The limitations of the reverse-engineering approach to cognitive modeling

doi:10.1017/S0140525X1200026X

Jay G. Rueckl

Department of Psychology, University of Connecticut, Storrs, CT 06268; and Haskins Laboratories, New Haven, CT 06511.

Jay.Rueckl@Uconn.edu

http://web.uconn.edu/psychology/people/Faculty/Rueckl/Rueckl.html

Abstract: Frost's critique reveals the limitations of the reverse-engineering approach to cognitive modeling – the style of psychological explanation in which a stipulated internal organization (in the form of a computational mechanism) explains a relatively narrow set of phenomena. An alternative is to view organization as both the explanation for some phenomena and a phenomenon to be explained. This move poses new and interesting theoretical challenges for theories of word reading.

Generally, models of skilled word reading are constructed via a process of *reverse engineering*: (i) A body of findings concerning a relatively small set of phenomena is identified (e.g., effects of word frequency, orthographic-phonological regularity, letter transpositions); (ii) an internal organization is hypothesized in the form of a system of computational or neural mechanisms; and (iii) the model is evaluated in terms of whether the hypothesized organization would generate the patterns of behavior that it was designed to explain. This form of theorizing is not entirely circular: The models are also evaluated in terms of their capacity to generate accurate predictions about new facets of the phenomena of interest and, less often, their capacity to address other kinds of phenomena. The reverse-engineering approach is not specific to the study of word reading, but theorists in the domain of skilled word reading are especially adept practitioners of this approach; there are many word reading models, and as a group they are perhaps as detailed and mechanistically explicit as can be found in any subfield of cognitive science.

Frost's article reveals the bankruptcy of the reverse-engineering approach. At one level, his article is largely a criticism of the "new age of orthographic processing" (sect. 1.1, para. 2) – the proliferation of models inspired by the discovery that letter position is coded far less rigidly (in some languages!) than previous models would have led us to believe. Frost demonstrates that the effects of letter transpositions (and other manipulations) are quite different for Hebrew readers than for readers of English (and Spanish and French), and thus, that flexible position coding is not a necessary consequence of how our minds/brains are predisposed to represent strings of letters, but instead depends on the interaction of the reader and his or her linguistic environment.

At a broader level, Frost's article is not simply about how readers represent the orthographic properties of printed words; rather, it is an exploration of how cross-language differences (and commonalities) in word reading should be explained more generally. To the extent that reverse-engineering models can account for these differences, it is by stipulating language-specific differences in the organization of the reading system (in the simplest case, differences in parameterization; in the more extreme case, by positing different sets of underlying mechanisms). In this approach, the impact of the structure of the writing system on the organization of the reading system is more a matter of rationalization than explanation; that is, the model provides no explanation of how experience with a given writing system results in the reading system having a particular organization. Relatedly, although reverse-engineering models can serve to generate hypotheses about the relationship between the organization of skilled and beginning readers, or about the relationship between skilled and disordered reading, they provide little insight about the processes that transform a beginning reader to a skilled reader or how these processes differ in typically developing and reading-disabled individuals.

Given these considerations, Frost's endorsement of learning models over the reverse-engineering approach ("structured models," in his terms) is precisely the right move. I would add to his analysis two key points: First, I believe the field has generally failed to appreciate that these two kinds of approaches represent different understandings of what counts as scientific explanation. For the reverse-engineering approach, the question is how to explain the behavior exhibited by readers in word recognition experiments, and the answer is the organization stipulated by the theorist, which describes the millisecond-scale processes by which the meaning and pronunciation of a printed word are computed. For learning models, the organization of the reading system plays a dual role. It describes the millisecond-scale processes by which a written word is read, and thus provides an explanation of the same kinds of phenomena addressed by reverse-engineering models. At a slower time scale, the organization itself changes as a consequence of learning, and the theory must explain how and why this happens. Thus, organization is both the explanation and the explanandum.

The second point I would add to Frost's analysis is that the acknowledgment that organization must itself be explained, and that learning is central to understanding this explanation, raises a new set of theoretical challenges. (1) We need to understand the nature of the learning process. For example, to what extent is reading acquisition a form of statistical learning? Are the mappings among orthography, phonology, and semantic learned independently, or does knowledge of one mapping constrain how the other mappings are learned? (2) How should the properties of a language or writing system be characterized? It has proven constructive to think that writing systems vary in their phonological transparency (Frost et al. 1987), the grain size of the mapping between orthography and phonology (Ziegler & Goswami 2005), and the richness of their morphology (Plaut & Gonnerman 2000). But these characterizations are imprecise; we need much better ways to quantify these and other dimensions of statistical structure. (3) The properties of an orthography are determined by the properties of the language it represents. Frost hypothesizes in the target article that orthographies are optimized to provide "*maximal phonological and semantic information by the use of minimal orthographic units*" (sect. 3.1, para. 1, italics in the original). Similarly, Seidenberg (2011) proposed that languages with complex inflectional morphologies generally have phonologically transparent orthographies. Our theories should provide a basis for understanding how and why orthographic systems are constrained by the properties of spoken languages. (4) Knowledge is not an all-or-none thing. Stipulated models typically assume otherwise: For example, a lexical unit either exists or not. But an impressive array of evidence indicates that the quality of lexical representations (their precision and stability) can vary substantially, even for skilled readers (Perfetti 2007). Our theories must provide the means to capture these "in-between" states. (5) The organization of the reading system differs for readers of different languages, but also among readers of the same language (Andrews & Hersch 2010; Yap et al. 2012). On what dimensions do these individual differences occur, and what gives rise to them?

Writing systems: Not optimal, but good enough

doi:10.1017/S0140525X12000337

Mark S. Seidenberg

Psychology Department, University of Wisconsin–Madison, Madison, WI 53706.

seidenberg@wisc.edu **http://lcnl.wisc.edu**

Abstract: Languages and writing systems result from satisfying multiple constraints related to learning, comprehension, production, and their

biological bases. Orthographies are not optimal because these constraints often conflict, with further deviations due to accidents of history and geography. Things tend to even out because writing systems and the languages they represent exhibit systematic trade-offs between orthographic depth and morphological complexity.

Frost's article is a discursive tour through some issues about the nature of writing systems, spoken languages, and reading. These issues have been extensively studied from linguistic, behavioral, neurobiological, and computational perspectives (see, e.g., Chomsky & Halle 1968; Daniels & Bright 1996; Hung & Tzeng 1981; Joshi & Aaron 2006; Sproat 2000; Perfetti et al. 2010). The target article achieves the appearance of originality by failing to credit much of this body of work, nor did it benefit from seriously engaging it.

For years, research on reading in different writing systems has focused on the putative advantages provided by more transparent, consistent representations of phonology (the Orthographic Depth Hypothesis; Katz & Frost 1992). For example, researchers have repeatedly demonstrated that children learn to read more quickly in shallow orthographies compared to English, which is notoriously "deep" (Joshi & Aaron 2006). This approach never worked for me (Seidenberg 1992; 2011). If shallow orthographies are easier to learn, why are so many deep ones represented among the highest-scoring countries on the Organisation for Economic Co-operation and Development (OECD), Pisa, literacy assessments? If reading Albanian is "a skill easily acquired" (Hoxhallari et al. 2004), how do the poor Anglos manage? The problem with this research is that it emphasized reading aloud rather than comprehension. People can read aloud with zero comprehension (cf. my Bar Mitzvah), and comprehend texts they cannot read aloud (cf. non-speaking deaf readers). The major limiting factor on reading comprehension is spoken language, not orthography (Hoover & Gough 1990).

In Seidenberg (2011) I have tried to nudge research toward considering both writing systems and the spoken languages they represent. It turns out that they are related in an interesting way. The languages with shallow orthographies (Finnish, Serbo-Croatian, and others) have complex inflectional morphology. Those with deep orthographies (Chinese, English) do not. This relation suggested to me the notion of "grapholinguistic equilibrium" (Seidenberg 2011). The writing systems that have survived support comprehension about equally well. Reading comprehension is a constant that is maintained via trade-offs between orthographic complexity ("depth," number and complexity of symbols, etc.) and spoken language complexity (particularly morphological). So, in Serbo-Croatian, you, the learner gets the spelling-sound correspondences for free, but then spends years mastering the ferocious inflectional system. English is deep, but the words are shorter, the irregularities are partial and concentrated among the high frequency words, and the inflectional system is trivial. Whereas Serbian would be too hard to learn if it were deep (Seidenberg 2011, pp. 164–65), English would be too hard to comprehend if it were shallow (all that abandoned morphology; Chomsky & Halle 1968). I summarized this conjecture by stating, with some hyperbole, "spoken languages get the writing systems they deserve" (Seidenberg 2011, p. 169).

This is a functionalist argument: The characteristics of both languages and writing systems result from satisfying a varied set of constraints related to our capacities to acquire, comprehend, and produce language for multiple communicative functions in characteristic environments. These constraints arise from different sources and often conflict. For example, elisions that promote fluency in speech production can increase comprehension difficulty. Including the vowels facilitates learning to read Hebrew but interferes with skilled reading, as Frost has shown. Billions of people read Chinese, but the writing system is under pressure because it is ill-suited for keyboarding. Writing systems and languages tend to come into alignment (or are placed there by fiat; see point 3 below), but these competing constraints ensure that the result is a compromise and inherently subject to ongoing modification.

Frost is correct in asserting that writing systems need to be understood in terms of the "full linguistic environment" (sect. 1, para. 5), which was the main point of Seidenberg (2011), a chapter in a book to which we both contributed, resulting from a conference we both attended. My chapter is also the proximal source for the "spoken languages get the writing systems they deserve" epigram (which Frost now attributes to a personal communication with the late Ignatius Mattingly). The wording is identical, but the claims are not. Whereas I think languages and writing systems are probably pretty well matched because they satisfy functional constraints arising from multiple sources, Frost claims that writing systems are optimal, their properties largely dictated by a language's phonological structure.

Among the many problems with Frost's account are the following:

1. The argument proceeds by analogy to a version of evolution whereby natural selection creates movement toward optimality, a basic misunderstanding of the theory (http://evolution.berkeley.edu/evolibrary/misconceptions_faq.php#a3). Orthographies evolved, but there is no magic hand directing progress and the outcomes were not as "inevitable" as Frost repeatedly asserts. Accidents of geography and history are to writing systems as mutation, migration, and genetic drift are to evolution.

2. There are ways to assess whether the solution to a problem is "optimal," but they require formalizing the problem and doing some math, which is what distinguishes Claude Shannon from Dr. Pangloss. Frost hasn't established that *any* writing system is optimal. To do so would require deciding, optimal for what? Acquisition? Comprehension? Texting? The erudite Mattingly (1992) wasn't careless enough to write that languages get the writing systems they deserve. Rather, he discussed the *mismatches* between languages and writing systems, and how they tend to diminish over time (because orthography changes the mental representation of spoken language as much as the opposite). This is satisficing, not optimizing.

3. Major changes to writing systems have repeatedly come about through legislative fiat – writing reform. These developments (e.g., Vuk's revision of Serbo-Croatian; the creation of Hangul in 15th-century Korea; character simplification in modern China) were planned rather than "natural," "inevitable" occurrences. Such abrupt innovations (punctuated equilibria?) have often led to great increases in literacy. Many countries have agencies that actively manage their writing systems (e.g., the Turkish Language Association, the Academy of the Hebrew Language).

4. Frost's descriptions of the five writing systems deviate from scholarly treatments (see especially Ramsey [1987, pp. 57–62] on the questionable status of "word" in Chinese; cf. Coulmas 2003; Daniels & Bright 1996). Solecisms abound – here are a few examples: Morphological variations in Serbo-Croatian do result in phonological variations, for example, systematic deformations of stems (Mirković et al. 2011), contrary to Frost's assertion. Writing systems that represent syllables are not alphabets. Frost writes that "nothing is arbitrary when it comes to orthographic structure" (sect. 3.2.3, para. 2), but many things are, starting with the arbitrary association between a visual sign and a pronunciation for many words in languages such as English and French, of which several possible spellings happen to be used).

Regarding Frost's characterization of recent psycholinguistic history, more research is now being conducted on orthographic representation, but there was no "paradigm shift" (sect. 1, para. 3). People did not change their fundamental assumptions about how reading works or how to study it; the science simply expanded. Frost is correct that orthography is shaped by its relations to phonology and meaning (Price & Devlin 2011; Seidenberg 2011, Fig. 9.1). The better accounts of orthographic knowledge that are indeed needed will emerge by taking a neurodevelopmental perspective on how such systems are learned and constraints on what is learnable, as well as differences between writing systems. Finally, the differences in sensitivity to letter

position that Frost emphasizes need to be assessed against statistical properties of writing systems – for example, the frequencies and distributions of letters in words – information that is lacking in most studies. Hebrew roots may resist letter transposition because of their statistical salience, derived from properties of the spoken language, which cause them to be more robustly encoded than most letter patterns in English. Given such statistical differences, the same underlying mechanisms can give rise to different sensitivities to letter position.

Frost and fogs, or sunny skies? Orthography, reading, and misplaced optimalism

doi:10.1017/S0140525X12000271

David L. Share

Department of Learning Disabilities and Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, Faculty of Education, University of Haifa, Mt. Carmel, Haifa 31905, Israel.

dshare@edu.haifa.ac.il

Abstract: I argue that the study of *variability* rather than *invariance* should head the reading research agenda, and that strong claims of orthographic "optimality" are unwarranted. I also expand briefly on Frost's assertion that an efficient orthography must represent sound *and* meaning, by considering writing systems as dual-purpose devices that must provide *decipherability* for novice readers and *automatizability* for the expert.

Frost has sounded a timely wake-up call to reading researchers and other cognitive scientists who are wont to draw universal generalizations on the basis of data collected from a specific culture, language, or orthography. He then asserts that the main goal of reading research is to develop theories that describe the *fundamental* and *invariant* phenomena of reading across orthographies. Among experimental psychologists, elucidation of the cognitive operations common to all readers, and, more generally, to human cognition, has always headed the agenda; "*variant* and idiosyncratic" (target article, Abstract, emphasis in original) factors are less important. But should invariance be our overriding concern? For biologically primary abilities such as depth perception or auditory localization that are acquired early, rapidly, and universally, invariance is unquestionably the rule; variability or individual differences is of lesser concern, often denigrated as the "noise" in the system. However, because learned skills such as reading and writing represent recent cultural innovations that are not part of humans' evolutionary heritage, *variability* rather than invariance is fundamental. Even in the field of spoken-language processing, which is widely regarded by reading researchers as biologically primary (in contrast to written-language processing), it has been argued that there are few, if any, language universals once we consider the full compass of spoken language variety (Evans & Levinson 2009; see also the discussion of WEIRD psychology in Henrich et al. 2010 – both in previous issues of *BBS*). If universals exist in reading – and this is a hypothesis, not an axiom – these are likely to be overshadowed by culture-specific, language-specific, and script-specific differences, as well as by massive inter-individual variance. As Evans and Levinson (2009, p. 429) argue, "Linguistic diversity then becomes the crucial datum for cognitive science."

Does every language get the orthography it deserves? Frost makes the strong claim that orthographies optimally represent speech and meaning, and that the evolution of writing systems is the culmination of a process of optimization. I suggest this note of finality and "optimality" is unwarranted. Every writing system, like spoken language, is a living, breathing organism that must adapt to the ever-changing needs of its users, their culture, and the technology of communication. Written language, like spoken language, ceases to change only when it dies. Frost's "optimalism" may be true of a few languages in societies with a long-standing literacy tradition, but is highly doubtful when it comes to the many developing societies which are relative newcomers to writing and literacy. For example, approximately one third of the world's languages are spoken in Africa (Bendor-Samuel 1996), yet only some 500 have a written form – the vast majority using a European Roman-based alphabetic orthography disseminated by missionaries who took it for granted that their own writing systems would be optimal for non-European languages. Indeed, many Western scholars still presume that European alphabets are inherently superior to non-alphabetic systems (see, e.g., Gelb 1952; Havelock 1982) But are they? The answer is we don't know yet, but as the following three illustrations suggest, Europe's "orthographic elitism," or rather "alphabetism," may be unfounded.

A study by Asfaha et al. (2009) investigated reading acquisition in four African languages in Eritrea that use either alphasyllabic (consonant-vowel [CV]) Ge'ez (Tigrinya and Tigre) or alphabetic Roman-based scripts (Kunama and Saho). All four languages are said to share a simple syllabic structure, a rich morphology, and a common national curriculum. All scripts, furthermore, are highly regular in either phoneme correspondences or CV (*fidel*) correspondences. The teaching of alphabetic Saho script focuses on CV units, whereas alphabetic Kunama is taught phonemically. A sample of 385 first-grade children who learned to read the alphasyllabic Ge'ez by far outperformed children who learned the alphabetic scripts, in spite of the larger number of signs/graphemes. Moreover, the CV-level teaching of alphabetic Saho produced superior results compared to alphabetic teaching of Kunama.

A second case in point comes from the Philippines, where the arrival of the Spanish in the 16th century lead to the marginalization of the indigenous Indic (alphasyllabic) scripts in all but the most remote regions. Kuipers and McDermott (1996) cite reports of unusually high literacy rates among the Hanunoo in the mountains of Mindoro.

A third example is from Southern Sudan, where the Dinka language is written in a European alphabetic orthography, which, according to some observers (John Myhill, personal communication, 2011), is almost impossible to read fluently. Myhill suggests this may be due to complex interactions between linguistic features not found in European languages, including voice quality and tone that can be both lexical and grammatical.

These few illustrations may not be isolated exceptions. There are documented cases of indigenous scripts invented *ex nihilo* by illiterate individuals aware only of the existence of writing systems among neighboring peoples or missionaries. Daniels (1996a) cites numerous examples (including the Cree and Vai syllabaries) and notes that almost all share a common design; signs for CV syllables alone (Daniels 1996a; see also Chen [2011], on Chinese).

A final comment relates to Frost's argument that an efficient writing system must represent sound *and* meaning. I have termed these two dimensions of orthography *decipherability* and *automatizability*. Orthographies can be regarded as dual-purpose devices serving the distinct needs of novices and experts (see Share 2008a). Because all words are initially unfamiliar, the reader needs a means of deciphering new letter strings independently (see Share [1995; 2008b] for more detailed discussion). Here, phonology and decipherability are paramount. To attain fluent, automatized reading, on the other hand, the reader needs unique morpheme-specific letter configurations that can be "unitized" and automatized for instant access to word meaning. Here morpheme-level representation takes precedence. (It may be morpheme *distinctiveness* [*know* versus *no*] rather than morpheme *constancy* [*know/acknowledge*] that is crucial for rapid, silent reading.)

This "unfamiliar-to-familiar" or "novice-to-expert" dualism highlights the developmental transition (common to all human skill learning) from slow, deliberate, step-by-step, unskilled performance to rapid, automatized, one-step skilled processing.