Journal of Institutional Economics (2023), **19**, 709–728
doi:10.1017/S1744137423000152

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Toward an economic theory of customary measurement

Glen Whitman

Department of Economics, California State University, Northridge, California, USA
Corresponding author. E-mail: glen.whitman@gmail.com

(Received 20 April 2022; revised 16 May 2023; accepted 17 May 2023; first published online 22 June 2023)

Abstract
The article proposes an economic theory of customary measurement systems. The form of such systems is driven by two transaction-cost factors: minimising costs of implementation, and coordinating on shared standards. These factors combine to yield seven principles of customary measurement, supported with illustrative examples from the traditional Anglo-American, Egyptian, Greek, Roman, Chinese, and Indian measurement systems. The theory illuminates various confounding features of such systems, including ubiquitous binary patterns, frequent appearance of duodecimal ratios, and persistence of trade-specific measures.

Keywords: binary; coordination games; measurement; transaction costs

Introduction

This article began with a T-shirt.

In 2007, I purchased a T-shirt with the Eye of Horus on the front (Figure 1a). I bought it only for its appearance, but afterward decided it would be wise to learn the symbol's meaning. Aside from its spiritual meanings, I was surprised to learn it had an economic meaning as well. In Ancient Egypt, the Eye of Horus consisted of six component parts, each of which corresponded to a fraction of the *hekat*, a measure of capacity. The fractions were ½, ¼, 1/8, 1/16, 1/32, and 1/64 (Figure 1b), which Egyptians would use to record quantities of grain bought and sold (Gyllenbok, 2018: 484).

This system is remarkable because it represents an application of *binary* occurring long before the Information Age, and despite the universality of nonbinary counting systems before modern times. But was this system unique, or part of a wider phenomenon? Consider a second example: the pirate-associated phrase 'doubloons and pieces of eight.' A doubloon, from the Spanish for *double*, was part of a currency sequence with denominations of one-half *escudo*, one *escudo*, two *escudos* (the doubloon), four, and eight (Hamilton, 1944: 22). 'Pieces of eight,' meanwhile, derives from the old practice of cutting coins into eight pieces – a practice also memorialised in the phrase 'two bits,' meaning a quarter of a dollar (Pieces of Eight, n.d.). Both of these exhibit binary patterns.[1]

One further example points to the potential significance of binary ratios in the analogue world. The US system of fluid capacity measure – whose ratios are perplexing to many modern users – is *almost entirely binary*, as shown in Table 1. The shaded diagonal shows that every pair of adjacent units has a ratio of either 2:1 or 4:1. Although US dry capacity and traditional UK capacity measures differ in some details, both exhibit essentially the same pattern.

If binary patterns are indeed common in customary measurement – as this article will demonstrate – the question is why. The practice of coin-cutting suggests a likely reason: when cutting something into pieces, halves are much simpler than other fractions. Anyone who has tried folding a sheet of paper into thirds (or fifths, or tenths) knows the principle at work here. Halving involves one simple

[1] I use 'binary' to refer to the integer powers of two and one-half.

Figure 1. The eye of Horus and its fractions.
Sources: Dahl (2007), Gyllenbok (2018: 484). In panel (b), each piece of the eye represents a fraction of the *hekat*, a unit of capacity.

Table 1. US Fluid capacity measures

Gallon									
2	**Pottle**								
4	2	**Quart**							
8	4	2	**Pint**						
16	8	4	2	**Cup**					
32	16	8	4	2	**Gill**				
64	32	16	8	4	2	**Jack**			
128	64	32	16	8	4	2	**Fl. Ounce**		
256	128	64	32	16	8	4	2	**Tablespoon**	
1,024	512	256	128	64	32	16	8	4	**Fl. Dram**

N. B.: This form of table is common in metrology. Each unit heading corresponds to both a column *and* a row. To find the ratio between two units, find the intersection of the larger unit's column and the smaller unit's row. For instance, a gallon is equal to eight pints.
Source: Gyllenbok (2018: 2,411).

comparison for equality of two quantities. For length, this can be done by folding or laying items side by side. For weight, it can be done with a double-arm scale. Volumes of a liquid can also be weighed; alternatively, with identical containers, equal volumes can be found by evening up fluid levels. Eyeballing is also easier with halves, as presumably happened with most coin-cutting. Doubling has advantages similar to halving, again because of the relative ease of establishing equality of two quantities.

These observations suggest that customary measurement systems, which often seem arbitrary and chaotic to modern eyes, may have possessed a hidden logic. This article's thesis is that customary weights and measures can be understood through an economic lens. Specifically, they are shaped by two factors related to transaction costs: the need to *minimise costs of implementation*, and the need to *coordinate on shared standards*. Taken together, these two factors explain many confounding aspects of such systems, including the ubiquity of binary patterns, the regular appearance of duodecimal (12:1) ratios, the limited use of decimal ratios, and the divergent measures employed in various trades.

The next section presents the theoretical framework, including the transaction-cost factors mentioned above. These concepts are then applied to arrive at seven predictions, or principles, that characterise the form of customary measurement systems. These principles are summarised in Table 2. For each principle, I show how it connects to the theoretical framework and how it interacts with the other principles, especially binary.

The section following the next connects the theory to related literature. The penultimate section offers illustrative evidence of these principles at work, often by reference to Anglo-American measures,

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

Table 2. Summary of customary measurement principles

Principle	When applicable	Supported by	Examples
Binary: units related by ratios in powers of two	When direct one-to-one comparisons are easy to make (e.g., balancing, levelling, folding)	Implementation	2 cups to the pint 2 pints to the quart 4 quarts to the gallon
Availability: units based on readily available objects and actions	When common objects and actions are similar in scale to the quantities measured	Implementation Coordination	Feet, hands, paces, rice grains, barleycorns
Comparability: units 'forced' into simple ratios with each other	When multiple availability-based units must be used together	Implementation Coordination	3 hands to the foot 3 feet to the yard
Divisibility: larger units that have multiple integer divisors	When there is frequent need for division into various numbers of equal parts	Implementation	12 inches to the foot 12 dozen to the gross
Counting: ratios that mirror those of counting systems	When counting substitutes for direct measurement, especially very large or very small quantities	Implementation	12 grains to the pennyweight 100 lbs to the hundredweight
Suitability: sizes/shapes corresponding to activities in which units are made or used	When production, consumption, and distribution generate natural sizes and shapes for units	Implementation Coordination	3 teaspoons to the tablespoon Trade-specific units: trusses, loads, sacks
Contact: unintuitive ratios from the meeting of trades/ regions	When high integration costs lead to the (partial or total) coexistence of disparate systems	Implementation Coordination	5.5 yards to the rod 14 pounds to the stone

but also the traditional Egyptian, Greek, Roman, Chinese, and Indian systems. The final section concludes with some concerns, caveats, and broader lessons the theory suggests.

Theoretical framework

Transaction costs: implementation and coordination

Transaction costs have been defined in various ways (Klaes, 2008), but Allen (2000) helpfully groups the definitions into two categories. The 'neoclassical' definition says transaction costs are 'the costs resulting from the transfer of property rights' (901). The 'property rights' definition says they are 'the costs [of] establishing and maintaining property rights' (898), including during transfer.

Measurement qualifies under either definition. Consistent with the neoclassical definition, measurement is part of the simple friction of transacting, as every transaction requires the parties to agree on terms of exchange. Measurement is thus one of the 'mundane transaction costs' associated with defining, counting, and paying for something transferred (Baldwin and Clark, 2002: 4).[2] But, consistent with the property-rights definition, measurement also influences and strengthens property rights, particularly on the margin of certainty, during exchange.

This article's theory is consistent with both definitions. The key point is that, other things equal, transactors have an incentive to minimise implementation costs, i.e., the practical costs of acquiring

[2]Baldwin & Clark (2002: 4) and Baldwin (2007) distinguish 'mundane' from 'opportunistic' transaction costs. This distinction does not map perfectly onto Allen's neoclassical/property-rights distinction, but it is similar; see Langlois (2006: 1,392).

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

and using measures. First, there is the obvious direct benefit of reducing any cost, provided nothing is lost by doing so. Second, when such costs vary with the number of transactions, reducing them enables a higher volume of exchange by shrinking the 'wedge' between buyer and seller prices. In this respect, measurement costs are functionally analogous to taxes or transportation costs (Allen, 2000: 902). *We should therefore expect people to employ measurement methods that achieve the same or similar transactions at lowest possible cost.*

The purpose of this article is *not* to explain why measurement happens at all. The theory takes as given that transactors will often, if not always, find measurement necessary to specify what is being transferred and to affect its certainty. Shared measures also facilitate comparison across vendors, and they ease intertemporal trade by allowing clearer specification of future quantities.[3] Swann (2009: 52), drawing upon Barzel (1982) and Akerlof (1970), further argues that measurement can help prevent market unravelling due to adverse selection. That, however, is all this paper will say on why measurement happens. The central question here is *why customary measurement systems take the forms they do*. The answer is that, under preindustrial economic conditions, measurement could be implemented in a lower-cost fashion if the system had certain features (such as binary ratios between units).

The problem may appear to be one of straightforward cost-minimisation, where the costs just happen to be a variety of transaction costs. In many respects, that is true. However, a firm or household cannot simply pick the measurement system that would minimise its *own* costs. To fulfil their purpose, measures must be shared. Efforts to establish and maintain shared measures also qualify as transaction costs (in the 'property rights' sense of that term). The process of selecting shared measures creates a collective action problem – specifically, a coordination game (Chuah and Hoffmann, 2003; Tirole, 1988: 408). This complicates matters in at least three ways.

First, coordination games have multiple equilibria. There are many possible measurement systems, but one must be chosen. Furthermore, some systems may be better than others. This makes it possible to get stuck in an inferior equilibrium, as emphasised in the network externalities literature (Arthur, 1989; David and Greenstein, 1990; Farrell and Saloner, 1985; Katz and Shapiro, 1985).[4] People may continue using an inferior system because everyone else does, as switching unilaterally is undesirable.

Second, coordination can happen at various levels: the market, the region, the nation, the trade or profession. Different groups may arrive at different 'local equilibria,' and later find they want to coordinate – e.g., between regions or across trades. This generates several related difficulties:

(a) Local equilibria naturally tend to be 'sticky.' Although the benefits of wider coordination may weaken the stickiness, there is still a question of which local equilibrium will prevail.
(b) People invest physical and human capital in their local equilibria, which creates further resistance to change. Transition costs will be asymmetric, with users of a discarded local equilibrium suffering more.
(c) Local equilibria will tend to reflect locally relevant factors such as implementation costs *within a specific trade*. If such an equilibrium is nevertheless discarded to achieve wider coordination, the losses to the losers may be ongoing, not merely transitional.
(d) In some cases, higher-level coordination may not in fact be efficient, even if some desire it. Sufficiently great transitional or ongoing losses can outweigh the benefits of wider coordination, thereby justifying the persistence of local equilibria.

Customary measuring systems will therefore reflect the influence of both minimising costs of implementation and coordinating on shared standards. In some cases, they will point in the same direction. But when they do not, customary measurement will tend to reflect a balance or tradeoff between the two.

[3] I thank an anonymous referee for this point.

[4] Leibowitz and Margolis (1990), however, use the QWERTY keyboard case to argue that this concern is overstated.

Principles of customary measurement

Taken together, implementation and coordination concerns yield several predictions – which I call *principles* – about the form of customary measurement. Table 2 summarises the main conclusions. These principles are not intended as universal truths. Rather, in Evans and Levinson's (2009: 437) taxonomy, they represent a combination of unrestricted and restricted tendencies: 'Most customary measurements systems will have property X' or 'Customary measurement systems with property X will tend to have property Y.' I did not derive these predictions through sheer deductive reasoning; rather, they were suggested by empirical patterns in real measurement systems (see the penultimate section) and the observations of historical metrologists.

Binary principle

The binary principle arises from implementation costs. Such costs may be fixed or variable. For a household or firm, the principal fixed costs are physical standards – i.e., weights, containers, and rulers of known and trusted provenance. In modern times, such standards are widely and cheaply available – but in preindustrial times, they were scarce, unreliable, and prone to deterioration (Zupko, 1990: 27). Given high fixed costs of standards, people would tend to acquire relatively few of them, relying instead on labour (a variable cost) to create divisions and multiples from the few they had.

However, variable costs of measurement could also be considerable. Dividing goods requires labour and time to assure equal subdivisions. Dividing a gallon into ten equal parts, for instance, would require many pairwise comparisons for equality. Binary ratios would reduce these costs; each halving involves one simple comparison. As indicated earlier, such a comparison could be performed by balancing, folding, levelling, and similar acts.

Furthermore, with one physical standard, a person could recreate all other units in a binary sequence with relative accuracy. Imagine a medieval merchant with only one standard capacity measure – say, a gallon. With a binary sequence like that in Table 1, he could reproduce any other unit in the sequence by repeatedly doubling or halving the one measure he had. Hence, binary patterns minimised the number of different devices that needed to be checked against a community's shared standard (such as one posted in the town square).[5] This system would have enabled people to convert specific variable costs into fixed costs *when justified by local circumstances* – such as for intermediate units that were used especially often. For instance, a tavern keeper might use their standard gallon to create many pint containers for beer. When they broke or deteriorated, they could be replaced by reference to the standard gallon.

Another advantage of binary sequences is their capacity to *fill the space of convenient and useful quantities*. A full six powers of two can fit within the range created by two powers of ten, yielding many more intermediate units. To illustrate, compare Anglo-American capacity to the metric system (which stands in here for a hypothetical decimal system, inasmuch as metric was not invented until the 1790s). One quart is approximately equal to one litre (1 qt ≈ 0.946 L), and a tablespoon approximately one centilitre (1 tbsp ≈ 1.5 cl). Between the litre and centilitre there is only one named unit, the decilitre, whereas between the quart and tablespoon we find the pint, cup, gill, jack, ounce, and half-ounce. All these intermediate quantities, useful in everyday life, could be produced with reasonable accuracy at low cost.

Binary sequences also may have entailed *lower cognitive costs*. Kula claims binary was convenient for uneducated people performing mental arithmetic (1986: 85). It is not instantly obvious why binary would make arithmetic easier, especially for people accustomed to decimal counting. But there is a subtle truth in Kula's argument: that binary lends itself to *behavioural algorithms that effectively mimic arithmetic calculations*. A worker might not remember that a gallon contains 32 gills. But she might remember that splitting a gallon five times yields a gill, or she might have a mnemonic device for recalling the sequence of unit names. Kula refers to such techniques as *mnemotechnics*

[5]Publicly posted standards were common in preindustrial times, particularly in wealthier times and places (Vincent, 2022: 75–76).

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

(1986: 85).[6] Mnemotechnics yield outcomes without the need for arithmetic; in this respect, mnemotechnics constitute routines assisted by artefacts (D'Adderio, 2011) and thus exemplify the notion of extended cognition (Clark and Chalmers, 1998).

These goals – producing accurate divisions, filling the space of useful measures, and reducing cognitive costs – are closely related. Any system, binary or decimal or otherwise, could in principle *express* intermediate values; e.g., 235 mL would be a unit similar to a US cup. However, ease of expression does not imply ease of *production*. Preindustrial people needed to produce multiple intermediate values with relative accuracy via simple behavioural algorithms, and that is where binary sequences had an advantage. Notice that the easiest intermediate units to produce in a decimal system would result from *implicit* binary, such as 0.5 and 0.25 of the base unit. It is not hard to imagine how such units would acquire names and outcompete harder-to-produce decimal alternatives, such as 0.6 or 0.2 base units.[7]

If binary was so advantageous, why wasn't it universal? Because, as the remaining principles will show, implementation costs and coordination needs varied across contexts.

Availability principle

The tendency of preindustrial people to use readily available objects, especially body parts, as measuring tools is widely acknowledged. This *availability principle* is supported by both implementation costs and coordination needs. It allowed people to measure with things they already possessed, rather than acquiring costly tools exclusively for the purpose. Furthermore, the same objects were possessed by nearly everyone, making them natural *focal points* (Schelling, 1960: 57) – i.e., intuitive solutions to coordination problems. They answered the question, 'What do *I* have that *everyone else* has, too?'

Body measures were convenient for relatively short lengths. Kula notes two other measurement categories arising from availability: First, measures associated with *actions*, such as the distance of a bowshot or the amount of labour a domesticated animal could perform in a day (Kula, 1986: 29). Second, measures based on commonly available *external objects*, such as a barleycorn or rice grain. The former were best for lengths much larger than the human body, the latter for lengths smaller than a human finger (Kula, 1986: 25). The latter were also useful for weight and capacity, which – barring dismemberment – could not easily be measured with body parts.

The Achilles' heel of available measures was variability. Not all feet are the same! Indeed, variance was a perpetual problem for most preindustrial units (Allen, 2012: 33). However, means existed to mitigate variability. For many body units, the same basic measure could be performed in multiple ways (Kula, 1986: 26). The cubit, for instance, could be measured from the elbow to the middle fingertip, or the index fingertip, or the first knuckle, and so on. This meant that, if some standard were publicly available, any given person could see how their 'personal' standard measured against it. A larger person could recreate the standard cubit using one of the shortening methods, while a smaller person might add a thumb's width.

Once a standard had emerged, it could provide the 'kernel' for a binary sequence: one-half the standard, one-quarter, etc. More than one standard might emerge from availability – such as a foot for short lengths (carpentry) and a pace for longer distances (land). These multiple standards could then provide the basis for multiple overlapping binary sequences.

Comparability principle

It's fine to measure length with either feet or paces, but sooner or later, someone will want to know their ratio, particularly when working on the same project (Watson, 1915: 25). This is a version of the coordination problem, inasmuch as different units could emerge as local equilibria for different

[6]Oddly, Kula does not use such mnemotechnics to explain binary sequences, despite discussing them in close proximity (1986: 83ff).

[7]Binary-like patterns sometimes appear even with modern metric, e.g., liquids offered in two-litre, one-litre, and half-litre quantities, presumably for similar reasons. I thank an anonymous referee for this point.

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

purposes, and it was easier to find a ratio between units than to abandon a useful unit. Furthermore, fixing the ratio between units could help to bind down the meaning of each one, thereby reducing variance (Allen, 2012: 33). The *comparability principle* says people tended to find such ratios.

But not just any ratio would do. To minimise cognitive costs, available measures had to be comparable without cumbersome calculations. This could be accomplished by forcing units into ratios 'with simple multiples and simple, fractionless divisors' (Kula, 1986: 26). In other words, ratios would typically be *unit ratios* – i.e., a ratio of one to an integer. Unit ratios had the added advantage of being natural focal points.[8]

Naturally occurring units don't in general have unit ratios. How did customary systems cope with this inconvenient fact? The simple answer is 'approximation.' But another method was available, identified by Rybakov and summarised by Kula (1986: 26–27). Again, many measures could be executed in slightly different ways. In a study of customary Russian measures, Rybakov found these slight adjustments were used to make ratios fit more accurately. If a fathom (outstretched arms) were measured in a longer way in a particular region, the ell (or European cubit) would be performed in a way that made it very close to one-quarter of this longer fathom. In short, small adjustments were used to preserve unit ratios, improving accuracy at minimal added cost.

Binary ratios are always unit ratios, but the reverse is not true. Thus, the binary and comparability principles could conflict. Binary ratios had the cost advantages discussed earlier. But comparability favoured the closest unit ratio, binary or otherwise, because it provided a more attractive focal point. Which would prevail? Notice that costliness of division rises with the number of nonbinary subunits. Binary division is simple; three-way division is harder but tolerable; five- or seven-way division is very difficult. Thus, when the most natural ratio was relatively small, such as 3:1, it could survive. But for larger natural ratios, binary would tend to prevail; for instance, 8:1 would tend to drive out 7:1.

Divisibility principle

There is an advantage to larger units having many divisors, thereby allowing many possible divisions into equal parts (Vincent, 2022: 263, 273). A unit of twelve equal parts, for example, can be divided into two, three, four, six, or twelve portions. This would have been helpful when something had to be split among a group – for instance, workers being paid in kind for a job, customers pooling funds for a purchase, or business partners dividing their gains. If the course of business created frequent need for some division, people would favour a system that accomplished it with minimal cognitive costs – i.e., without difficult fractions. The more frequent the need for a particular division, the more likely a measure facilitating it would be cost-justified.

However, as the binary principle makes clear, the fact that a unit *has* a given divisor does not mean it's easy to *effect* that division. Each additional divisor (other than twos) would have required either another physical standard or more labour spent on creating difficult subdivisions. As with comparability, therefore, the advantages of divisibility had to be weighed against binary. Consequently, there was a special advantage for duodecimal ratios (12:1, 24:1, etc.), which resulted from a single 3:1 ratio along with multiple 2:1 ratios, thereby accommodating many possible divisions while remaining *mostly* binary (and thus *mostly* low-cost).

Counting principle

We might expect a society's counting system to dictate its measurement system directly. But that is the puzzle we started with: that many customary systems did *not* mimic counting, relying instead of powers of two. Once we grasp the binary principle, the real question is why decimal ratios appeared as often as they did. A plausible answer is that in some circumstances *counting was a low-cost substitute for direct measurement*.

[8]Cf. Allen and Lueck (2009: 886 n20) and Young and Burke (2001), who present models in which information or bargaining costs yield simple fractions in agricultural contracts.

Consider a merchant with a single standard vessel for dry capacity. He regularly uses sequential halving to find smaller units. He may also use doubling to find larger units. But the latter is not as necessary as the former. Dividing a unit into ten parts is hard, but counting up ten units is easy. The merchant can simply measure units with his lone standard, stacking or loading them as he goes. Counting does not have the same problem with accuracy that division does.

Furthermore, binary's advantage of filling the space of convenient measures was less applicable for very large quantities, as even binary measures would have been widely spaced in that region. To achieve very large quantities via doubling, it would have been necessary to have additional costly equipment – extra-large scales, extra-large vessels, etc. – with minimal added benefit relative to counting. Therefore, cost minimisation predicts a category of deviations from binary at the upper end of measurement scales, where people would have tended to rely on counting instead.

This counting principle applies at the bottom of the scale as well, especially when small indivisible units such as grains were involved. Rather than starting with one tiny grain and doubling to reach every higher-valued unit, it would be less cumbersome to simply count a given number of grains to construct the next highest unit, after which doubling could take over. While filling more of the available measure space was an advantage for typical quantities, it became a disadvantage at the extreme low end of the scale, again giving the edge to counting.

Going a step further, counting systems are 'technological objects… devices for figuring things out and for tackling recurrent coordination problems' (Harper, 2010: 171). As such, they should be shaped by the same pragmatic concerns as direct measurement. Recall that many cultures have used non-decimal counting systems – including duodecimal, which may have arisen from using the thumb to count the segments of the other fingers; vigesimal, which may have arisen from using toes as well as fingers; and even sexagesimal, famously used in Babylon and Sumer (Macey, 2010: 90–92). The involvement of fingers and toes exemplifies the availability principle. Moreover, twelve, twenty, and sixty all have many divisors. With direct measurement, the advantages of binary could outweigh divisibility. But when counting substituted for direct measurement, divisibility concerns would prevail. We should therefore expect some counting measures to have used twelves and twenties rather than tens.

Furthermore, twelve and twenty have advantages of spatial configuration. Ten items can be arranged in a single line, or two lines of five each – relatively elongated shapes. But twelve can be arranged in three lines of four each, twenty in four lines of five each. These shorter-wider arrangements can be visually apprehended as blocks, and may also fit better in spaces such as storage rooms, carts, and cargo holds. These configurations would thus economise on cognitive costs while dovetailing with storage and transportation needs. As such, they illustrate again how artefact-assisted routines (D'Adderio, 2011) and extended cognition (Clark and Chalmers, 1998) can reduce cognitive burdens. (This process of 'fitting' measures to typical use also gives rise to the next principle.)

Notice that multiple principles combine to support duodecimal. Comparability naturally creates the occasional 3:1 ratio, which together with binary yields duodecimal sequences. Divisibility indicates that duodecimal ratios are practically useful, and the counting principle provides a set of cases where costs of divisibility are relatively low. Empirically, distinguishing these principles' separate contributions will be difficult because they reinforce one another. For instance, if not for the advantages of divisibility, some natural 3:1 ratios might have been displaced by 2:1 or 4:1. Moreover, allowing a single 3:1 ratio would diminish the marginal value of further nonbinary ratios.

Suitability principle

Suitability refers to measures having sizes and shapes appropriate to the activities in which they were used. This allowed measurement to piggyback on production, distribution, and consumption needs, thereby avoiding added costs of measuring.

At the production stage, suitable measures were driven by physical capital. As Kula observes, 'the width of the piece of cloth is determined by the width of the loom' and the size of a pane of glass 'by that of the milling equipment in the glassworks' (1986: 6). At the distribution stage, similar concerns yielded 'transport-determined measures' such as the basket and wagonload (Kula, 1986: 6). At the consumption stage, when goods were packaged in quantities suitable for household use, such quantities naturally doubled as measuring units.

While implementation costs are central here, coordination plays a key role. Suitability relates to *specific* industries and goods. The physical capital used in casting iron differs from that used in baling wool; consumers' desired quantities of wine differ from those for milk. Therefore, coordinative equilibria driven by suitability would tend to be highly local, i.e., specific to trades and products. Resistance to alternative measures would be significant, as switching to another system would involve either (a) changing physical capital to match the new standards, thus incurring transition costs and possibly ongoing costs from using less suitable equipment; or (b) maintaining existing capital while incurring added measurement costs. Therefore, customary measurement would tend to tolerate many trade-based local equilibria.

Finally, suitability should interact with binary. The advantages of matching units to production, distribution, and consumption needs could trump binary's advantages. At the same time, binary sequences could appear *within* a given trade, but with suitability-driven units as their bases.

Contact principle

The contact principle is driven by coordination concerns. People had good reason to keep local standards to maintain suitability and avoid transition costs. Yet the expansion of commerce, as well as cooperation across trades for joint projects or shared transport, brought differing systems into contact. What happened when they met?

Sometimes one system prevailed. Other times, strong incentives to maintain existing standards led to the accommodation of two systems side-by-side. This mechanism, I suggest, explains the most peculiar ratios between units. *Full* integration of competing systems produces intuitive unit ratios like 2:1, 3:1, and 10:1. *Partial* integration results in unintuitive ratios like 7:1 and 5.5:1. Finally, when systems are *not integrated at all*, we see ratios like 3.785411784:1 (litres to the gallon).

These patterns should correlate with the costliness of integration: the higher the cost, the lower the level of integration. Larger and better established systems, with more adherents and more capital devoted to them, would be more resistant to integration. This seems especially likely when coordination was required only at contained points in the process, such as border crossings, because there existed a lower-cost alternative to society-wide conversion: having merchants who specialised in making conversions, developing the skills needed for that purpose (Kula, 1986: 96).

Spontaneous order and the role of government

The theory presented is, in many respects, a spontaneous-order story. Coordination games do not require central direction to reach equilibrium. Players face significant incentives to converge on shared standards, especially when facing similar costs and benefits with built-in focal points. Even conceding the possibility of persistent suboptimal equilibria – per the network externalities literature – such equilibria will nevertheless tend to be *functional*. In this sense, the theory resembles spontaneous-order stories like Menger (1892) on the emergence of money and Demsetz (1967) on the evolution of private property rights.

The archaeological record supports the notion that measurement standards can arise without central control (Vincent, 2022: 54–55). Nevertheless, governments have been involved in promulgating measurement standards from time immemorial. Even measures that long preceded centralised governments were likely influenced by local authorities. What role have governments played in the processes described earlier?

In some respects, government actors' interests were aligned with those of private actors. They stood to share in the expanded commerce and improved living standards that would come from lower transaction costs. To that extent, governments could be expected to reinforce coordinating equilibria, encouraging higher-level coordination if and only if its benefits exceeded whatever losses it imposed

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

on users of local equilibria. For example, governments may have favoured simpler and more widely shared measures because they increased the efficiency of contract enforcement, which would (or could) redound to the benefit of the governed.[9]

However, government involvement needn't always have been salutary. State actors had an interest in simplifying tax collection, minimising tax avoidance, and extracting added revenue for the ruling class – something they could do by requiring differing units (Kula, 1986: 55–58), even if these were otherwise inconvenient. Some state actors may also have wished to rationalise measures according to an abstract scheme that seemed more logically consistent and harmonious. Ironically, governments trying to foster uniformity often inadvertently contributed to the proliferation of standards (Zupko, 1990: 8).

A key factor easing the difficulty of accounting for state involvement is that it was so rarely effective. Preindustrial governments often lacked the means to enforce their metrological designs. Kula documents failure after failure in European measurement reforms (1986: 16). Standardisation efforts foundered due to poorly written laws, scarcity of physical standards, and difficulty of gaining local officials' cooperation (Zupko, 1990: 26–28). Consequently, 'Local populations grew accustomed to ignoring government directives' (1990: 8).

State interventions seem to have been most successful when they codified existing measures rather than overriding them (see, e.g., Kula [1986: 111] and Owen [1966: 129]). Reforms were more likely to gain traction when they kept the existing system's essential features while clearing away the underbrush created by confusion, uncertainty, and competing versions of the same basic units. States were also well-positioned to provide focal points when a coordinative consensus had not yet been reached. I therefore tentatively presume that *effective* state interventions tended to reinforce the processes described above. However, sufficiently powerful governments may have created exceptions to this generalisation.

Related literature

This article's theory fits comfortably within the New Institutional Economics (NIE) pioneered by Coase (1937, 1960), North (1981), Nelson and Winter (1982), Williamson (1985), Barzel (1982), and many others. Specifically, it exemplifies NIE's tendency to show how historical practices that seem bizarre to modern eyes were actually efficient, or at least functional, given relevant conditions. See, e.g., Allen (2012) and Leeson (2009). As Allen puts it, 'societies are driven to find institutions that get the job done best under the circumstances faced at the time' (2012: 98). Some in this tradition would even say *all* historical institutions were constrained efficient, perhaps tautologically so (Leeson, 2020).

The importance of measurement has been widely recognised in NIE (Allen, 2012; Barzel, 1982; North, 1991), with particular emphasis on how costly measurement affects organisational forms. One conclusion in the literature is that lower-cost measurements are more likely to become standardised and thus reduce the need for complex contracting (Barzel, 2005). However, the literature has tended to discuss measurement in the abstract, focusing on *when* and *whether* to measure, without much attention to *how* to measure – i.e., the form taken by measures. This article aims to fill that void.

In treating measuring standards as equilibria (and often focal points) of coordination games, the theory is consistent with Denzau and North's (1994) notion of institutions as shared mental models, Nelson and Sampat's (2001) notion of institutions as social technologies, and Lachmann's notion of institutions as points of orientation that promote coordination (Foss and Garzarelli, 2007). A coordinative equilibrium provides a shared mental toolkit – including units and behavioural algorithms to generate them – that promotes social coordination while also potentially conveying useful knowledge (such as how to contain costs).

The theory is also congruent with the literature on standards and modularity, particularly Langlois (2006) and Baldwin (2007). Households and firms are the primary modules in the economic system, and measurement standards are 'a special module whose function is to coordinate the other modules'

[9] I thank an anonymous referee for this example.

(Langlois, 2006: 1,396). This module can emerge via a bottom-up process, although public authorities may also be involved. As Langlois observes, there may be an inverse relationship between costs incurred at different levels (2006: 1,393). For example, the state might promulgate a new top-down system with the intention of reducing coordination costs for households and firms. One insight of this article, cast in modularity terms, is that doing so could inadvertently *increase* measurement costs for households and firms by obliging them to use less convenient and suitable units.

Finally, the theory illustrates Rizzo's (1999) distinction between logical and praxeological coherence. *Logical coherence* refers to the character of a system that is internally consistent in an abstract sense; all terms are well-defined and related to each other by invariant rules. *Praxeological coherence* refers to the functionality of a system in practice; it is concerned with usefulness, convenience, and suitability. Although logical coherence *may* contribute to praxeological coherence, the former is neither necessary nor sufficient for the latter. This insight is helpful for understanding how customary measurement systems could function despite logical inconsistency – and also why the metric system could not have taken hold earlier in history. The metric system, with its universal decimal ratios and naming system, is a model of logical coherence. It is also quite functional in the present day. But that functionality is historically contingent, dependent not only on a sufficiently educated populace and governments powerful enough to enforce metric, but also on industrialisation having reached a point where reliable standard measures (metric or otherwise) can be made cheaply and widely available, thereby obviating problems of costly division. Even so, customary measures maintain a grip in some highly developed corners of the world, most notably the US, but also the UK – where, post-Brexit, the government has decided to allow some Imperial measures to make a comeback (Gross, 2021). Praxeological advantages of customary measures could be part of the reason why.

Illustrative evidence

Binary

Historical metrologists confirm that binary sequences are ubiquitous in customary measurement. Gyllenbok observes that three 'bases' occur more than any other in the division of units, the first of these being 'the binary sequence, which uses 2 as its base, with the first numbers in the sequence consequently being 2, 4, 8, 16, 32, and 64' (2018: 3). (The other two common bases are decimal and duodecimal; more on these later.) Zupko describes the medieval custom of creating additional units by dividing units into 'halves, thirds, and fourths' with prefixes indicating their origins: 'The most important of these units were the demi (= half) series in France such as the demi-arpent, demi-aune, and the like'; in England, 'such renderings were preceded by *farthing*-, *fer*-, *fur*-, or *quart*-'; and in Germany, '*Achtel*- or *Achteling*- (1/8), *Drittel*- (1/3), *Halb*- or *Halbe*- (1/2), *Quart*- (1/4), and *Viertel*- (1/4)' (1990: 14).[10] Kula observes, 'The system of dichotomous divisions and successive dichotomous multiples constitutes, arguably, a universal phenomenon of the primitive mentality' (1986: 83). Although other divisions – particularly thirds – make regular appearances, 'the commonest dichotomous division was of the pure variety,' meaning an uninterrupted binary sequence (Kula, 1986: 85). Many sequences that *appear* non-binary reveal their binary character once we recognise that a single 3:1 ratio has crept in.

The following examples should drive home the frequency of binary patterns in customary measurement.

Early Indus Valley civilisation weights. Among the very earliest measurement artefacts are stone cubes used for weighing in the Indus Valley civilisation. One set found in the Mohenjo-Daro region, dating to 2300 BCE, consisted of 'weights doubled in accordance with the binary sequence, with the following multiples of the base unit of c. 13.65 g: 1/16, 1/8, 1/4, 1/2, 1, 2, and 4' (Gyllenbok, 2018: 3).

Ancient Chinese length measures. The Ancient Chinese traditional length system (c. 1100 BCE–c. 221 BCE) initially appears nonbinary and almost chaotic. As shown in Table 3, units are related by a plethora

[10]Punctuation and formatting altered for clarity.

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

Table 3. Ancient Chinese length measures

yin											
1 ¼	**phi**										
2 ½	2	**liang**									
5	4	2	**tuan**								
6 ¼	5	2 ½	1 ¼	**chhang**							
10	8	4	2	1 3/5	**chang**						
12 ½	10	5	2 ½	2	1 ¼	**hsün**					
20	16	8	4	3 1/5	2	1 3/5	**mo**				
25	20	10	5	4	2 ½	2	1 ¼	**jen**			
100	80	40	20	16	**10**	8	5	4	**chhih**		
125	100	50	25	20	12 ½	10	6 ¼	5	1 ¼	**chih**	
1,000	800	400	200	160	100	80	50	40	10	8	**tshun**

Source: Gyllenbok (2018: 474).

of ratios, including such exotic ratios as 1–3/5 and 6–1/4. But closer inspection shows that the system consists of two overlapping binary sequences. One sequence (shown with dark shading) starts with the *phi*, which by successive halving yields the *liang*, *tuan*, *chang*, and *mo*. The other (shown with light shading) starts with the *chhang*, which by successive halving yields the *hsün*, *jen*, and *chhih*.

These two sequences are connected by one decimal point of contact: one *chang* equals ten *chhih* (marked in boldface). At the upper end, ten *chang* make a *yin*; at the lower end, ten *tshun* make a *chhih*. This decimal sequence links the opposite ends of the scale (*yin-chang-chhih-tshun*), but the two binary sequences dominate the table's centre – as would be expected when binary does a better job of filling the space of convenient quantities.

Most exotic ratios in the table fall out from the one point of contact between these three sequences. But it is implausible that common people regularly converted, say, *chhang* directly into *mo* at a 3–1/5 ratio. More likely, they made indirect conversions using behavioural algorithms consisting of simpler ratios, usually 2:1.

Pre-Akbar weights in North India. The pre-Akbar system of weights in North India emerged sometime before 1556 and persisted until the introduction of metric (Gupta, 2020: 46). As shown in Table 4, two binary sequences are apparent: one at the higher end (dark shading) and one at the lower end (light shading). Furthermore, if we allow the 3:1 ratio between the *maashaa* and *taak*, the lower sequence extends imperfectly up through the diagonally hatched region. The point of contact between the two sequences is the multiply-divisible *chhataank*, consisting of either four *kancha* or five *bhaari*. This yields some peculiar ratios in the table – but again, such conversions were likely accomplished indirectly through sequences of simpler ratios.

English weight and capacity. In the UK, there is a close historical relationship between weight and capacity measures, with units often sharing names. As shown in the introduction, US fluid capacity measures – derived from English measures – display a nearly unbroken binary structure. However, those measures have been supported by powerful modern governments. A better test would be seeing whether the binary pattern persisted over time – and this turns out to be true. Similar binary patterns appeared in virtually every English weight system surveyed by Ross (1983: 20–35), including the Tower pound weight system (791–1527 CE), the Hanseatic merchants' pound system (pre-13th c. – 1582), the avoir-du-pois weight system (1340–1582), the Henry VII Winchester corn weight system (1497–1601), the troy pound weight system (1497-present), the troy corn weight system (1497-?), and the avoirdupois pound weight system (1582 onward). Each system had notable exceptions and discontinuities, but their binary character is nevertheless persistent and usually obvious.

English capacity measures show very similar patterns. English dry capacity systems used mostly the same names as the capacity-like weight systems, followed the same binary pattern, and broke from the pattern in the same ways (Ross, 1983: 37–39). Liquid capacity employed different unit names and exhibited more deviations from binary; nevertheless, binary sequences predominated (Ross, 1983: 42–50).

To summarise, binary ratios are so common in customary measurement systems as to constitute the default against which exceptions are defined. The remaining principles will help explain the exceptions.

Availability and comparability

These two principles are best addressed jointly. The availability principle is widely acknowledged, and the preceding tables provide various examples:

- Needham (1959: 83–84) affirms that early Chinese length measures derived from parts of the body such as 'the finger, the woman's hand, the man's hand, the forearm, [and] the foot.' The *chhih* was the span between thumb and index finger when outspread; the *hsün* was the width of outstretched arms; and the *chang* was (possibly) an adult man's height (Baidu, 2018).[11] Plausibly – and consistent with the earlier prediction – the upper binary sequence could have arisen from halving/doubling the *chang*, and the lower from halving/doubling the *chhih* or *hsün*, thus generating an overlapping pattern.
- Table 4's Indian weight measures include the *chawal*, a grain of rice; the *dhan*, a wheat berry; and the *ratti*, a certain plant's seed (Shrivastava, 2017: 40, 43).

The comparability principle is not obvious in cases, like those above, where the natural ratios of body parts and other objects apparently permitted binary ratios. Clearer illustrations occur in cases where natural ratios were markedly nonbinary. In England, the barleycorn was forced into comparability with the inch – originally a thumb's width – at a ratio of 3:1 (Zupko, 1985: 199), as were the foot and yard (whose origin is disputed but may have been an arm's length [Connor, 1987: 83]). Similarly, in Mesopotamia, the cubit was divided into two feet of three palms each (Willard, 2008: 2,244).

The origin of the 12-inch English foot shows the tension between comparability and binary. The 3:1 hand-to-foot ratio (Zupko, 1985: 177) exemplifies comparability. Yet the hand itself, defined as 'the breadth of the palm including the thumb,' consists of four inches (Zupko, 1985: 177), with each inch divided into binary fractions that still appear on rulers today. Moreover, the foot was sometimes forced into 4:1 comparability with the *palm*, a unit notionally equal to a palm's width without the thumb (Zupko, 1985: 273–274), and the palm itself was divisible into four *digits* (Zupko, 1985: 109). In a possible confusion of the palm with the hand, we even find a legal rule of 1566 specifying 'foure grains of barley make a finger [digit]; foure fingers a hande [palm?]; foure handes [palms?] a foote,' which together imply a 64-barleycorn or 16-digit foot (Robinson, 2007: 51, insertions mine). These ratios were ultimately eclipsed by the 12-inch foot, a change that seemingly resulted from the thumb-wide inch outcompeting the finger-wide digit (Watson, 1915: 129).

The English inch and foot have roots in the Roman system. The word 'inch' derives from the Latin *uncia*, meaning a twelfth part. But the Romans, too, felt the pull of binary measures. The Roman foot (*pes*) could be divided into *either* sixteen *or* twelve parts, with the former (*digitus*) apparently being the earlier division (Gyllenbok, 2018: 551). The 16-part Roman foot may have been inherited from the Ancient Greeks, whose foot (*pous*) consisted of sixteen fingers (*dactylos*) (Gyllenbok, 2018: 488).

A similar pattern appears in customary Indian length measures, which Shrivastava (2017: 40) avers were often based on body parts. In a notably binary sequence, the *dhanush* (height of a bow) consisted of four *aratni* (possibly a cubit), and the *aratni* consisted of two *vitasti* (hand spans) (Gyllenbok, 2018: 536).

[11]I thank my colleagues Yue Zhang and Zhong-Guo Zhou for their help with translations.

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

Table 4. Northern India, Pre-Akbar weight measures

maund										
40	**seer**									
160	4	**pav**								
640	16	4	**chhataank**							
2,560	64	16	4	**kancha**						
3,200	80	20	5	1 ¼	**bhaari***					
12,800	320	80	20	5	4	**taak****				
38,400	960	240	60	15	12	3	**maashaa**			
307,200	7,680	1,920	480	120	96	24	8	**ratti**		
1,228,800	30,720	7,680	1,920	480	384	96	32	4	**dhaan**	
4,915,200	122,880	30,720	7,680	1,920	1,536	384	128	16	4	**chawal**

*Or *Tolaa*. **Or *siki* (inferred from other ratios).
Source: Gupta (2020: 46–47, Tables 1.43 and 1.44). Gupta took these weights from Wikipedia, but also seems to have vetted them for accuracy (for example, he says the *chawal:dhaan* ratio should probably be 2:1 rather than 4:1).

However, the *vitasti* consisted of three *dhanugrana* (bow grips), each of which was four *angula* (finger breadths), resulting in a *vitasti* of twelve *angula* (Gyllenbok, 2018: 536). It seems the natural ratio of a bow grip to a span approximated 3:1, and that ratio resisted the pull of binary.

Divisibility

The duodecimal measures in Roman, English, and Indian length systems above were likely supported by divisibility. As discussed earlier, distinguishing the contributions of comparability and divisibility can be difficult. But independent support for divisibility is provided by duodecimal in systems where natural ratios, and thus comparability, were less salient: weight and capacity. One weight example is the 12:1 *maashaa*-to-*bhaari* ratio (arising from the 3:1 *maashaa*-to-*taak* ratio) in Table 4. In Ross's survey of English weight and capacity systems, a 12:1 ounce-to-pound ratio often displaced 16:1 in otherwise binary sequences (1983: 20–39). The absence of an intermediate 3:1 ratio between *named* units in these English examples supports divisibility's role even where natural ratios were not a factor.

What about other non-binary ratios? The 5:1 ratio turns up occasionally, but generally as a consequence of decimal sequences whose appearance will be addressed later. The 6:1 ratio appears automatically in any duodecimal sequence.

Because of its minimal value in divisibility, 7:1 should be (and is) rare. One famous exception is the Egyptian royal cubit, which consisted of seven palms rather than the six palms of the common or 'small' cubit. Reimer (2014: 94) wryly speculates on how this happened: 'Everything was easy until some pharaoh demanded that his royal cubit have one more palm than everyone else's. I imagine that he made this proclamation to two scribes, the first of whom declared that 7 palms in a royal cubit was no good since division by 7 was awkward. After the first scribe was decapitated, the second agreed that the royal cubit was a wonderful idea.' If this story resembles the truth, the royal cubit shows that powerful governments could override the measures of the common man, particularly for state-sponsored projects. Nevertheless, the more practical small cubit remained in everyday use until it was replaced by a 'reformed' royal cubit of only six palms (Hirsch, 2013: 1–2). Notably, the palm was divisible into four digits, meaning the small cubit had twenty-four digits – duodecimal again.

Once duodecimal ratios were present, the marginal utility of further divisions seems to have declined rapidly. As Gyllenbok summarises, duodecimal divisions 'often turned out to be a sufficient subdivision for most cultures in history' (2018: 3).

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

Counting

The section on 'Counting principle' predicted a category of deviations from binary at the upper end of measurement scales, where people were inclined to rely on counting over measuring. These deviations are expected to be decimal, duodecimal, or vigesimal. This is what we observe:

- In Ancient Egyptian capacity, the *hekat* was divided in binary fashion downward as described in the introduction – but moving upward, the progression was largely decimal (Gyllenbok, 2018: 483).
- In Ancient Chinese length (Table 3), the highest unit – the *yin* – was either ten *chang* (from one binary sequence) or 100 *chhih* (from the other).
- In England, the troy and avoirdupois pound weight systems both included the hundredweight, defined as 100 pounds; in avoirdupois, this was called the 'short' hundredweight to distinguish it from the 'long' hundredweight, which fell within a binary pattern. Vigesimal also appears here; in both cases, twenty (short or long) hundredweights made one (short or long) ton. (Ross, 1983: 25, 29)
- In several English corn (i.e., grain) weight and capacity systems – specifically, Henry VII Winchester, Elizabeth I Winchester, and William III Winchester – the binary pattern gave way to decimal at the very high end. There were ten cooms to the wey, ten quarters to the last, and two weys to the last – yielding a 20:1 coom-to-last ratio (Ross, 1983: 24, 34–35). Thus, overlapping binary and decimal ratios yielded a vigesimal one.
- In the Pre-Akbar North Indian weight system shown in Table 4, the highest unit – the *maund* – was 40 times the *seer*, the highest unit of the upper binary sequence.
- In the above systems, vigesimal and decimal predominate. However, in continental Europe, Kula finds duodecimal to be dominant: 'As far as transactions *involving counting* are concerned, it would appear that the duodecimal system prevails throughout Europe: the dozen rules, assisted by its divisions and multiples. The unit of twelve dozen, or 144, has its own names, for example, 'the large dozen'' (Kula, 1986: 83, emphasis added).

The section on 'Counting principle' predicted a similar set of deviations at the lower end of measurement scales. This, too, is evident in the systems discussed:

- In the Ancient Chinese length system shown in Table 3, the *tshun* is the smallest unit of the decimal sequence. Ten *tshun* yield one *chhih* (smallest unit in the lower binary sequence), while 100 *tshun* yield one *chang* (smallest unit in the upper binary sequence).
- In English weight systems, when pennyweights were present, the ounce was defined as 20 pennyweights, as the penny coin was often used as a weight (Connor, 1987: 125). Various ratios of grains to the pennyweight occurred, but ultimately the duodecimal 24:1 prevailed (Ross, 1983: 20–21, 24–25).

In short, we see ample evidence of 10:1, 12:1, and 20:1 ratios in cases where counting would tend to replace direct measurement, even when binary ratios otherwise dominated.

Suitability

The suitability principle is apparent in some consumption-driven units, such as the tablespoons, cups, and pots (pottles) in Anglo-American capacity measure (Table 1). As discussed, these units follow a mostly binary pattern. But one famous exception is the teaspoon, which is one-third of a tablespoon. At one time, the teaspoon was a dram (one-quarter tablespoon) and thus consistent with binary. But during a time of falling tea prices and rising tea consumption (Smith, 1992), the teaspoon increased in size to hold more sugar (Griffith, 1859: 25). In this instance, suitability trumped binary.

In some English wine and ale capacity systems (Ross, 1983: 43–47), a jarring exception to otherwise markedly binary patterns is the 'reputed quart,' equal to one-fifth gallon. This unit derived from an

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

alternate definition of the gallon as eight pounds of *wine* rather than *wheat*, with the reputed or unofficial quart being one-quarter of this (Connor, 1987: 187). This alternative quart outcompeted the official quart as the customary size of a wine bottle, perhaps as a more desirable quantity for consumption, perhaps as a means of minimising the excise tax on glass (Moody, 1960: 65). Because bottles of this approximate size were cheaply available (Jones, 1986: 11), the division problem wasn't a binding constraint; merchants could simply pour other measures into these containers. Despite its disagreement with other capacity measures, the reputed quart nevertheless generated its own binary pattern in the reputed pint and reputed half-pint (Jones, 1986: 108).

Suitability is even more evident on the production side. Among English systems of weight, the most significant deviations from binary occurred in specific trades such as wool, hay, lead, and precious metals (Ross, 1983: 26–34). Binary appeared occasionally in these systems but was far less common. Fully explaining the ratios used would require examining the production and distribution practices of these specific trades. To take one example, the avoirdupois old hay weight system had 56 pounds to the truss and 36 trusses to the load (Ross, 1983: 30). 'Truss' comes from the Old French word for packing, while a 'load' was the amount that could be loaded into a cart (Zupko, 1985: 237, 421). It seems reasonable to assume these units and their ratios derived from the physical constraints of packaging and shipping hay.

As expected, suitability can interact with binary. Consider cloth measurement. The 12-inch foot, as discussed earlier, is a notable exception to binary in English measurement. But cloth is the exception to the exception. Cloth was measured by the yard, then sequentially halved into the half-yard, quarter, half-quarter, nail, and half-nail (Connor, 1987: 84; Zupko, 1985: 256). The insistence on binary in this trade makes sense because folding is especially convenient with cloth, which heightens the usefulness of binary comparisons. Using a yard as the base unit was helpful because of its being equal to half a fathom, the width of two outstretched arms – a natural movement in manipulating cloth. These advantages did not apply with equal strength to other trades. 'Thus the foot and inch are used to the exclusion of the yard in building, while the yard and its binary subdivisions to the exclusion of the foot and inch in measuring cloth, and surveyors in surveying public land use neither the yard, foot nor inch' (Stratton, 1904: 822).

Contact

The contact principle, which encompasses various encounters between different trades' and regions' measuring systems, helps explain some of the most unusual ratios in customary measurement.

An encounter between trades is discernible in English length units. As noted above, different trades relied on different units. But eventually, it became desirable to make them comparable, possibly to allow greater precision in land measurement (Connor, 1987: 82). The yard was made comparable to the rod in the highly unusual ratio of 5.5:1. The origin of the rod's length is disputed; see Connor (1987: 43–44). Whatever its origin, it was presumably used to measure land by tipping it end-over-end or walking it forward repeatedly. This meant it had to be long enough to make the process speedier than foot-to-toe walking, but not so long as to become unwieldy (Connor, 1987: 44). Its length was thus consistent with its purpose. When comparability became necessary, its original length had to be maintained to avoid upsetting the established land-measuring system – and so it was simply redefined in terms of the newer yard, yielding the 5.5-yard rod (Connor, 1987: 83). The ratio was surely awkward, but not practically important given the different typical uses of these measures.

A similar encounter between different coordinative equilibria may explain the overlapping binary sequences in Ancient Chinese length (Table 3). Needham says that the table 'includes several independent systems' (1959: 84), which may have arisen in different trades or regions. The two distinct binary sequences in pre-Akbar weights in North India (Table 4) similarly suggest an encounter between systems.

The 14-pound stone is the most notable deviation from binary in English weight, and its origin reflects the intersection of availability, suitability, contact, and binary principles. For weighing heavy

https://doi.org/10.1017/S1744137423000152 Published online by Cambridge University Press

objects, using a large stone was a natural (available) option. Because stones vary widely in size, different localities and trades could coordinate on quite different stones. Thus, the stone historically ranged from four to 32 pounds (Zupko, 1985: 391), with an 8-pound stone persisting in some uses well into the 20th century (Connor, 1987: 336). But the now-familiar 14-pound stone derived from the wool trade, where its value was codified in 1389 to facilitate wool exports to Florence, which of course had a different weight system (Britannica, 2020); this is the contact principle at work. Despite its complex origin, the stone has nevertheless generated its own binary sequence, with two cloves/nails to the stone, two stones to the tod/quarter, and four quarters to the (long) hundredweight (Ross, 1983: 22, 29).

Concerns, caveats, and conclusions

> The confusions and contradictions of historical unit usage defy the most ingenious present-day attempts to harmonize them or to explain them away.
>
> Arthur Klein, *The World of Measurements* (in Robinson, 2007: 51)

As an economist, I have ventured into the field of historical metrology with some trepidation. As Klein implies, many have tried and failed to rationalise customary measurement systems. In this final section, it is therefore appropriate to offer some caveats and concerns.

In the illustrative examples of the penultimate section, I should acknowledge a degree of cherry-picking. My research led to numerous metrological tables, and although they frequently had patterns conforming to the principles discussed, not all cases exemplified them as clearly as those presented here. Moreover, some metrological systems resisted any attempt to make sense of them. To take one example, consider the measures of medium length in the Aztec Empire (Table 5). Per Gyllenbok's description (2018: 465), availability is certainly at work, as at least three units derived from bodily measures (albeit exotic ones from a Western perspective).[12] Comparability is likely at work in the 2:1 and 3:1 ratios. But no binary *sequences* are apparent, and overall the ratios are highly unintuitive.

From this, it is tempting to say the Aztec length system falsifies the theory. Then again, there may be relevant factors invisible to someone who doesn't speak Aztec and hasn't worked with these units. The suitability or contact principle might explain some of the more confusing ratios. There might be two or more distinct systems overlaid atop each other. There may be missing units whose presence would make ratio sequences more apparent. Political or religious factors may have influenced the system. Reporting error may have contributed to the confusion. And, of course, the Aztec system did not survive, though how long it lasted is unclear.

Given cases like this, I should emphasise that the theory explains *many* seemingly peculiar features of customary measurement systems, but not all. Some features of these systems may fall within this article's theoretical framework but only reveal their secrets upon further research.

A different concern relates to the kind of evidentiary support needed. I have supported the binary principle mainly through binary patterns in real-world customary systems. Historians agree such sequences were ubiquitous. However, the reasons offered in this article are more speculative: the relative ease of halving and doubling, the advantage of filling the space of useful measures, and the cognitive ease of binary behavioural algorithms. These reasons are intuitive and supported by circumstantial evidence, such as the well-known scarcity of physical standards before modern times. Nevertheless, I have found no historians who explain binary sequences on these grounds (though Kula, 1986 comes close). Nor am I aware of direct evidence such as narrative accounts of merchants describing a process of halving and doubling to create desired units from the standards they had. Similar concerns apply to other principles; for instance, I am not aware of narrative accounts of merchants substituting counting for measurement at large quantities. Perhaps future research will uncover such narratives.

[12]E.g., the *niquizantli* is 'the distance from the ground beneath one foot to the extended fingers of the upraised opposite arm' (Gyllenbok, 2018: 465).

Table 5. Aztec empire length measures (middle scale)

matlacicxitla					
1–1/9	**maitlneuitzanantli**				
1–1/3	1–1/5	**niquizantli**			
1–2/3	1–1/2	1–1/4	**maitl**		
2–2/9	2	1–2/3	1–1/3	**mitl**	
3–1/3	3	2–1/2	2	1–1/2	**yollotli**

Source: Gyllenbok (2018: 465).

The role of powerful governments also warrants further research. In China, consistent decimal measurement systems arose by 200 BCE and possibly much earlier, long before metric in the West (Gyllenbok, 2018: 474–477); powerful dynasties surely played a role. Further research on the influence of educated elites, including mathematicians and architects, would also be helpful – especially in understanding the use of sexagesimal in measuring angles and time (see Macey, 2010: 92).[13] Relatedly, it would be helpful to explore whether the theory applies better, or perhaps worse, to more literate and numerate cultures.

Notwithstanding these concerns, explaining customary measurement systems with an approach similar to this article's seems natural and almost inevitable. Facilitating commerce is a principal advantage of measurement. It stands to reason that the needs of commerce, including coping with transaction costs, would have shaped the form of measuring units. Such transaction costs include both the everyday costs of implementation and the challenge of coordinating on shared measures with other users. The seven principles of customary measurement follow naturally from these two factors.

Aside from its historical interest, I hope this article's thesis helps advance a more widely applicable idea: the distinction between logical coherence and praxeological coherence (Rizzo, 1999). Academics and intellectuals naturally gravitate toward abstract logical modes of thought – and then chafe when they do not describe the world. But the logic of the mind is not always the logic of life. The rules that guide real-world behaviour do not necessarily need to be consistent with each other; they need only be consistent with the pragmatic purposes they serve.

References

Akerlof G. (1970) The market for "Lemons": quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **84**(3), 488–500.

Allen D.W. (2000). Transaction costs. In Bouckaert B and De Geest G (eds.), *Encyclopedia of Law and Economics, Volumes I–V*. Cheltenham, UK: Edward Elgar Publishing.

Allen D.W. (2012) *The Institutional Revolution: Measurement and the Economic Emergence of the Modern World*. Chicago: University of Chicago Press.

Allen D.W. and Lueck D. (2009) Customs and incentives in contracts. *American Journal of Agricultural Economics* **91**(4), 880–894.

Arthur B. (1989) Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal* **99** (March), 116–131.

Baidu (2018) What are the ancient units of length? How come? In Baidu Baike. Retrieved March 8, 2023, from https://zhidao.baidu.com/question/424960617732677652.html.

Baldwin C.Y. (2007) Where do transactions come from? Modularity, transactions, and the boundaries of firms. *Industrial and Corporate Change* **17**(1), 155–195.

Baldwin C.Y. and Clark K.B. (2002) Where Do Transactions Come From? A Perspective from Engineering Design. Working Paper, Harvard Business School.

Barzel Y. (1982) Measurement cost and the organization of markets. *Journal of Law and & Economics* **25**(1), 27–48.

[13]My tentative guess: Divisibility trumped binary because angles and (especially) periods of time were not as susceptible to direct comparison for equality.

Barzel Y. (2005) Organizational forms and measurement costs. *Journal of Institutional and Theoretical Economics* **161**(3), 357–373.

Britannica T. Editors of Encyclopedia (2020). stone. Encyclopedia Britannica. https://www.britannica.com/science/stone-unit-of-weight.

Chuah S.H. and Hoffmann R. (2003) The Evolution of Measurement Standards. Working Paper, Nottingham University Business School.

Clark A. and Chalmers D. (1998) The extended mind. *Analysis* **58**(1), 7–19.

Coase R.H. (1937) The nature of the firm. *Economica* **4**(16), 386–405.

Coase R.H. (1960) The problem of social cost. *Journal of Law and Economics* **3**(1): 1–44.

Connor R.D. (1987) *The Weights and Measures of England*. London: Her Majesty's Stationery Office.

D'Adderio L. (2011) Artifacts at the centre of routines: performing the material turn in routines theory. *Journal of Institutional Economics* **7**(2), 197–230.

Dahl J. "Eye of Horus bw." (2007) Wikimedia.org. Wikimedia Commons. Web. 23 November 2021.

David P.A. and Greenstein S. (1990) The economics of compatibility standards: an introduction to recent research. *Economics of Innovation and New Technology* **1**(1): 3–41.

Demsetz H. (1967) Toward a theory of property rights. *The American Economic Review* **57**(2), 347–359.

Denzau A.T. and North D.C. (1994) Shared mental models: ideologies and institutions. *Kyklos* **47**(1), 3–31.

Evans N. and Levinson S.C. (2009) The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* **32**(5): 429–492.

Farrell J. and Saloner G. (1985) Standardization, compatibility, and innovation. *The RAND Journal of Economics* **16**(1), 70–83.

Foss N.J. and Garzarelli G. (2007) Institutions as knowledge capital: Ludwig M. Lachmann's interpretative institutionalism. *Cambridge Journal of Economics* **31**(5), 789–804.

Griffith R.E. (1859) A Universal Formulary: Containing the Methods of Preparing and Administering Officinal and Other Medicines. New ed., revised and expanded by Robert P. Thomas. Blanchard and Lea.

Gross J. (2021) Britain Signals Intent to Revert to the Imperial System. *The New York Times*. https://www.nytimes.com/2021/09/17/world/europe/imperial-measurements-pounds-ounces-return.html.

Gupta S.V. (2020) Units of Measurement: History, Fundamentals and Redefining the SI Base Units. 2nd edition. Cham, Switzerland: Springer Nature Switzerland AG. Accessed November 23, 2021 at https://books.google.com/books?id=R_7sDwAAQBAJ&newbks=0&printsec=frontcover#v=onepage&q&f=false.

Gyllenbok J. (2018). *Encyclopaedia of Historical Metrology, Weights, and Measures*, Vols. **1, 2, and 3**. Cham, Switzerland: Springer International Publishing.

Hamilton E.J. (1944) Monetary problems in Spain and Spanish America 1751–1800. *The Journal of Economic History* **4**(1), 21–48.

Harper D.A. (2010) Numbers as a cognitive and social technology: on the nature of conventional number sequences used in economic systems. *Journal of Institutional Economics* **6**(2), 167–190.

Hirsch A.P. (2013). Ancient Egyptian Cubits – Origin and Evolution. Dissertation, Near and Middle Eastern Civilizations, University of Toronto.

Jones O.R. (1986). Cylindrical English Wine and Beer Bottles: 1735–1850. National Historic Parks and Sites Branch, Environment Canada – Parks.

Katz M.L. and Shapiro C. (1985). Network externalities, competition, and compatibility. *American Economic Review* **75**(3), 424–440.

Klaes M. (2008). Transaction costs, history of. In Durlauf SN and Blume LE (eds.), *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan. Available online at https://discovery.dundee.ac.uk/en/publications/transaction-costs-history-of

Kula W. (1986). *Measures and Men*. R. Szreter, trans. Princeton, N.J.: Princeton University Press.

Langlois R.N. (2006) The secret life of mundane transaction costs. *Organizational Studies* **27**(9), 1389–1410.

Leeson P.T. (2009) *The Invisible Hook: The Hidden Economics of Pirates*. Princeton: Princeton University Press.

Leeson P.T. (2020) Logic is a harsh mistress: welfare economics for economists. *Journal of Institutional Economics* **16**(2), 145–150.

Leibowitz S.J. and Margolis S.E. (1990) The fable of the keys. *Journal of Law & Economics* **33**(1), 1–25.

Macey S.L. (2010) *The Dynamics of Progress: Time, Method, and Measure*. Athens, Ga.: University of Georgia Press.

Menger C. (1892) On the origin of money. *Economic Journal* **2**(6), 239–255.

Moody B.E. (1960) The origin of the 'Reputed quart' and other measures. *Glass Technology* **1**(2), 55–68.

Needham J. (1959) *Science and Civilisation in China: Volume 3, Mathematics and the Science of the Heavens and the Earth*. Cambridge: Cambridge University Press.

Nelson R.R. and Sampat B.N. (2001) Making sense of institutions as a factor shaping economic performance. *Journal of Economic Behavior & Organization* **44**(1), 31–54.

Nelson R.R. and Winter S.G. (1982) *An Evolutionary Theory of Economic Change*. Cambridge: Belknap Press of Harvard University Press.

North D.C. (1981) *Structure and Change in Economic History*. New York: W.W. Norton & Co.

North D.C. (1991) Institutions. *Journal of Economic Perspectives* **5**(1), 97–112.

Owen W.J. (1966) The history of the English system of weights and measures. In Chishold LJ (ed), *Report of the 50th National Conference on Weights and Measures 1965, National Bureau of Standards Miscellaneous Publication 272*, Washington, D.C.: US Government Printing Office, pp. 128–140.

Pieces of Eight (n.d.) Encyclopedia.com. Cengage. Accessed on November 23, 2021, at https://www.encyclopedia.com/history/united-states-and-canada/us-history/pieces-eight.

Reimer D. (2014) *Count Like an Egyptian: A Hands-On Introduction to Ancient Mathematics*. Princeton, N.J.: Princeton University Press.

Rizzo M.J. (1999) What kind of legal order? Logical coherence and praxeological coherence. *Journal des Economistes et des Etudes Humaines* **9**(4), 497–510.

Robinson A. (2007) *The Story of Measurement*. London: Thames & Hudson.

Ross L.A. (1983) *Archaeological Metrology: English, French, American and Canadian Systems of Weights and Measures for North American Historical Archaeology*. Hull, Quebec: Supply and Services Canada.

Schelling T. (1960) *The Strategy of Conflict*. Cambridge: Harvard University Press.

Shrivastava S.K. (2017) Measurement units of length, mass and time in India through the ages. *International Journal of Physical and Social Science* **7**(5), 39–48.

Smith W.D. (1992) Complications of the commonplace: tea, sugar, and imperialism. *Journal of Interdisciplinary History* **23**(2), 259–278.

Stratton S.W. (1904). Weights and measures. In Beach FC (ed), *Encyclopedia Americana*, vol. **16**, New York: The Americana Company, pp. 820–827.

Swann, G.M.P. (2009). The Economics of Metrology and Measurement. Report for National Measurement Office, Department for Business, Innovation, and Skills. United Kingdom.

Tirole J. (1988) *The Theory of Industrial Organization*. Cambridge, Mass: The MIT Press.

Vincent J. (2022) *Beyond Measure: The Hidden History of Measurement From Cubits to Quantum Constants*. New York: W.W. Norton & Co.

Watson C.M. (1915) The origin of English measures of Length. *Journal of the Royal Society of Arts* **64**(3293), 125–131.

Willard R.H. (2008) Weights and measures in Egypt. In Selin H (ed), *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*, Dordrecht: Springer, pp. 2244–2251.

Williamson O.E. (1985) *The Economic Institutions of Capitalism*. New York: The Free Press (MacMillan).

Young H.P. and Burke M.A. (2001) Competition and custom in economic contracts: a case study of Illinois agriculture. *American Economic Review* **91**(3), 559–573.

Zupko R.E. (1985) *A Dictionary of Weights and Measures for the British Isles: The Middle Ages to the Twentieth Century*, Philadelphia: American Philosophical Society.

Zupko R.E. (1990) *Revolution in Measurement: Western European Weights and Measures Since the Age of Science*. Philadelphia: American Philosophical Society.

Cite this article: Whitman G (2023). Toward an economic theory of customary measurement. *Journal of Institutional Economics* **19**, 709–728. https://doi.org/10.1017/S1744137423000152