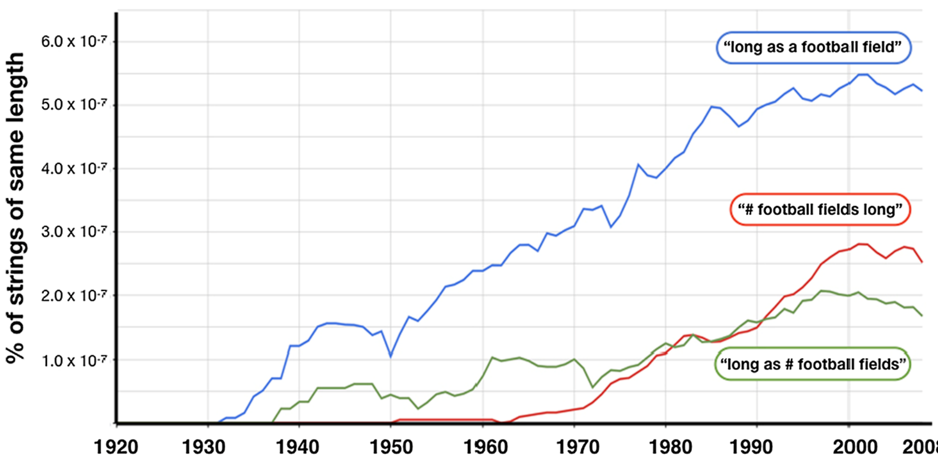

Fig. 2. An analysis of "football field" used as a length comparator from the Google Ngram database (1920–2008). The blue line tracks use of the phrase in the basic property comparison construction used for equal comparisons ("long as a football field"), which first appeared in the corpus (i.e., registered more than 40 tokens) in 1932. The green line tracks use of the phrase in the same construction but for unequal comparisons ("long as two/three/four football fields"), a usage first registered in 1938. The red line tracks its use in the unit construction for unequal comparisons ("two/three/four football fields long"), first registered in 1951. The y-axis shows the percentage, out strings of the same length (e.g., out of all 5-word strings), that each string of interest accounts for. Data is continuously averaged over an eleven-year window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

suspect the issue of imprecision is widely recognized in traditional societies (e.g., Saxe & Moylan, 1982), few small-scale cultures appear to have introduced standards. The HRAF database provides evidence of a handful of cases. Among the Maori, it was common to enshrine one man's arm-span in a wooden rod, called a *rauru*; such rods would be used throughout house-building projects and would sometimes be passed down through generations (Best, 1924; see also Best, 1918). The Tzeltal used a measuring rod known as a *jaubtic*, corresponding to a fathom (Villa Rojas, 1969). The Sinhalese used a "carpenter's rule" known as a *vadu riyana*, based on the cubit (Leach, 1961).

We hypothesize that standards become more common with intensive commerce and industrialization. British history offers many examples. Such standards could be based on known concrete lengths. For example, Henry I of England (1100–1135) standardized the yard as the length between his nose and the thumb of his outstretched arm (Macey, 1989). Standards could also be representative of a class of comparators. For instance, King David I of Scotland (1124–1153), standardized the Scottish inch by averaging the thumbs of a small, medium, and large-sized man (Macey, 1989). Such standards could then be made from wood, stone, or metal. Wooden "cubit rods" were common in Ancient Egypt and represented the standard cubit as well as other units (Scott, 1942). King Edward I (1272–1307) introduced the "Iron Ulna" as a standard measure (Whitelaw, 2007). Importantly, we suggest, the adoption of a standard precipitates a key change in how the comparator will come to be conceived. Decoupled from bodily spans or everyday objects, the comparator becomes something significantly more abstract and more recognizable as a unit (Hallowell, 1942). However—at least at first—the link to the original concrete source remains clear; the unit retains its approximate length and concrete label.

We suggest that standardization also promotes further abstraction. For instance, we speculate that it is at this stage that the comparator becomes more broadly used across contexts. Part of the reason for this is simply that a physical standard may be easier to align with a target than its concrete source. For example, one's actual foot is not easy to use to measure height or depth, but a rod that is the length of an idealized foot is. Further, it is only at this stage that it makes sense for people to propose new definitions of the standard, which may have little to do with its original basis. For example, the unit may be defined in terms of other units, as we discuss in the next section. Other extensions also become possible. Once the foot was not only an appendage but also an idea, it made sense for John Locke to propose the "philosophical foot" and for others to introduce the "hour foot" (Anstey, 2016). That the foot could exist in these different versions suggests it had become an abstraction—a standard that could be redefined and extended. Moreover, once the idea of standards is established, one can invent entirely new units, abstract from birth. The meter, first proposed in 1670, is the preeminent example. The idea of the meter—that is, the idea of a new basic unit of length—was widely discussed before its physical basis or even its label had been decided upon (see, e.g., Alder, 2002, pp. 87–96).

3.4. Stage four: systematization

After the notion of a standard is established, the stage is set for further changes. One such change is *systematization*. This occurs when people begin to compare standard units to each other—abstractions to abstractions—and thereby create an interconnected system of units. This may first happen within the same dimension and scale, when people begin to define units as nested within other units. Although most small-scale societies seem not to have developed measurement *systems*, as exemplified by the metric or imperial system, some had inventories of units with pockets of systematicity. Out of the 352 measures extracted from the HRAF, only 15 (from 10 different cultures) were inherently relational, that is, based on multiples or divisions of another unit.[4] For instance, the Maori had a unit—the *kumi*—which was equivalent to 10 fathoms and which an ethnographer described as "the first step toward a scientific system of measurement" (Best, 1924, p. 190); the Amhara had a unit based on 40 elbow-lengths (Messing, 1985), which was enshrined in a particular rope. In other cases, units retained their grounding in tangible phenomena while also being understood as part of a system. In Burma, the smallest unit was the 'hair's-breadth'; ten of these equaled a 'sesame seed'; six 'sesame seeds' equaled one 'barley-corn'; four 'barley-corns' equaled a fingers-breadth; and so on (Scott, 1910). The Saami developed set of three inter-related distance measures: a day's journey for a human; a day's journey for a reindeer, which was said to be ten times as long; and a day's journey for a wolf, which was said to be a hundred times as long (Itkonen, 1984). Thus, while there is evidence from a range of small-scale societies for some degree of systematization, most conventional units in such societies seem not to have been either standardized or systematized.

The process of systematization overlaps with the process of standardization. As discussed earlier, the two may interact, as when a unit is standardized by defining it with respect to a system. For example, a German treatise on surveying and geometry from 1522 described a practical procedure for simultaneously determining a standard 'foot'

[4] Not included in this count are cases of the "half-fathom" (found in 17 cultures). Because the fathom is naturally divisible in at the center of the chest, it is unclear whether this unit is motivated by a drive to systematize or merely by a salient anatomical span. Multiples of finger-breadth (e.g., three fingers wide) were excluded for the same reason.

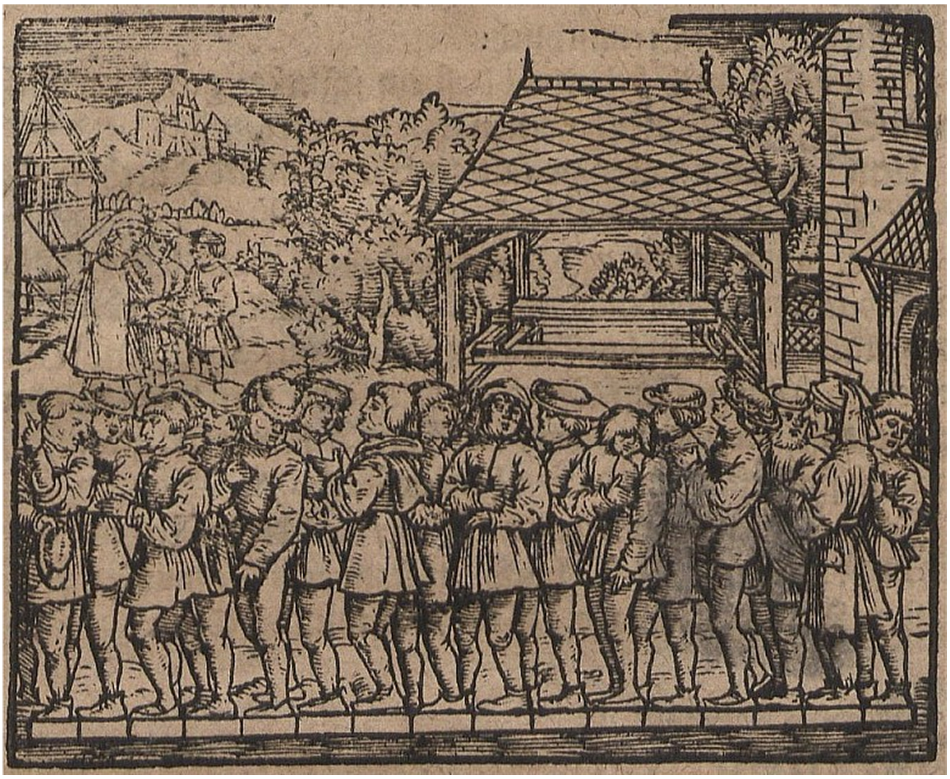

Fig. 3. An illustration from *Geomtrei* (1608 edition), by Jacob Köbel, of a procedure for determining a "lawful foot" and, at the same time, a "lawful rood" (a measure equivalent to 16 feet). He describes having sixteen men "of different sizes," upon leaving church, arrange their left feet in a row. According to Köbel, the resulting full length will be a standard rood, and one 16th of it will be a standard foot.

Table 6
Summary of stages.

Stage	Linguistic aspects	Practical aspects	Cognitive aspects
Ad hoc comparison			
Everyday objects are used as ad hoc comparators	No independent sense of comparator separate from everyday object. Basic comparison constructions are favored	Ad hoc comparisons also made using gestures, props, and found, easy-to-customize materials	No clear distinction between cognitive representation of objects as comparators and objects as objects
Conventionalization			
Certain classes of objects become conventionally used as comparators	Emerging sense of comparator as separate from everyday object (i.e., polysemy). Increasing use of unit construction instead of comparison construction	As comparators become more frequent, they are more likely to be used in unequal comparisons	Conventional comparators are frequently used and become more cognitively accessible, offsetting the difficulty of unequal comparisons
Standardization			
Conventional comparators are standardized. Often occurs alongside systematization	Terms for standardized comparators become fully polysemous, with a clear unit sense in addition to their object sense. The link between the two senses often remains clear	Standards are enshrined physically (metal, wood, stone), and copies of these standards are created and circulated	Some comparators now act as standard units (though they may also continue to be seen as concrete objects)
Systematization			
Standardized comparators are nested within each other, both within and across orientations. Often occurs alongside standardization	Terms for units may or may not retain an etymological connection to their original source. Some directly reference the broader system. People may not know the etymology of units	Measurement tools (e.g., rulers, tape) are designed and labeled with reference to the broader system	Units are understood as part of broader systems. It is not necessary to know the original concrete sources of the units because units can be understood relationally

(1/16th of a rood) and its cousin, the standard 'rood' (a measurement originally based on the length of a rod) (Fig. 3). Although systematization may begin by linking units along the same orientation and at the same scale (as in the German example), it eventually links units more broadly. Instead of merely comparing two different units of similar extent, it becomes possible to compare length units at fundamentally different scales (e.g., 'foot' and 'mile'), and orientations (units of distance and units of depth), and so on (e.g., Hallowell, 1942). The process depends on comparison, of course, but also on proportional and hierarchical reasoning. The end result of this process of systematization is a coherent set of units that can be understood as a relational system. Indeed, modern dictionaries define many units, not in terms of their basis in concrete, observable phenomena, but in terms of *other units*. The OED defines the 'inch' as "the twelfth part of a foot." Once such a system is in place, people can use units without an exact idea of their concrete grounding, as long as they can relate at least some of the units to sizes in the world.

3.5. Summary

The four stages in the career of measurement just described involve changes in how measurement units are used and understood in language and everyday life (see Table 6). Units continue to be refined longer after the stages of standardization and systematization are reached, but such developments are beyond the scope of our paper (see Alder, 2002; Astin, 1968; Crease, 2011). Also, as noted earlier, these are not historical stages that cultures pass through and then leave behind. Ad hoc comparisons, for instance, remain widely observable today, even in cultures with fully abstract unit systems. Nor are these stages that every new unit must inevitably pass through. Once the later stages have been reached for at least some units and dimensions, people begin to recognize abstractions such as 'unit' and 'system of units'; accordingly, new units can be introduced that bypass the earlier stages altogether.

Our account has focused on the case of length, but we believe it will extend to other dimensions such as weight, volume, or area. There is abundant evidence from the HRAF, for instance, that early units for such dimensions were tied to concrete phenomena. The Chagga had a variety of volume measures based on pots, gourds, and other vessels and specialized for measuring milk, butter, and honey (Marealle, 1963); Bedouins in Libya measured land area in camel-days, i.e., the area that a camel could plow in a day (Behnke, 1980). Indeed, historians of measurement have often noted that the qualitative shift from concrete to abstract units is evident across dimensions (see e.g., Crease, 2011; Kula, 1986). In short, we do not have any reason to suspect that the career of measurement differs substantially across different dimensions.

4. Cognitive origins of measurement: the cognitive tool account

As we have shown, over the course of history, measurement units have developed gradually and somewhat unevenly. Several cultural traditions—not only in Europe but also, for instance, in China (Crease, 2011) and in Ancient Egypt (Scott, 1942)—have developed elegant systems of abstract units. Across these traditions the evolution of measurement proceeded at different paces. For example, decimal measurement existed in China for four centuries before it was proposed in the West in the 16th century (Crease, 2011, p. 82). In many smaller-scale societies, conventional length units appear never to have been fully standardized or systematized; in others, conventional length units appear to have been confined to particular uses. (Of course, it is possible that other small-scale societies developed such systems, but that these are not well documented in the historical or ethnographic record.) This slow and spotty emergence invites questions about the cognitive underpinnings of measurement. In particular, it suggests that the very idea of units—not just specific unit systems—may have been culturally evolved. This is what we will refer to as the "cognitive tool" account of units (see, e.g., Gentner, 2010; Norman, 1993). On this view, units—much like the alphabet (O'Connor, 1996), numbers (Ifrah, 1985), cardinal direction systems (Brown, 1983), map-making techniques (Uttal, 2000), and the abacus (Srinivasan, Wagner, Frank, & Barner, 2018)—are tools that have to be developed.

Other accounts are possible, of course. It could be that humans come pre-equipped with the idea of units—that is, the intuition that the physical world is composed of distinct dimensions and that these dimensions are divisible into quanta. On this view, the gradual emergence of particular systems of measurement is merely a matter of people slowly converging on conventional means for packaging and communicating their antecedent concepts. And, moreover, the unevenness with which measurement systems have emerged in different places may reflect the fact that our antecedent ideas about units are only expressed given the right cultural pressures. This might be termed the "natural intuition" account of units.

It is hard to adjudicate between the cognitive tool and natural intuition accounts based on historical evidence alone, but other kinds of evidence bear on these proposals. In particular, the "cognitive tool" account entails two corollaries that the "natural intuition" account does not. If units are not natural intuitions but culturally evolved cognitive tools then: first, they may be difficult for children to learn; and, second, learning to wield such tools may have cognitive consequences, both for individuals and cultures. We now examine whether the evidence supports these corollaries.

4.1. Developmental acquisition of measurement

At least two lines of research suggest that measurement concepts do not come easily to young children. A first is studies by Piaget and colleagues on "spontaneous measurement" in young children (Piaget et al., 1960). Children were asked to build a tower the same height as a model tower, with smaller blocks on a lower table. Then they were asked to check whether the towers were the same height. (In some conditions a screen was placed between the two tables.) Objects such as sticks and paper strips were available for use. Piaget et al. reported several stages in children's behavior. In Stage I, until the age of 4 or 5 years old, children simply made qualitative visual comparisons. In Stage II (4; 6 to 7), realizing that visual comparison was inadequate, some children took the route of rebuilding their towers to be next to the model. When they did use a separate comparator, it was generally based on their own body. Some children used their arm to measure their tower; others placed their two hands at the top and bottom of their tower and tried to hold this gesture as they walked to the other tower. In Stage III (7 and older), children used external comparators such as sticks. At first children only considered such a stick useful if it was the same length as one of the targets (i.e., one that afforded an equal comparison). Later, they were willing to use a comparator longer than the target, marking the height of the tower with their hand. Children initially resisted using a shorter comparator, but by roughly 8 years of age, they could use a shorter comparator stepwise to measure the tower. Follow-up studies on spontaneous measurement find that children can be induced to use a comparator at younger ages if the inadequacy of visual comparison is more obvious (Bryant & Kopytynska, 1976), or when the task is couched in a particular, motivating context (Miller, 1989). Interestingly, children induced to measure in one context will not necessarily spontaneously measure later, in a superficially different context. For example, Bryant and Kopytynska (1976) found that a majority of 5–6 year-olds spontaneously used a stick to measure the depth of a hole (73% of children overall, ranging from 35% to 95% across five task variants). 120 of these children were also given a miniature version of the original Piagetian tower-measuring task, both before and after completing the hole task; not one child spontaneously measured in this context.

These findings underscore the slow developmental acquisition of measurement. Parallels between children's development and cultural evolution should be viewed with caution. Nonetheless, it is interesting that bodily comparators—the first comparators used by children in Piaget et al.'s study—are the most frequent type of comparator in small-scale measurement in the HRAF data (Table 2). Likewise, Bryant and Kopytynska (1976) finding that children would use a stick to measure depth but not height resonates with the historical evidence that measurement is often first tied to specific contexts.

A second line of research suggests that children also have trouble mastering conventional measurement practices, such as the use of rulers (Kellman & Massey, 2013; Lehrer, 2003; Solomon et al., 2015; Szilágyi, Clements, & Sarama, 2013). By age six, children in the US are readily able to measure a target object (e.g., a crayon) when the base of a ruler and the base of the target are aligned. However, when the bases are shifted (e.g., the base of the crayon is aligned with the 2-inch mark rather than the base)—even 2nd graders perform quite poorly (Solomon et al., 2015; see also Congdon, Kwon, & Levine, 2018). A common error in these shifted problems is to count the hash marks rather than the spaces. (This yields an answer one more than the correct response.) Solomon et al. (2015) suggest that children have difficulty conceptualizing continuous spatial intervals as countable, despite readily seeing discrete entities in this way. As in Piaget et al.'s work, the idea of concatenating small units of space to measure a large extent seems to be particularly challenging.

In sum, measurement understanding is slow to emerge in development, much as it has been over the course of history. In both cases, it may begin in specific contexts and proceed through a series of stages. Of course, there are major differences between children's course of learning and the historical evolution of measurement. Children in western societies inherit a highly systematic, elegant system of length units and measurement practices. This makes it all the more striking that these children have considerable problems grasping these practices—and even grasping the very notion that spatial extents are divisible into units. Such observations are consistent with first corollary of the cognitive tool account—that the development of measurement understanding involves a process of constructing new cognitive representations. An interesting parallel can be drawn to the development of number understanding. English-speaking children inherit a highly systematic set of number words, but it takes a long time for them to perceive this systematicity (Carey, 2004; see also Gentner, 2016; Mix, Sandhofer, & Baroody, 2005).

4.2. *Cognitive consequences of acquiring length units*

The second corollary of the cognitive tool account is that mastering units has cognitive consequences. There may be a number of such consequences, but here we consider only two. First, acquiring length units leads to fluency with a particular system. For example, Americans tend to find inches and yards easier to use than centimeters and meters, despite the advantages of the metric system. This sense of fluently thinking in a particular system involves, first, representing the connections between units (e.g., one yard = three feet = 36 in.) and, second, representing at least some correspondences between those units and the world (e.g., that wall is two feet high). Developmental research suggests that acquiring a firm sense of external correspondences involves developing internal standards that can be used as comparators (Duffy, Huttenlocher, Levine, & Duffy, 2005; Vasilyeva, Duffy, & Huttenlocher, 2007). For example, in one study, children were shown a wooden dowel inside a clear container (Duffy, Huttenlocher, & Levine, 2005). They were later asked to select which of two dowels was the same size as the target. When the size of the containers was changed from study to test, 8-year-olds correctly chose the dowel with the same height as the target, suggesting that they could use an internalized standard as a comparator. But 4-year-olds chose the dowel whose size was the same relative to its container, suggesting that they were reliant on the container as an external comparator.

A second consequence of mastering length units is that it lays conceptual groundwork for measuring further aspects of the world. At the cultural-historical level, certain basic dimensions of the world—notably length, volume, area, and weight—have been measured for thousands of years, at least in some cultures (Friberg, 1984; Morley & Renfrew, 2010). But other dimensions have become unitized only recently. Part of the reason for this is differences in the available science and technology: for example, specialized knowledge and instruments were needed before people could measure electric current (in amperes) or radioactivity (in Strontium units). But beyond this, we suggest, the key conceptual tools of measurement—for instance, an abstract notion of a unit, of a hierarchically organized system of units, and even of a measurable dimension—must first be developed for relatively accessible aspects of the world before they can be applied to less accessible aspects. Once these notions become established, they can be applied by analogy to new physical dimensions, such as the heat of peppers (measured in Scoville units) or the bitterness of beer (measured in International Bitterness Units). Going further, the idea of measurement can also be applied to nonphysical dimensions—such as mortality risk (in micromorts), the strength of chess positions (in centipawns), and intelligence (in I.Q. points). Of course, some question the notion that intelligence or other abstract constructs can be measured on one-dimensional scales. In some cases, it seems indisputable that the idea has been extended too far: for example, 14th century scholars at Oxford proposed measuring the constructs of certitude, virtue, and grace (Crosby, 1997, p. 14). Our modern profusion of units and of measurable dimensions would probably not have been possible without the conceptual groundwork that was laid initially for basic, accessible dimensions like length.

5. Conclusions

Units as we now know them—highly abstract, relationally understood, broadly shared, and precisely defined—have not always been that way. Rather, they started out concrete—tied to tangible phenomena and often confined to particular practices. Notably, many units across eras and cultures have been drawn from the human body. The career of measurement thus provides a powerful illustration of the role of the body in supporting abstract thinking (Bender & Beller, 2012; Gibbs, 2005; Lakoff & Núñez, 2000), and of the historical derivation of many abstract concepts from body-part terms (Heine, 1997; Ifrah, 1985). Just as remarkable as the fact that units were once concrete and often embodied is the fact that they managed to escape these humble origins. Our goal here has been both to document these origins and to show how such a transition may have happened. Our account thus complements a number of recent accounts of how abstract ideas evolve from concrete beginnings (e.g., Gentner & Asmuth, 2017; Jamrozik, McQuire, Cardillo, & Chatterjee, 2016; Xu et al., 2017).

People in modern, industrialized societies are so accustomed to parsing the world in terms of quantified dimensions that it is tempting to see the idea of abstract units as self-evident. The evidence reviewed here suggests, on the contrary, that measurement units do not come easily, either in history or in child development. They are thus best considered products of cultural evolution. In important respects, measurement units are analogous to other powerful abstractions, such as numbers (Frank, Everett, Fedorenko, & Gibson, 2008; Gordon, 2004), spatial prepositions (Gentner, Ozyürek, Gürcanli, & Goldin-Meadow, 2013; Heine, 1997), cardinal direction terms (Brown, 1983), and maps (Uttal, 2000). Like these other concepts, units have decidedly down-to-earth origins, but have now become so abstract and so ubiquitous that it is easy to take them for granted and to forget they have a history at all. And, like these other abstractions, measurement units may be considered "cognitive tools" (Gentner, 2003; Miller, 1989; Norman, 1993), with potentially far-reaching cognitive consequences for the individuals and cultures that wield them.

Declaration of interests

None.

Funding

This work was supported by the Office of Naval of Research (ONR N00014-16-1-2613 to Gentner) and by the NSF Spatial Intelligence and Learning Center (SBE 0541957).

References

Alder, K. (2002). *The measure of all things: The seven-year odyssey and hidden error that transformed the world.* New York: Simon & Schuster.
Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113.
Alkire, W. H. (1970). Systems of measurement on Woleai atoll, Caroline Islands. *Anthropos, 65*(1970), 1–73.
Anderson, M. (1978). *Saami ethnoecology: Resource management in Norwegian Lapland.* Ann Arbor, MI: University Microfilms International.
Anstey, P. R. (2016). Locke on measurement. *Studies in History and Philosophy of Science Part A, 60*, 70–81.
Astin, A. V. (1968). Standards of measurement. *Scientific American, 218*(6), 50–63.
Behnke, R. H. (1980). *Herders of Cyrenaica: Ecology, economy and kinship among the bedouin of Eastern Libya.* Urbana, IL: University of Illinois Press.
Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition, 124*(2), 156–182.
Best, E. (1918). The Maori system of measurement. *The New Zealand Journal of Science and Technology, 1*, 26–32.
Best, E. (1924). *Maori, Vol. 2*. Wellington, N.Z.: H.H. Tombs.
Birket-Smith, K. (1953). *Chugach Eskimo.* Kobenhavn: Nationalmuseets publikationsfond.

Blanco Izaga, E. (1975). *Emilio Blanco Izaga: Colonel in the Rif. Hraflex Books ethnology series.* New Haven, CT: Human Relation Area Files.
Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*(1), 193–216.
Brown, C. H. (1983). Where do cardinal direction terms come from? *Anthropological Linguistics, 25*(2), 121–161.
Bryant, P. E., & Kopytynska, H. (1976). Spontaneous measurement by young children. *Nature, 260*(29 April), 773.
Bybee, J. (2003). Mechanisms of change in grammaticalization: The role of frequency. In B. D. Joseph, & J. Janda (Eds.). *The handbook of historical linguistics* (pp. 602–623). Oxford: Blackwell.
Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus, 133*(1), 59–68.
Congdon, E. L., Kwon, M. K., & Levine, S. C. (2018). Learning to measure through action and gesture: Children's prior knowledge matters. *Cognition, 180*, 182–190.
Crease, R. (2011). *World in the balance: The historic quest for an absolute system of measurement.* New York: W.W. Norton.
Crosby, A. (1997). *The measure of reality: Quantification and western society, 1250–1600.* New York: Cambridge University Press.
Crump, T. (1990). *The anthropology of numbers.* New York: Cambridge University Press.
Damm, H. (1938). *Central Carolines. Part II: Ifaluk, Aurepik, Faraulip, Sorol, Mog-Mog.* Hamburg: Friederichsen, De Gruyter, and Co.
Davison, J. (1996). *Voices from Mutira: Change in the lives of rural Gikuyo Women, 1910–1995.* Boulder: Lynne Rienner Publishers.
Duffy, S., Huttenlocher, J., & Levine, S. (2005). It is all relative: How young children encode extent. *Journal of Cognition and Development, 6*(1), 51–63.
Duffy, S., Huttenlocher, J., Levine, S., & Duffy, R. (2005). How infants encode spatial extent. *Infancy, 8*(1), 81–90.
Emmons, G. T., & De Laguna, F. (1991). *Tlingit Indians. Anthropological papers of the American museum of natural history.* Seattle: University of Washington Press.
Ennis, M. W. (1962). *Umbundu: Folk tales from Angola.* Boston: Beacon Press.
Epps, P. (2006). Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica, 2*(0111550), 259–288.
Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108*(3), 819–824.
Friberg, J. (1984). Numbers and measures in the earliest written records. *Scientific American, 250*, 110–118.
Gentner, D. (2003). Why we're so smart. In D. Gentner, & S. Goldin-Meadow (Eds.). *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.
Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775.
Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist, 71*(8), 650–657.
Gentner, D., & Asmuth, J. (2017). Metaphoric extension, relational categories, and abstraction. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2017.1410560.
Gentner, D., & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor and Symbol, 16*(3&4), 223–247.
Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science, 9*, 672–693.
Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*, 263–297.
Gentner, D., Ozyürek, A., Gürcanli, O., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition, 127*(3), 318–330.
Gibbs, R. W. (2005). *Embodiment and cognitive science.* Cambridge: Cambridge University Press.
Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.
González, R. J. (2001). *Zapotec science: Farming and food in the Northern Sierra of Oaxaca.* Austin, TX: University of Texas Press.
Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496–499.
Hallowell, A. I. (1942). Some psychological aspects of measurement among the Saulteaux. *American Anthropologist, 44*(1), 62–77.
Heine, B. (1997). *Cognitive foundations of grammar.* Oxford University Press.
Hilger, M. I. (1957). *Araucanian child life and its cultural background.* Washington, DC: Smithsonian Institution.
Honigmann, J. J. (1949). *Culture and ethos of Kaska society.* New Haven, CT: Yale University Press.
Huffman, R. (1931). *Nuer customs and folklore.* London: International Institute of African Language and Culture.
Hurford, J. R. (1987). *Language and number: The emergence of a cognitive system.* Oxford: Basil Blackwell.
Ifrah, G. (1985). *From one to zero: A universal history of numbers.* New York: Viking.
Itkonen, T. (1984). *The Lapps in Finland up to 1945, Vol. 2.* Porvoo, Helsinki: Werner Söderström Osakeyhtiö.
Jamrozik, A., McQuire, M., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review, 23*(4), 1080–1089.
Jenness, D. (1935). *The Ojibwa Indians of Parry Island: Their social and religious life.* Ottawa: National Museum of Canada.
Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. *Psychology of Learning and Motivation, 58*, 117–165.
Kroeber, A. L. (1925). *Handbook of the Indians of California. Bulletin.* Washington: Government Printing Office.
Kula, W. (1986). *Measures and men.* Princeton, NJ: Princeton University Press.
Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*(4), 417–446.
La Barre, W. (1948). *Aymara Indians of the Lake Titicaca plateau. Memoirs.* Menasha, Wis.: American Anthropological Association.
Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York: Basic Books.
Le Clercq, C. (1910). *New relation of Gaspesia: With the customs and religion of the Gaspesian Indians.* Toronto: The Champlain Society.
Leach, E. R. (1961). *Pul Eliya, a village in Ceylon: A study of land tenure and Kinship.* Cambridge: Cambridge University Press.
Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin, & D. E. Schifter (Eds.). *A research companion to Principles and Standards for School Mathematics* (pp. 179–192). Reston, VA: National Council of Teachers of Mathematics.
Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS One, 5*(1), e8559. https://doi.org/10.1371/journal.pone.0008559.
Macey, S. (1989). *The dynamics of progress: Time, method, and measure.* Athens, GA: University of Georgia Press.
Majid, A., & Levinson, S. C. (2010). WEIRD languages have misled us, too. *Behavioral and Brain Sciences, 33*(2–3), 103.
Man, E. H. (1932b). *The Nicobar Islands and their people.* London: The Royal Anthropological Institute of Great Britain and Ireland.
Man, E. H. (1932a). *On the aboriginal inhabitants of the Andaman Islands.* London: The Royal Anthropological Institute of Great Britain and Ireland.
Marealle, P. (1963). Notes on Chagga customs. *Tanganyika Notes and Records, 60*, 67–90.
Marshall, H. I. (1922). *Karen people of Burma: A study in anthropology and ethnology.* Columbus: The University of Ohio Press.
Messing, S. D. (1985). *Highland Plateau Amhara of Ethiopia.* New Haven, CT: Human Relations Area Files.
Miller, K. F. (1989). Measurement as a tool for thought: The role of measuring procedures in children's understanding of quantitative invariance. *Developmental Psychology, 25*(4), 589–600.
Mix, K. S., Sandhofer, C. M., & Baroody, A. J. (2005). Number words and number concepts: The interplay of verbal and nonverbal processes in early quantitative development. In R. V. Kail (Vol. Ed.), *Advances in child development and behavior: Vol. 33*, (pp. 305–346). New York: Elsevier.
Morley, I., & Renfrew, C. (Eds.). (2010). *The archaeology of measurement: Comprehending heaven, earth, and time in ancient societies*. Cambridge: Cambridge University Press.
Morris, C. (2017). The size of things: An ngram experiment [Blog post]. Retrieved from <http://colinmorris.github.io/blog/size-of-things>.
Naik, T. B. (1956). *Bhils: A study.* Delhi: Bharatiya Adimjati Sevak Sangh.
Nooy-Palm, H. (1979). *The Sa'dan-Toraja: Organization, symbols and beliefs, VoL. 1.* University of Washington Press.
Norman, D. (1993). *Things that make us smart.* Boston, MA: Addison-Wesley.
O'Connor, M. (1996). The alphabet as a technology. In P. T. Daniels, & W. Bright (Eds.). *The world's writing systems* (pp. 787–794). New York: Oxford University Press.
Oliver, D. L. (1955). *Solomon Island society: Kinship and leadership among the Siuai of Bougainville.* Cambridge: Harvard University Press.
Pankhurst, R. (1969). A preliminary history of Ethiopian measures, weights, and values. *Journal of Ethiopian Studies, 7*(1), 31–54.
Paques, V., & Turner, T. (1954). *Bambara. Monographies Ethnologiques Africaines.* Paris: Presses Universitaires de France.
Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLoS ONE, 10*(10), 1–24. https://doi.org/10.1371/journal.pone.0137041.
Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry.* New York: W.W. Norton.
Reichel-Dolmatoff, G. (1949). *Kogi: A tribe of the Sierra Nevada de Santa Marta, Colombia. Revista, Vol. 1.* Bogota: El Instituto.
Richards, A. I. (1939). *Land, labour and diet in Northern Rhodesia: An economic study of the Bemba tribe.* Pub. for the International Institute of African Languages & Cultures by the Oxford University Press.
Saxe, G. B., & Moylan, T. (1982). The development of measurement operations among the Oksapmin of Papua New Guinea. *Child Development, 53*(5), 1242.
Schebesta, P. (1954). *Negritos of Asia. Ethnography of the Negritos: Half-Vol. 1, economy and sociology, Vol. 2.* Wien-Mödling: St. -Gabriel-Verlag.
Schmandt-Besserat, D. (1996). *How writing came about.* Austin, Texas: University of Texas Press.
Scott, J. G. (1910). *Burman: His life and notions.* London: Macmillan and Co.
Scott, N. E. (1942). Egyptian cubit rods. *The Metropolitan Museum of Art Bulletin, 1*(1), 70–75.
Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia, 20*(6), 615–627.
Senft, G. (1986). *Kilivila: The language of the Trobriand Islands.* Berlin: Mouton de Gruyter.
Solomon, T. L., Vasilyeva, M., Huttenlocher, J., Levine, S. C., Solomon, T. L., & Vasilyeva, M. (2015). Minding the gap: Children's difficulty conceptualizing spatial intervals as linear measurement units. *Developmental Psychology, 51*(11), 1564–1573.
Srinivasan, M., Wagner, K., Frank, M. C., & Barner, D. (2018). The role of design and training in artifact expertise: The case of abacus and visual attention. *Cognitive Science, 42*(S3), 757–782.
Szilágyi, J., Clements, D. H., & Sarama, J. (2013). Young children's understandings of length measurement: Evaluating a learning trajectory. *Journal for Research in Mathematics Education, 44*(3), 581–620.
Talayesva, D. C., & Simmons, L. W. (1942). *Sun chief: The autobiography of a Hopi Indian.* New Haven, CT: Pub. for the Institute of Human Relations by Yale University Press.

Thompson, L. (1940). *Southern Lau, Fiji: An Ethnography. Bulletin.* Honolulu, Hawaii: Bernice P. Bishop Museum.
Tree, E. F. (2009). Meemul Tziij: An indigenous sign language complex of Mesoamerica. *Sign Language Studies, 9*(3), 324–366.
Uttal, D. H. (2000). Seeing the big picture: Map use and the development of spatial cognition. *Developmental Science, 3*, 247–264. https://doi.org/10.1111/1467-7687.00119.
Vasilyeva, M., Duffy, S., & Huttenlocher, J. (2007). Developmental changes in the use of absolute and relative information: The case of spatial extent. *Journal of Cognition and Development, 8*(4), 455–471.
Villa Rojas, A. (1969). The Tzeltal. Ethnology, Part One. *Handbook of Middle American Indians, 7*, 195–225.
Whitelaw, I. (2007). *A measure of all things: The story of man and measurement.* New York: St. Martin's Press.
Wiese, H. (2007). The co-evolution of number concepts and counting words. *Lingua, 117*(5), 758–772.
Xu, Y., Malt, B. C., & Srinivasan, M. (2017). Evolution of word meanings through metaphorical mapping: Systematicity over the past millennium. *Cognitive Psychology, 96*(June), 41–53.
Young, A. L. (1972). *Medical beliefs and practices of Begemder Amhara.* Ann Arbor, Michigan: University Microfilms.