early writing

A cognitive archaeological perspective on literacy and numeracy.

Karenleigh A. Overmann
Abstract

This inquiry seeks to understand how the original form of writing in Mesopotamia—the small pictures and conventions of protocuneiform—became cuneiform, a script that could not be read without acquiring the neurological and behavioral reorganizations understood today as literacy. The process is described as involving small neurological and behavioral changes realized, accumulated, and distributed to new users through interactions with and concomitant incremental changes in the material form of writing. A related inquiry focuses on why and how numerical notations differ from other written signs. Crucially, numerical signs instantiate their meaning, a representational mode that contrasts with the signification used to represent non-numerical language and which makes numerical notations contiguous with their unwritten precursors, technologies like fingers, tallies, and counters. Instantiation is related to the perceptual system for quantity; this so-called number sense influences the function and form of numerical signs. Reading is then discussed as a cognitive activity that necessarily involves a material form, a plausible example of extended cognition. Because numerical notations share function and often form with precursor technologies, if the former participate in extended cognition, the latter likely do as well. In conjunction with the contiguity between numerical notations and their unwritten precursors, this complicates the idea that (all) writing is (just) language. Finally, potential follow-on research is suggested.

Keywords

early writing systems; Mesopotamia; literacy; numeracy; extended cognition
Introduction

Literacy is so fundamental to our experience in the 21st century that we often take it for granted. We see writing as a tool that encodes, preserves, and transmits knowledge across space and time, and which subjects ideas to analysis and revision (Olson, 1994). We forget that widespread literacy is mere centuries old, a consequence of the printing press and its ability to make writing more generally available. We ignore the fact that becoming literate means acquiring a set of specific neurological and behavioral reorganizations, the various ways our brains and behaviors change when we learn to read and write. And we think of written characters as symbols or graphemes, ignoring the fact that they are also material forms. As the material component of reading, writing is not just individual pieces of materials with writing on them or the particular characters they might contain. Rather, writing is sets of characters and conventions like the modern Latin alphabet and ancient cuneiform, the system of writing invented in Mesopotamia about 6000 years ago.

In the following analysis of how literacy emerged in Mesopotamia, the focus is the process whereby writing as a material form changed from protocuneiform, which consisted of small pictures and conventions, to cuneiform, a script that could only be read by acquiring the specific neurological and behavioral reorganizations understood today as implicit to literacy. In this discussion, there is a fundamental terminological issue, since *glottographic script*, which means writing that records language, does not readily distinguish between early forms meaningful through resemblance and convention and later forms whose morphological change implies the need to acquire specific neurological and behavioral reorganizations in order to read them. The first affords substantial differences in the choice of vocabulary and syntax and can be read with similar ease by people who speak different languages, while the second represents semantic and phonetic elements of a particular language in a way that restricts choice and is most easily read by someone who knows the language. However, the difference between early and later forms is not reducible to their phoneticity, the degree to which they specified vocabulary and syntax, noting that the later form had more ability in this regard because it incorporated techniques that were not initially available. More central is the morphological change in the material form of writing, something that is tractable to cognitive analysis and interpretation. Here the two states will be distinguished terminologically as *writing*, the early material form that approximated meaning through resemblance and convention, and *script*, the later material form that emerged through and accumulated incremental neurological and behavioral changes until it could no longer be read without them.

Seeing writing as a material form lets us examine how it changed over time. The case of Mesopotamian writing is analyzed here, noting that other cases of original writing (e.g., Egypt, China,
Mesoamerica) have not yet been subjected to similar analyses. They are nonetheless mentioned briefly as appropriate to emphasize that the Mesopotamian case is not likely to be universal in all respects. However, the Mesopotamian case does exemplify a process in which writing by hand causes neurological and behavioral changes in the writers that influence the material form. In response, the material form changes incrementally in a way that serves to accumulate the neurological and behavioral changes, and it acts as a mechanism for distributing them to new users when they learn to read and write. Early Mesopotamian writing (Fig. 1) consisted of small pictures and conventions. Reading and understanding them was a matter of recognizing the objects or forms they depicted; these signs meant what they looked like, or things reasonably related to them (e.g., a picture of a head might mean “head” or “person”). Reading and writing these early characters would not have required acquiring the set of neurological and behavioral reorganizations required for later forms. In contrast, later forms no longer resembled or depicted; reading and writing them involved and depended upon a set of neurological and behavioral reorganizations acquired through training and practice.

Interpreting morphological change in the material form of writing as indicating neurological and behavioral change in writers requires insight into literacy from the perspective of contemporary neuroscience. In a literate brain, the portion of the fusiform gyrus known as the visual word form area (VWFA) recognizes written characters (Cohen & Dehaene, 2004; Dehaene & Cohen, 2011; McCandliss et al., 2003). It also interacts with Exner’s area, the motor region that plans the movements specific to handwriting (Klein et al., 2016; Roux et al., 2009). This region and its functionality are thought to enable us not only to produce written characters, but recognize them as well (Dehaene, 2009; Konnikova, 2014). The fusiform gyrus also interacts with Broca’s and Wernicke’s areas, the brain centers for producing and comprehending speech (Nakamura et al., 2012; Perfetti & Tan, 2013; Tremblay & Dick, 2016). The interactions between these regions associate written characters with language, and the involvement of these regions is highly consistent across individuals, languages, cultures, and writing systems (Bolger et al., 2005; Carreiras et al., 2007; Frost, 2012).

Today, being able to read a script means acquiring these reorganizations, since a script cannot be read without them. These reorganizations are acquired when someone learns to read and write, and being able to read and write demonstrates that these reorganizations have been acquired. These reorganizations enable someone to interact with a specific material form, which is script, and interacting with the material form is how these reorganizations are acquired. For ancient writing, changes in the material form, including the decreased depictiveness often called “increasing abstractness” and the increased ambiguity that characterizes cursive, attest that these reorganizations occurred in ancient peoples.
A significant difference between the two states is that writing could be invented in ancient times, but scripts could not be. For a script to function, the necessary reorganizations in the brains and behaviors of the users would already need to be in place, which of course they were not and could not be until writing had first been invented and then interacted with long enough for brains and behaviors to change. And with no idea of what literacy was until it emerged, there was no teleology, no inventive purpose or goal, to guide its realization. Instead of invention, the material form of script emerged gradually from interactions with writing sustained over generations, a process in which the material form functioned to realize, accumulate, and distribute incremental neurological and behavioral change (Overmann, 2016a, 2021c). Notably, modern script inventions (e.g., the Vai script of Liberia; see Kelly et al., 2021) achieve literacy in generally the same way, but with a significantly compressed timeline that is plausibly related to knowing the purpose, use, effects, and outcomes of writing ahead of time.

Getting from \textit{state}_1, pictures and conventions meaningful through resemblance, to \textit{state}_2, script that could be read only with extensive training and practice, required change in brains, behaviors,
and the material form of writing. Sustained, collaborative interaction with writing as a material form—communities writing across generations of time—incrementally changed the brains and behaviors of the writers. Changes in brains and behaviors, in turn, allowed for incremental adjustments of the material form. And as that material form changed, the neurological and behavioral reorganizations needed to use it became increasingly important, until at some point, writing—which by this time had become a script in the terminology adopted here—could no longer be read without them. Interactions with the material form that is writing influenced change in brains and behaviors, and change in brains and behaviors enabled writing to be manipulated into novel forms that potentialized more change in the system.

A Case Study in the Emergence of Literacy: Mesopotamia

Early writing in Mesopotamia consisted of small pictures that meant what they resembled, or things related to them, so a picture of a head could mean head, person, or capital (Fig. 1). There were also conventions; these meant what everyone agreed they meant, so the quartered circle meant sheep. Early signs (written characters) conveyed ranges of related meanings but gave no clues to pronunciation or type, which made them ambiguous regarding the words or morphemes intended. In fact, they were so inarticulate that scholars are still not sure which language was associated with the earliest writing. It was probably Sumerian, but there are also reasons to think that it could have been Akkadian (Englund, 1998b; Veldhuis, 2014). As can be seen in sign chronologies (Fig. 2), about six centuries after they had emerged, Mesopotamian characters had lost much of their resemblance to the things they once depicted. This change in form, which is also found in Egyptian and Chinese writing (Fig. 3), can be understood neuroscientifically, specifically, in how written characters become recognized.
Figure 2.

Chronology of Mesopotamian signs.

Morphological changes imply neurological and behavioral reorganizations like topological recognition and automaticity.

The diagram does not include the 90 degree rotation that has been variously dated to the mid-third (Studevent-Hickman, 2007), late third (Powell, 1981), mid-second (Picchioni, 1980), and late second (Nissen et al., 1993) millennia.

Data from Nissen et al. (1993, Fig. 106, p. 124).
Mesopotamian

Archaic

Later

Egyptian

Archaic

Later

Chinese

Archaic

Later

Figure 3.

Change in form in early writing.

Archaic forms are meaningful because they resemble objects (from left to right: fish, bird, axe, arrow, and vase). Written forms that no longer resemble imply topological recognition; understanding these “increasingly abstract” forms requires training and practice. Notably, once they emerged, Egyptian forms with increased ambiguity (hieratic) and decreased resemblance (demotic) were used alongside hieroglyphs, which persisted throughout.

Image adapted from De Morgan (1905, Fig. 38, p. 243). (A version of this chart will appear in Overmann, 2022.)
Characters are recognized topologically, through combinations of their structural features and spatial relations. Topological recognition is a function of the fusiform gyrus, a portion of the temporal lobe with an evolutionarily provided functionality for recognizing objects and faces (Coltheart, 2014; Dehaene & Cohen, 2007; McCandliss et al., 2003). As written objects became recognized topologically, the need for them to resemble was relaxed (Overmann, 2016a, 2021a). Once written objects were recognized topologically, their forms could deform and still be recognized. Character recognition and meaning became functions of (a) the characters themselves, even if deformed; (b) the context provided by adjacent characters; (c) and learned associations between the material form and language. This, in turn, freed the set of elements to converge on features and points of contrast that enabled them to be discriminated and individuated. Contrastive features were likely lines, orientations, and shapes that activated the visual system more strongly (Hodgson, 2007, 2012, 2019), a matter currently being explored at the University of Bologna, Italy.

Discrimination is telling the characters apart from one another, while individuation is identifying a character for what it is. As is true of writing today, discrimination and individuation involve recognizing characters through their features. The features of early writing were more obvious than was true of script (see Figs 1 and 2). Discriminating and individuating characters with subtle differences depends on familiarity. With familiarity, objects become easier to discriminate and individuate, so the clues that distinguish them can become more subtle. In turn, subtle clues mean that specific training and practice are needed to gain the familiarity that enables discrimination and individuation. This same effect is found today in facial recognition, especially in differences between same- and cross-race identifications (Brigham et al., 2007; Hayward et al., 2017). The difference in writing as a material form is that its features are subject to manipulation and change. The features differentiating early written characters were relatively obvious, while those differentiating characters of script had become more subtle. The characters of script still resembled the original forms of early writing. Nonetheless, the characters of script were becoming much closer to each other than the originals were. Topological recognition let characters deform, which freed them to differentiate in ways that enhanced the ability to tell them apart and identify them, while allowing the clues that did this to become more subtle.

The chronology of specific signs (Fig. 2) shows them starting as small pictures and conventions whose resemblance had a range of meanings. As writers practiced, they produced signs more efficiently and with greater standardization. Feature recognition of signs relaxed the need to preserve the original forms, so characters became less depictive. This let them converge on points of contrast, under the influence of things like movement automaticity and the biomechanics of production. Signs continued to simplify, suggesting that proficiency in using global cues may have reduced the need
for local detail, explained below. Signs changed from forms whose meanings could be approximated with relative ease to forms that required acquiring specific neurological and behavioral reorganizations to recognize and produce. In essence, written characters became too much alike. It had been relatively easy to tell the earlier pictures and conventions apart, but as characters became less depictive and more alike, the differences between them were too subtle without familiarity. This intensified the need for training, which became more formal, and formal training, in turn, intensified systemic change. Overall, the set of characters lost variability, while the remaining variability converged on features that helped to tell characters apart and identify them.

In becoming simpler, characters lost some of their detail, suggesting an optimization or balance of local details and global cues (Overmann, 2021a). Local detail helps novices discriminate and individuate characters, but it also slows proficient readers, who make greater use of global cues, something novices find difficult to do. The more detail characters have, the longer it takes not just to write but also to read them. The term “compression” expresses a similar idea, the reduction of information that minimizes the physical effort of producing characters (Kelly et al., 2021); this focuses on biomechanical aspects of production and principles like least effort (Zipf, 1949). Today, the tension between local detail and global cues is suggested by diacritics as used within African tonal languages (Bird, 1998, 1999). These have two written forms, one with diacritics for novices and one without for masters; a single form, in comparison, implies a compromise between proficiency levels. Simplification also suggests an effect associated with the adaptation of writing to new languages; further study is needed on this point. Interestingly, simplification is not characteristic of Chinese writing, where characters tend to become more complex over time, particularly in adding phonetic elements (Han et al., 2021). Notably, Chinese writing is also associated with distinct neurological activation patterns that suggest differences in phonological processing (Tan et al., 2005) and working memory (Cantlon & Brannon, 2007), matters that plausibly relate to the additional demands imposed by the greater visual complexity of the characters.

Writing by hand was critical to these developments. Today, we still learn to recognize written characters by interacting with the material form of writing. That is, we practice reading and writing until we become proficient, just like ancient writers did. And just like it did for them, handwriting changes our brains and behaviors, for example, improving our fine motor skills, hand–eye coordination, lexical retrieval, recognition and recall functions, and tolerance for ambiguity in written forms (James & Engelhardt, 2012; Longcamp et al., 2005; Mueller & Oppenheimer, 2014; Sülzenbrück et al., 2011). In the historical development of writing, handwriting afforded the practice and habituation that standardized how characters were formed and automated their production. Most importantly, handwriting allowed for the continual adjustment of the material form. This
was essential to the system’s ability to change, as the material form changed in ways that reflected, accumulated, and distributed incremental change in brains and behaviors.

About 15 centuries after writing began in Mesopotamia, a form of cursive developed, characterized by “abbreviated signs, crowded writing, and unclear sign boundaries” (Veldhuis, 2011, p. 72). Today, we think of cursive as a formal handwriting in which all the letters of a word are connected to each other; it contrasts with block print, where the letters are not connected (Bringhurst, 2004). Cursive can be written more quickly, so it tends to be sloppy and thus ambiguous in its form. Being able to recognize characters even when they are ambiguous is a training effect gained by handwriting practice (James & Engelhardt, 2012; Longcamp et al., 2005). Greater tolerance for ambiguity is simply the ability to recognize characters topologically, despite their increasing deformation. Because cursive can be written quickly, it is writing that is sloppy and ambiguous but fast. Being able to be produced more quickly means that writing keeps up better with the pace of thought, which makes it a tool that can interact with mental content, an important aspect of literacy.

Standardization is forming each character with particular strokes in a particular order. When writing first began, there was no such protocol. Over time, the strokes used in the characters and the order in which they were made became increasingly regular and standardized (Bramanti, 2015; Taylor, 2015); standardization, in turn, shows handwriting behavior becoming automatic. Automaticity frees up cognitive resources like attention and working memory (Logan, 1992). Simply, automated behaviors do not require the kind of sustained, dedicated attention that unfamiliar behaviors do. We experience this same thing when we learn to drive. At first, we must attend closely to operating the car and conditions on the road; as we gain proficiency, however, we pay less attention to these things, becoming alert only when conditions change (Charlton & Starkey, 2011). In writing, automaticity lets us focus on what we are writing, its content, rather than how we are writing, its production. This would have helped transform writing into a tool that could engage mental content directly (Tucha et al., 2008).

Phonographic script is writing that records pronunciation clues, which not all writing does. For example, the early pictures and conventions in Mesopotamian writing were associated with ranges of semantic meanings related to what they looked like, as distinct from expressing particular vocabulary. Mesopotamian writing reduced the ambiguity of its early pictures and conventions by incorporating techniques for specifying them, an important step toward phonographic script. Techniques for specification included determinatives, signs that determined or classified other signs, identifying the type of sign they were. Two such classifiers identified the signs they modified as the name of a god or a geographic place (Hayes, 1990, pp. 30–31). Determinatives are not believed to have been pronounced, and
they could appear before or after the signs they classified (Edzard, 2003; Jagersma, 2010), effectively increasing the visual complexity of signs for non-numerical language.

Another technique for specification was phonography, the visual depiction of pronunciation clues. Sumerian had a lot of single-syllable words and morphemes. This meant that pictures of things like fish could also be used for their syllabic values (Woods, 2010, p. 43). Sumerian also had a lot of homonyms, words and morphemes that are pronounced the same but have different meanings (Hyman, 2006). Using a picture of a fish as a syllable leverages the rebus principle, the use of signs as homonyms, independent of their visual meanings. It is often illustrated by signs whose sequence of names are pronounced as “I can see you” in English. Understanding the meaning of a homonym depends on its context, and it also requires knowledge of the language being written. Otherwise, a sign for a fish used as a syllable appears in contexts where its visual meaning makes little sense. A later technique was the use of phonograms, signs with no meaning other than a phonetic value, typically a syllable.

In Mesopotamia, signs for numbers and then non-numerical language emerged in the late fourth millennium BCE (Krispijn, 2012; Nissen et al., 1993; Veldhuis, 2012). Handwriting at the repetition needed to administer a state-level bureaucracy initiated neurological and behavioral change, like the increased recognition of characters by their features and increased standardization and automaticity in their production (Overmann, 2016a, 2021a). By the third millennium BCE, writing had achieved enough specificity in representing language to identify it as Sumerian (Englund, 1998b; Veldhuis, 2014). The ability to represent phonetic values facilitated the adaptation of writing to other languages with different phoneme inventories, like Akkadian (Cooper, 1996, 2004; Krispijn, 2012). By 2000 BCE, a literacy analogous to how we understand the term with respect to ourselves as readers is signaled by several phenomena. Words and morphemes were no longer being split between lines of text (Cooper, 1996), an integrity of form that implies written objects were being recognized by their features. Cursive writing developed (Veldhuis, 2011), demonstrating tolerance for ambiguity and gaining speed of production. Writing was applied to many new purposes (Krispijn, 2012; Veldhuis, 2011), showing it had achieved significant expressive power. Training was highly formalized (Veldhuis, 2014), because script could no longer be read without acquiring neurological and behavioral reorganizations. And finally, the types and rate of change also decreased around this time.

These factors suggest a general “recipe” for developing a script from writing: It is a process in which the neurological and behavioral changes that writers acquire by interacting with writing, in turn, influence its material form; this then changes incrementally to accumulate and distribute those changes to new users. For characters to
become recognized topologically, there had to be an initial repertoire of conventional signs. These had to be simple, and the material form malleable enough, to enable their production, repetition, recombination, and change. Signs had to be written by hand, enough hours per day and days over years, that brains became trained to recognize them and associate them with language. The behavior had to be repeated and sustained for generations, a production demand associated with state-level bureaucracies. Techniques for specifying words and morphemes, including pronunciation clues, had to be included. And signs could not be numerical, since numbers lack the qualities that motivate increased phoneticity. The next section discusses why numbers are so different.

Numbers

Numerical notations are often thought of as just another part of writing and script, and certainly, signs for numbers are involved in many of the processes just described. However, in actuality, there are many differences between signs for language and signs for numbers (Overmann, 2019), and these are important in order to understand how and why written signs for numbers behave the way they do, as well as their continuities with unwritten forms like tokens and quipus. Literacy is the ability to read and write, which means interacting with a material form that can represent the meanings and perhaps pronunciation clues to a particular language. In contrast, numeracy is reasoning with numbers, which can be written but do not have to be, since they can be represented with objects like the fingers, the notches on a tally, or the beads on an abacus. Literacy and numeracy are severable phenomena, as attested by the fact that there are several numerical notation systems that are not associated with any script, and a few scripts that lack any signs for numbers (Chrisomalis, 2010). For literacy to develop, signs for language must be handwritten, while for numeracy, numbers need not be written. Writing is arguably the first time that language takes material form, and antecedent forms like imagery and iconography on pottery are not engaged with the repetition frequency needed to realize neurological and behavioral reorganizations. In contrast, numerical notations follow and directly develop from material forms like fingers, tallies, and counters (Overmann, 2016b) and share function and often form. Writing for language potentially represents the entire lexicon and phoneme inventory, which requires many symbols. In comparison, numerical notations represent only a subset of the lexicon, and then without phonetic values, and this requires very few symbols. Most importantly, signs for language signify, or represent by depicting, indicating, or suggesting what they mean. In contrast, signs for numbers instantiate, or represent
quantity by being an instance of it, though this is more apparent in early forms, as will be discussed.

Signs for numbers instantiate quantity because we perceive quantity, another evolutionarily ancient ability. Our ability to perceive quantity through the so-called number sense means that we can recognize one, two, and three without counting, an ability known as subitization (Penner-Wilger et al., 2007; Piazza et al., 2011; Railo et al., 2008). Without counting, quantities larger than about four are just “many”; here we can appreciate bigger and smaller in groups, assuming the difference is above a threshold of noticeability, an ability called magnitude appreciation (Brannon, 2006; Dehaene, 2011; Piazza, 2011). Besides the part of the brain that perceives quantity (intraparietal sulcus), numbers involve the parts that know and count on the fingers (angular gyrus), plan motor movements (cerebellum), and make decisions (prefrontal cortex; temporo-parietal junction) (Balsters et al., 2013; Brooks et al., 2014; Frank & Barner, 2012; Gracia-Bafalluy & Noël, 2008; Kringinger et al., 2011; Marinthe et al., 2001; Penner-Wilger et al., 2007; Roux et al., 2003; Vandervert, 2017). This neurological infrastructure means that our perception of quantity is integrated with, and thus directly informs, our interaction with the material forms used to represent numbers. The language functions are not highlighted here because numerical thinking does not significantly involve language (Amalric & Dehaene, 2016; Brannon, 2005; Varley et al., 2005).

As noted earlier, pictures represent by resembling, conventions through social agreement. Both convey ranges of likely meanings, rather than particular vocabulary choices, and both achieve specificity through strategies like determinatives and phonology, techniques that add new information to written signs, thereby increasing their visual complexity. In comparison, numerical signs instantiate quantity. Three fingers, three tally notches, three abacus beads, three cuneiform wedges, and three vertical strokes in the Roman numeral three all mean three by virtue of having three elements. Instantiation makes numerical signs unambiguous in their numerical meaning, so they do not need to be specified further with pronunciation clues. It also makes written forms of number contiguous with their unwritten precursors (e.g., fingers, tallies, tokens), something with no parallel in signs for language at any stage.

The same distinction between signification and instantiation was made by the Belgian surrealist René Magritte, who labeled his 1929 painting of a pipe, *La Trahison des Images*, with the phrase, “*Ceci n’est pas une pipe*” [“This is not a pipe”]. That is, the painting is an image of a pipe, which is signification. It is not itself a pipe, which would be instantiation. So, in the same way a picture of a pipe is not itself a pipe, a sign for sheep is not itself a sheep, even when it looks like one. In contrast, for numbers, four cuneiform wedges are four, six protocuneiform cones are six. A sign for a number is that number because it has that number of
elements. Instantiation means that numerical meaning is unambiguous, across languages and even without language, which is why numerical signs do not need to be specified phonetically. This is true even when numerical signs include conventions like those distinguishing integers and fractions or expressing grouped or “bundled” quantities.

Instantiation involves repetition. However, because of the way the number sense works, we cannot appreciate quantities above about three or four without counting. This means we must count, and counting is laborious, unreliable, and assumes a system of counting already exists. The strategy used to mitigate this is known as bundling. Some specified amount of repetition is replaced and represented by a convention understood as a consolidated value. For example, in protocuneiform, in the general counting system known as sexagesimal system S (Nissen et al., 1993), one small cone meant one. Ten small cones were bundled and replaced by one small sphere that meant ten. Six small spheres were bundled and replaced by one large cone that meant six tens, which we would call sixty. The highest number of repetitions without replacement is known as the unbundled maximum. By consolidating and reducing repetition, bundling makes written numerical information more concise, relative to its unwritten precursors; concision, in turn, makes numerical information more accessible, but requires the user to learn the conventions (Overmann, 2019).

Numerical notations instantiate, either by repeating, like three vertical strokes mean three, or by bundling, like the circle means ten in Minoan Linear A, identifiable through the unbundled maximum of nine vertical strokes. These distinctive properties mean that numbers can be identified in otherwise undeciphered scripts, like Minoan Linear A (Corazza et al., 2021; Packard, 1974). Repetition and bundling are not just unambiguously numerical, they are uniquely numerical. That is, because they instantiate quantity, they embody numerical states and relations that have no counterpart in signs for non-numerical language, and this lets them be identified as numbers. The difference in representational mode is found in the earliest writing. In Mesopotamia, signs for quantity instantiated number through repetition and bundling, while signs for commodity signified objects through resemblance or, in the case of the tablet shown in Fig. 4, convention. Bundling is more apparent in some number systems, like the one from Mesopotamia, than in others, like our familiar Western numerals. In cuneiform, instantiation through repetition was used to a specified extent, where a bundle was inserted and used. This point will be returned to later, because how a system of numerical notations develops in this regard is influenced by the number sense. For higher values, bundling and repetition are combined.
Figure 4.

The obverse face shows the difference in representational mode as it appeared in early Mesopotamian writing (c. 3500–3350 BCE).

Our familiar Western numerals are ciphered forms that descended from the numerical notations of Mesopotamian cuneiform and Egyptian hieroglyphs, which were instantiated and bundled (Chrisomalis, 2004, 2010). The number sense influenced how these numerical notations developed. Of the unbundled groups shown in Fig. 5, the smallest quantities can be appreciated or understood as the quantities they are without counting because they are subitizable. More than about three or four becomes increasingly difficult to appreciate. Beyond the subitizing range, we appreciate differences in magnitude, like seeing there is less to the left and more to the right. As notations, groups with few elements are subitizable. This gives them an inherent identifiability that means they change little over time. Groups beyond the subitizing range are organized into subitizable subgroups. These also become difficult to appreciate, so over time, they are encoded as bundles, and they can simplify as conventional (ciphered) forms that avoid the need to count elements or element
subgroups (Overmann, 2021c). For hieroglyphs, this visual simplification began with hieratic, a cursive form of hieroglyphs that emerged around 2600 BCE (Hoffmann, 2012; Lopriano, 1995; also see Fig. 6).

Consider change in the form of written numbers across five millennia and multiple languages (Overmann, 2021c; also see Figs 6 and 7). The forms of signs for subitizable numbers are conserved because their elements instantiate quantities that the number sense can appreciate. Numbers higher than about four are outside the subitizing range. Like subitizable numbers, they instantiate quantity. Unlike subitizable numbers, they are not appreciable by the number sense. To overcome this limitation, their elements are grouped into smaller, subitizable subgroups (Chrisomalis, 2010). Over time, and because of topological recognition, handwriting effects like automaticity, and so on, these simplify as forms that avoid counting elements or element subgroups; they become conventions whose semantic meanings are typically the numbers one through ten. This does two things. First, ciphered forms are more subject to the mechanisms that change written forms, so their forms change more than those of subitizable numbers do, but much less than the forms of characters for non-numerical language. Second, ciphered forms preserve their numerical relations, so they remain identifiable as numbers.
Figure 6.

Change in the form of written numerals over five millennia and multiple languages.

Numerical notations for the small (subitizable) numbers one and two instantiate quantity and essentially remain one and two linear strokes throughout. In comparison, numerical notations for the large (non-subitizable) numbers five and six change to a greater extent, though not as much as characters used for non-numerical language (compare with Figs 2 and 3).

Cuneiform (not shown) and hieroglyph numbers instantiated quantity; hieratic was a cursive form of hieroglyphs whose instantiative elements are recognizable, and demotic and later scripts are ciphered forms that no longer instantiate. The complex links between Western numbers and their Mesopotamian and Egyptian ancestors are shown in Fig. 7.

The data were sourced from Chrisomalis (2010).

Figure 7.

The complex ancestry of Western numerals.

The diagram simplifies the links between Western numerals and protocuneiform and hieroglyph numbers, omitting, for example, potential interactions with Nabatean, Coptic, Ethiopic, Greek, and Roman numbers. Note that when the practice of using clay tokens as counters might have emerged in Mesopotamia is unknown; some authors have placed it as early as the tenth (Moore & Tangye, 2000) or ninth (Schmandt-Besserat, 1992) millennia; however, the use of small clay objects as counters is unverifiable prior to the emergence of numerical impressions found in assemblage with tokens in the mid-fourth millennium BCE (Amiet, 1966, 1972; Broman, 1958; Oppenheim, 1959; also see criticisms and discussion in Englund, 1998a; Friberg, 1994; Overmann, 2019; Zimansky, 1993). Data sourced from Chrisomalis (2004, 2010).
Instantiation has other consequences. One is that while numbers can be written with phonetic clues to their pronunciation in a specific language, they become much less usable as numbers when this happens. This is because phonetic values make signs visually complex, and visual complexity degrades or destroys the semantic concision that makes numerical relations and patterns apparent and accessible. This can be seen by comparing, for example, the numbers one and two written out phonetically in Arabic (دحاو نانثا), Greek (ένα δύο), and Bosnian (jedan dva) with those represented by Chinese (一二), Western (1 2), and Roman numerals (I II). The lack of any need for clues to phonetic values can also be illustrated with Roman numerals, which most people use without knowing their names in Latin. The lack of phonetic values means that numerical notations work with any language. This is not true of non-numerical language, where reading a phonetic script, one that specifies particular words and morphemes in a particular language, requires both knowledge of the language and training and practice with the script.

Numerical signs are unambiguously meaningful and more usable as numbers without phonetic clues. As a result, phonetic forms of numbers emerge much later. For example, after writing emerged in Mesopotamia in the mid-fourth millennium BCE, the earliest phonetic transcriptions of the Sumerian numbers two through ten appeared centuries later, in the late third millennium (Edzard, 1980; Friberg, 1986; Pettinato, 1981a). As for large Sumerian numbers, their phonetic values were not recorded for over a thousand years (Damerow, 1988). Photonically specified forms of numbers emerged not just later, but also only under specific conditions. Consider the conditions needed to produce the Ebla artifact TM.75.G.2198 (Fig. 8). Writing for non-numerical language had to be in place, as did the ability to depict phonetic values. Most importantly, there had to be a reason to add phonetic clues to numbers, since the visual complexity this adds degrades or destroys their semantic concision as numbers (Overmann, 2021c). The place where this artifact was found, the Semitic city of Ebla, suggests that it was made by Semitic-speaking scribes who wanted to learn the Sumerian vocabulary for numbers in addition to the Sumerian notations, much like we might today learn Latin vocabulary for numbers in addition to the Roman numerals. The artifact was also made just prior to adapting cuneiform to Akkadian, a Semitic language. Perhaps recording phonetic versions of numbers was an initial means and motivation for adapting cuneiform to Akkadian (Overmann, 2021c).
In the historical emergence of a number system, multiple material forms precede written notations (Overmann, 2019). Cross-culturally, the fingers appear to be the earliest device used for counting, presumably because of the neurological interaction between the parts of the brain that appreciate quantity and know the fingers (e.g., Roux et al., 2003). Finger-counting influences number systems toward properties like linearity and stable order (Gelman & Gallistel, 1978), as well as grouping by tens, fives, and twenties (Epps, 2006; Epps et al., 2012), though many other variations are possible (Overmann, 2021b). Since the hands cannot represent quantity for long, material forms that accumulate like the fingers do, but that can do so for longer and with greater capacity (e.g., tallies), may be incorporated. Devices accumulating to quantities higher than the hands encounter the perceptual limits on non-subitizable quantities, motivating the use of bundling. Written numerical notations garner the handwriting effects discussed earlier in conjunction with literacy, with the differences related to and influenced by instantiation as noted. In Mesopotamia, handwriting yielded the cuneiform numbers, which eventually added the convention of place value (Friberg, 2005; Robson, 2007).

As numbers elaborate, their representation becomes more concise (Overmann, 2019). For example, representing the number 75 takes the hands of eight people but a single tally with 75 notches, seven Mesopotamian tokens, three cuneiform signs, and two Western
numerals. Concision increases the volume of data that can be brought together for simultaneous visual inspection, which increases the likelihood that relations and patterns will be noticed. Concision is achieved by minimizing the information that is explicitly represented, while increasing the implicit knowledge the user must supply, concepts like sign meaning and place value (Overmann, 2019). This means the need for training increases, just like it does for the system of writing generally.

For language, handwriting is critical to developing literacy because it is the mechanism that enables the material form of writing to change incrementally in response to neurological and behavioral change in the users. This is not true of numbers, since numbers do not need to be written. Nonetheless, handwriting has distinct neurological and functional effects on numbers. Neurologically, because written objects are recognized topologically, written numbers become objects recognized through their features. They are no longer collections of countable discrete objects, the way numbers represented with technologies like fingers, tallies, and counters are (Overmann, 2019). Functionally, written notations are concise to a degree that their precursors are not, and this enables the collection of relational data in previously unfeasible volumes. For example, multiplication tables are not feasible with fingers, tallies, or counters in the way they are with notations. The concision that notations added to the system for numbers allowed the collection of such numerical relations into tables, which scribes learned as part of their training. This would have caused the scribes to begin to think of numbers in terms of these relations. These new material forms and behaviors would eventually redefine numbers through their relations. Through handwriting and topological recognition, and through concision and relations, numbers would become objects in a relational system, and calculation would ultimately become a matter of manipulating learned relations, rather than physical counters, a transition that would continue for several millennia (Pullan, 1968; Reynolds, 1993; Stone, 1972; Woods, 2017).

Non-numerical writing was also important to the development of complex mathematics, as it allowed calculations to be recorded, initially as narrative descriptions of the steps involved, much like a cookbook recipe (Devlin, 2003). This documentation helped to codify the calculations as algorithms, or series of specific steps to be performed in certain order. This in turn allowed calculations to become longer, more complex, and more accurate, since they no longer had to depend on human memory. This too would have helped numbers to be reconceptualized as objects, ones that could be manipulated by processes and operations. It would take several more millennia before these narrative descriptions would eventually become signs without phonetic clues, like the familiar plus and minus signs used today for operations like addition and subtraction (Schulte, 2015).
Cognition as Extended and Enacted

The insights into literacy and numeracy presented here were realized by analyzing them as systems composed of brains, bodies, and material forms. This model enables the examination of how these systems change over time and the inference of changes in brains and behaviors from changes in the material forms used in writing and numbers. The model itself is relevant to understanding the insights drawn from it, and it has potential utility in understanding how writing develops and the contiguity between written and unwritten forms of numbers. It also serves to complicate the view that (all) writing is (just) language.

What does it mean to say that cognition is a system? After all, cognition is generally conceived as activity in the brain, a construct in which activity in the brain is synonymous with the mind. However, this view of mind and brain is a recent historical development. By the early 19th century—a mere two hundred years ago—the idea that the brain might actually have something to do with cognition was starting to gain traction. Today, it is so uncontroversial to associate mind with brain that the two are often equated. Besides trying to figure out how a material organ like the brain is even capable of phenomena like sensation and consciousness (Block, 2002; Chalmers, 1995, 2017, 2020; Jackson, 1982), debate now centers on what else the mind might include, in addition to the brain. Several claims have become commonly accepted. Cognition is seen as embodied and embedded, influenced by being in a body and an environment (Lakoff & Johnson, 2008; Prinz, 2009; Smith, 1999; Wilson & Clark, 2009). Cognition is also understood as evolving, continuing to change in an evolutionary sense, including the last 10,000 years (Malafouris, 2013). Other claims, however, are minority positions found mainly in philosophical work, and almost never in the mainstream cognitive sciences. These see cognition as extended, as including materiality as an integral component (Clark, 2008; Clark & Chalmers, 1998), and as enacted, as consisting of the interactivity between brain, body, and world (Hutto, 2013; Hutto & Myin, 2013).

It is easy to link neural activity to environmental phenomena in causal terms, and this preserves the Cartesian distinction. For example, vision is a perceptual modality in which neural activity is causally linked to light waves and refractive properties. Simply, light entering the eye excites the neurons in the retina and brain. This chain of causes yields an effect, which we call vision. However, extension and enaction claim that material forms are more than just causally linked. Extension says that materiality is part of the mind, and enaction says that cognition is the interactivity between brain, body, and world. These claims are much more difficult to establish, for reasons that include the fact that we tend to think of things in the world as distinct phenomena that are only causally linked to cognition.
One issue is that not all researchers seem to understand what extension actually means. Some have claimed extension means that objects think, a position known as panpsychism (e.g., Johnson & Everett, 2021). This, of course, is not what extension claims, which is, rather, that objects are an essential component of our thinking. Another issue is that historical discussions of extended cognition were not entirely convincing. For example, an early thought experiment compared different forms of memory (Clark, 2008; Clark & Chalmers, 1998). Otto uses a notebook to help him remember things, and Inga does not. How Otto remembers by writing notes differs qualitatively from Inga's reliance on her brain's innate capacity for storage and recall. However, both systems arguably have the same functionality. Debate then centers on whether these forms of memory are truly equivalent, and the basis on which external aids like notes can be considered as a part of cognition, rather than just causally linked to it.

In the Cartesian model, Otto externalizes some of his mental content onto a notebook, which acts as a passive repository of that mental content. He consults it later, which allows him to recover the information it contains. The notebook appears to be causally linked to cognition, rather than a part of it. But now consider what happens when Otto reads his notes. A person is only reading while he interacts in a specific way with a specific material form, which of course is writing. Reading as a cognitive state does not, and indeed, cannot exist without that material component and its active engagement. Thus, the material form that is writing is critical to the cognitive state that is reading, and reading is the interactivity of brain functions, the material form that is writing, and the behaviors that interface the two. Reading is extended because it includes a material form, and it is enacted because it exists only when brain, body, and writing interact.

Once reading is accepted as a cognitive state that involves brain, body, and a material form, the question becomes what it is that assumes the cognitive state. If we think it is the mind that reads, then we must consider the mind as a system, of which the brain is a component, along with the material form of writing and the behaviors that actively engage it. Redrawing the boundaries of cognition to include material forms and behaviors does not mean the human brain is not critical to cognition, or special in an evolutionary sense. Nor does it mean the material or behavioral components function within the system in the same way the brain does. Certainly, for any individual person, brain and body are always part of the system, while material forms come and go in a way that reinforces the impression that the brain is where cognition occurs. For their part, material forms are malleable in their form and function, durable and persistent in their existence, and publicly accessible in ways that mental content, psychological states, and even behaviors are not. But once material and behavioral components are recognized as integral to cognitive states like reading, we have a valid model of cognition that includes more than the brain, and which possibly applies to more cognitive states than just reading.
The material component of reading—the sets of characters that comprise writing and scripts—includes numerical notations. Consider the earliest unambiguous numerical notations, impressions made in clay with counters made of clay some 6000 years ago in Mesopotamia. Because the impressions were made by counters, the two had the same form (shape and size) and function (instantiation through repetition and bundling). Now, if notations are recognized as part of the extended state that is reading, shared form and function imply that counters are likely part of an extended state as well. Granted, there are important differences between notations and counters, so the extended states they participate in likely differ. For example, counters are manipulable, while notations are fixed. This means that calculating with counters is a matter of physically moving them, while calculating with notations is much more knowledge-based. Such distinctions suggest there are likely to be qualitative differences in the extended states that notations and counters participate in, in terms of matters like the types and amounts of knowledge needed, the physical movements involved, and the focuses and durations of cognitive resources like attention and working memory. But these are the kind of details that can be established empirically, rather than questions that need to be settled.

Modeling cognition as extended and enacted lets us consider cognitive change at the level of groups and across spans of time that exceed what neuroscientific theories and methods can measure, but which are tractable to the theories and methods of cognitive archaeology, as was shown in the previous sections. The extended model lets us look at the historical dimension of literacy and numeracy, which were described as a process in which material forms accumulate cognitive effort, distribute it across space and time, and become increasingly adept at eliciting specific neurological and behavioral responses in their users. The extended model lets us examine how the material form of writing becomes capable of expressing particular languages fluently, and how and why numbers differ in this regard. It also complicates the view that (all) writing is (just) language. For numbers, writing is yet another device in a chronology of material forms that enable their realization and elaboration from the perceptual experience of quantity; the sequence of devices constitutes a manuovisual medium for representing and manipulating numerical information that can be expressed in language but does not reduce to it. Spatial perception is arguably similar in being instantiated by writing but expressible in language, the difference between “ ” and “space." While these issues of categories and boundaries will not be settled here, they suggest that writing is a manuovisual medium capable of instantiating visual perceptions and expressing language.
Conclusion

The model of how literacy emerges as presented here is not concerned with writing’s ability to store and distribute knowledge, nor the effects that the increased knowledge availability and the ability of writing to represent and organize knowledge have on a society’s conceptual content, as these concepts and consequences are well-covered in other work (e.g., Goody, 1986; Goody & Watt, 1963; Olson, 1994; Olson & Cole, 2006). Nor is it concerned with social variability in degrees of literacy between individuals (Veldhuis, 2011). Instead, contemporary neuroscientific and pedagogical insights into the ways that brains and behaviors change through the embodied practice of writing are used to interpret change in the material form of an original writing system over time. The goal is explaining how a script might emerge from sustained, communal interaction with writing as a material form. It is nonetheless recognized that while the number of people actively engaged in writing and becoming literate in the manner described may have been relatively few, the concepts and conceptual changes realized through their engagement would likely have had a wider societal impact. This is because the availability of mechanisms like reading aloud and discussion provide opportunities for concepts and conceptual changes to be relayed far beyond the individuals who actually performed the reading and writing.

Mesopotamia was an appropriate initial case study because the material record presents a unique view of the emergence of writing from practices like seals (Shendge, 1983) and tokens and bullae (Schmandt-Besserat, 1992). While several aspects of how writing emerged in Mesopotamia appear to be common to other original writing systems, further study is warranted, as there are also noticeable differences. For example, in Egypt, writing appears to have increased in ambiguity before losing resemblance and detail. Hieroglyphs maintained their ideal forms throughout, likely the effect of an artistic mandate. Hieratic, a cursive form of hieroglyphs, emerged about six centuries later (Hoffmann, 2012; Lopriano, 1995) and was thereafter used alongside hieroglyphs. While cursive implies topological recognition, tolerance for ambiguity, and increased speed of production, hieratic continued to resemble hieroglyphs, with the latter serving as both visual anchor and ideal. Demotic, a cursive much less depictive of hieroglyphs, emerged about two thousand years later; its “more abstract” form suggests a relaxed social imperative for preserving form and an increased opportunity to develop contrastive elements. Systemic change would also have been influenced by use. In Mesopotamia, writing is seen as having emerged as a bureaucratic tool, while in Egypt, it served purposes like honoring and communicating with the gods, matters that continue to be studied. Use would have affected things like the number of writers and the amount of writing behavior, influencing the rate of systemic change.
Including more writing systems in the model may ultimately gain additional insight into the critical changes, temporal sequencing, and functional interdependencies inherent in the process, beyond the general outline realized from studying Mesopotamian writing presented here. The degree to which other original writing systems (e.g., China and Mesoamerica) are tractable to this kind of analysis will be determined by the availability of early material.

Acknowledgements:

I would like to thank four anonymous reviewers, whose respective expertise in ancient writing systems, grapholinguistics, psycholinguistics, and neurobiology greatly improved the manuscript.
References


Karenleigh A. Overmann is an Associate Professor of Anthropology (adjunct) and co-director of the Center for Cognitive Archaeology at the University of Colorado, Colorado Springs. She is interested in how societies become numerate and literate by using and modifying material forms over generations of collaborative effort, the effect this elaborational mechanism has on conceptual content, how material forms become increasingly refined to elicit specific behavioral and psychological responses, and what this might augur about the future of human cognition. In June 2020, she completed two years of postdoctoral research at the University of Bergen (MSCA individual fellowship, EU project 785793), and she was a visiting scholar at the University of Pittsburgh from Sept. 2020 to June 2021. She earned her doctorate in archaeology from the University of Oxford in 2016 as a Clarendon scholar.