Putting Humpty Together Again: Synthetic Approaches To Nonlinear Variable Effects Underlying Lexical Access

Chris F. Westbury & Geoff Hollis

Department of Psychology, University of Alberta

P220 Biological Sciences Building

T6G 2E9 Edmonton, Alberta, Canada

Correspondence on this chapter should be addressed to Chris Westbury at the above address or:

Tel: 780-492-8518, Fax: 780-492-1768 E-mail: chrisw@ualberta.ca

Acknowledgements: This work was made possible by a Natural Sciences & Engineering Research Council grant from the Government of Canada to Chris Westbury.

Abstract

A large number and variety of variables have been implicated in lexical access by experimental, observational, and clinical methodologies. New variables (such as co-occurrence distance, pink noise in reaction times, and derivational entropy) continue to be defined. The problem of reconciling the different contributions of these variables to lexical access is already overwhelming, and increases as ever more sophisticated variable definitions are invented. In this chapter, we provide an overview of the nature of this problem, and look at work that has attempted to synthesize the contributions of multiple variables to lexical access. In the first section in the paper, we lay out the theoretical nature of the problem, considering the role of constructs in psychological explanations and how to conceive of the potential problem of overfitting the data. In the second section we discuss the importance of nonlinearity and briefly summarize work that has used nonlinear methods to synthesize variable contributions related to measures of lexical access. In the third section, we present our own work that uses genetic programming to find non-linear relations between fifteen variables relevant to lexical decision reaction times.

In his charming book on the role of fictional entities in science and other cognitive enterprises, the philosopher Hans Vaihinger wrote that "Psychological conditions in particular are so intricate that, a priori, just those fictions are on the whole possible and conceivable which [sic] in the main emphasize only one point and neglect others in order this to make the treatment more practicable" (Vaihinger, 1939/1949, p. 21). The observation serves to remind us that experiments in psychology must always be a compromise. We experimental psychologists do our best to find the best dependent measure, the best experimental paradigm, the best model of the phenomenon under study, and the best control of nuisance variables that we can. However, we are always aware that a slight change in any of these might have given us better insight. Perhaps we controlled for a factor that is actually interacting in a relevant way with our independent variables. Perhaps another paradigm would give us a more sensitive measure of the effects of our independent variables. Perhaps our predictors would have been better at explaining variance in the dependent measure if we had transformed them, by taking their logarithm, their square root, or their inverse. Perhaps the constructs guiding our choice of experiments are too simplistic or non-optimally defined in some other way. Perhaps a complex nonlinear regression would reveal that our predictors are together much better at accounting for variance than any linear model suggests they are. As Vaihinger suggests, there are too many possibilities for us to test all of them. Every experiment in psychology is a compromise that makes experimental treatment practical. In this chapter we consider in detail some implications of and ways of dealing with this necessary compromise in experimental psycholinguistics.

Psycholinguistics abounds with well-defined independent variables that have been shown to impact on lexical access: frequency, orthographic and phonological neighbourhood, summed bigram frequency, morphological family size, pink noise, and word entropy measures, to name just a few. In the first section of this chapter, we will consider an ontological question that has been pursued in psychology since it beginnings: In what sense are these variables ‘real’? Are they pure constructs- fictions, in Vaihinger’s sense? Is their apparent causal role an epiphenomenon of the psychological computation, or do they somehow actually have a causal role to play? Are they perhaps proxies for some correlated factor that really does play a role in the computation? In the second section we review some of the work that has been done recently in constructing new predictors of lexical access that are explicitly non-linear, and consider why this matters. In the penultimate section before concluding, we will focus in more detail on work we have done ourselves, that uses computational methods to synthesize non-linear combinations of predictors that maximize the amount of variance we can account for in behavioral measures of lexical access.

Psycho-ontology: On the Nature of Psychological Variables

Discussions of the ontological status of psychological variables go back to the beginnings of experimental psychology (see, e.g., Carnap, 1936; Tolman, 1938; Hull, 1943). The first scientific psychologists made a distinction between two kinds of variables: *intervening variables* and *hypothetical constructs*.

A hypothetical construct is an entity for which an existence claim has been made. An example of an intervening variable is a subatomic particle. Some particles have not yet been observed, even indirectly, but it is claimed that they exist, with enough evidence to justify the construction of very expensive machines to try to detect them. Perhaps a clearer historical example is genes, which were for many decades a hypothetical construct. It was known that *something* had to exist to carry characters in the observed distributions, but until the discovery of DNA no one knew exactly what that thing was (Westbury & Dennett, 2000).

An *intervening variable* is defined as any variable that is *reducible* to an empirical event or property, but for which no claim of existence has been made. An example of an intervening variable is temperature, which we know to be a proxy measure of the motion of particles. In this sense, we could *theoretically* (however *inconveniently*) discard the construct of heat and instead speak directly of the measured motions of real particles. This kind of elimination is by definition always possible for intervening variables. They summarize or otherwise represent an observable phenomenon that could theoretically be directly observed and described without ever mentioning the intervening variable.

A contemporary reviewer of the distinction between hypothetical construct s and intervening variables has noted that "a failure to distinguish these two leads to fundamental confusions" (Meehl and MacCorquodale, 1991, p. 262). The two must be treated in very different ways. To disprove hypothetical construct, one can only appeal to empirical evidence. Many hypothetical entities (for example: ether, the four humours, Cartesian hydraulics) have been proven not to exist, by finding observable entities that could do all the same work as the postulated non-observable entities.

Disproving intervening variables is a very different matter, since an intervening variable is nothing more than a *description* of a state of affairs; a way of looking at things. For this reason, Meehl and MacCorquodale (1991) claimed that "the only consideration which can be raised with respect to a given proposed intervening variable . . . is the question of convenience" (p. 260; see also Meehl, 1977). Refuting an intervening variable would be like refuting a graph or painting. Assuming that there has been no outright fraud (that is, assuming that the representation does accord with the known facts), one cannot refute a graph or a painting. The best one can do is to suggest that a different representation would be better for some purpose. The reasons for offering such a suggestion are many, ranging from the purely aesthetic (you prefer the pristine simplicity of line graphs to the visual heaviness of bar graphs; you prefer impressionism to realism) to the highly pragmatic (the way data were graphed exaggerates what is in fact a very small difference; an architectural rendering specifies relevant details for constructing a building that a water color painting of the building does not).

The bane of cognitive psychology (in particular) as a science is that the variables available for the study of cognitive processes are *always* intervening variables[1]. Assuming that materialism is true and dualism is false, the actual empirical entities that underlie cognitive variables are streams of charged particles careening at great speed between and within neurons. No one who understands the matter could believe that these streams could be observed directly in their entirety or that those streams are even constant within an individual over time or between individuals (see Quine, 1960, Rorty, 1970, Churchland, 1981, 1988 for discussion of the eliminative materialist position that comes closest to making such a claim). For this reason psychology has to content itself with descriptions that may accord better or worse with the observed facts, but that ultimately are only convenient fictions.

Construct extravagance: The problem of over-fitting.

We have found that the claim that all cognitive variables are intervening variables is met with strong resistance from many quarters. Some scientists simply do not like to be told that their

[1] Many would claim that it is not just psychology, but all of science that is limited to intervening variables. Reichenbach (1938) memorably referred to the existence claim for hypothetical constructs as “surplus meaning”.

work consists of constructing what the philosopher Ludwig Wittgenstein called ‘perspicuous representations’, those representations that produce “just that understanding which consists in 'seeing connections'” (1953, S122). Others object for more pragmatic reasons, suggesting that the kind of construct profligacy the claim implies will open up issues that are not just philosophically distasteful but actually anti-scientific. In particular, we are often told that the claim that if all cognitive variables are intervening variables, then over-fitting our data is inevitable. Since this claim would be the end of cognitive *science* if it were true, it is worth considering in some detail.

Over-fitting refers to the problem of handling *degrees of freedom*, the number of free parameters in a model. If a model contains too many degrees of freedom- in particular, as many or more degrees of freedom as the phenomenon being modeled- then it is possible to build models that are extremely accurate but trivial. If a model developed on a specific dataset is allowed to have too many degrees of freedom, the result will be a model that is accurate for the data on which it was produced, but which generalizes very poorly to other data sets collected in similar circumstances. In essence, over-fit models of this type model a specific dataset related to a phenomenon, instead of modeling a general principle underlying that phenomenon in all its myriad manifestations. Modeling a general principle is science. Modeling a specific dataset is tautology.

The argument against construct profligacy goes roughly as follows. If all of cognitive science’s variables are intervening variables, and if intervening variables are descriptions that can vary as we please, and if things than can vary as we please are degrees of freedom, then cognitive science has an infinite number of degrees of freedom and therefore all its theories are over-fit. If this argument were correct, all of cognitive science would be waste of time.

Fortunately for cognitive scientists, it is false.

The error in the argument stems for the equation of intervening variables with degrees of freedom. An intervening variable is not a degree of freedom, because an intervening variable is a description. Descriptions are not degrees of freedom. A reductio ad absurdum will make this clear. We may sometimes think about what kind of descriptive device we will use to present our data to others. A table? A bar graph? A line graph? A box-and-arrow graph? A pie chart? If descriptions of data were degrees of freedom, then each device that was considered for presenting a dataset would add one degree of freedom to that dataset. The absurd outcome would be that the more educated, creative, or thoughtful a scientist was- the more options he or she pondered for data description- the more likely it would be that his or her data was over-fit. Moreover, degree of freedoms would accrue invisibly to others, since there could be no way of knowing by looking at the final data presentation how many different descriptions were considered before it was chosen.

An intervening variable has exactly the same relation to a dataset as a graph: it describes the dataset in a certain way. If we don't like one graph, we are free to choose another. Changes in how a dataset is described cannot affect the dataset itself.

Although there is no relation between the number of intervening variables we use to describe a dataset and the degrees of freedom of that dataset, there are nevertheless constraints on the number of intervening variables that should be used in any description. However, the constraints are not mathematical or otherwise formal, nor are they written in stone. They are aesthetic and pragmatic. William of Ockham famously stated in the fourteenth century that "entities should not be multiplied beyond necessity". He was not stating a law. He was giving sound advice. Most of us accept that accounts of data that have fewer variables are better than those with more variables. There may be exceptions, however. Perhaps a description with more variables might more mathematically transparent, in which case we might choose to go against Ockham's Razor and present the more accessible description. Perhaps (as we will see below) the variables that best describe the dataset may be mathematical constructs that have no obvious real-world counterpart, which we might prefer to keep.

Nonlinearity: A necessary evil

Another reason for resistance to the idea that psychology is limited to coming up with coherent and useful descriptions is that very few psychologists go beyond categorical and linear analyses. Having just one good description at hand makes it easier to assume that the description is actually reality. The decision to stick to categorical and linear analyses is often made unconsciously- this is what we are trained to do in graduate school and we are rarely informed why we should do otherwise. Our reliance on linear analyses has the unfortunate consequence of valuing (short term) convenience at the expense of (the best approximation to) correctness. In this section consider some empirical evidence that shows why non-linear, continuous descriptions are vital in psychology in general and psycholinguistics in particular.

One important reason that all psychologists should be interested in working with nonlinear models was framed by Minsky & Papert (1968). Around this time, a type of neural network known as a *perceptron* was garnering a great deal of interest from psychologists. Perceptrons are a type of neural network that have two parallel input nodes chained to a third output node. Banks of perceptrons are capable of psychologically relevant tasks such as pattern recognition and classification. However, Minsky & Papert provided a mathematical proof showing that traditional perceptrons are unable to solve a certain class of problems known as linearly non-separable problems. The definition of these problems requires understanding of some technicalities of representation in neural networks that would take us outside the scope of this article. The standard example of a linearly non-separable problem is the exclusive-OR function, which outputs true if just one of its inputs is true, and outputs false otherwise. Minsky & Papert's proof that this very simple problem could never be solved by a perceptron rendered perceptrons uninteresting in the context of complex psychological behavior.

Since Minsky & Papert's proof it has been realized that other types of neural networks are powerful enough to offer insights to psychology. Whole perceptrons can be chained together to provide more complex behavior. However, their utility is contingent on the nodes in each perceptron having *nonlinear activation functions*. Chains of perceptrons with nodes only employing linear activation functions can always be reduced to a single bank of perceptrons (Dawson, 2004, pp. 170-173) and, thus, are uninteresting by Minsky & Papert's proof.

The lesson is that computational power does not necessarily increase with structural complexity in any system that only performs linear transformations on its inputs. If a system is to be psychologically interesting – if it is to be more than merely the *sum* of its environment – the system must necessarily be a nonlinear one.

Several researchers working more directly in language-related fields have found other evidence suggests nonlinearity should be at the forefront of psycholinguistics. Van Orden, Holden & Turvey (2003) stress the importance of nonlinear dynamics in cognitive systems. Nonlinear, dynamical systems depend on *reciprocal causality* (as opposed to *domino causality* that psychologists typically assume). In a system with nonlinear dynamics, processing does not unfold in a straightforward, sequential, orderly manner. Instead, each subcomponent or sub-process is fundamentally tied to every other one, in such a way that the manipulations or calculations of one can fundamentally augment the state or trajectory of all others. The end result is a system where the individual contribution of any one component cannot be worked out in the behavior of the whole. Such systems are fundamentally nonlinear and no amount of diligent linear analysis can adequately capture the complexity of the system.

Such systems can be behaviorally complex, and even chaotic. However, they have a tendency to organize their behavior around *attractors* in their space of possible behavioral states. These attractors might be point attractors – singularities in behavioral state space (*phase space*). They might be cyclic attractors – unbreaking loops of behavior in phase space. Or they might be strange attractors. Strange attractors are volumes in a multidimensional phase space that the system converges to. However, once in that attractor, the system does not necessarily follow a recurrent, cyclic pattern as with cyclic attractors. Chaotic behavior can persist within the volume of the strange attractor, so long as the system's behavior does not leave the volume of the strange attractor. Insofar as they organize around attractors, nonlinear dynamic systems are said to be self-organizing – no matter what their original behavior, they tend to reorganize their behaviors into a specific region of space.

This may seem like an abstract philosophical idea on which to found any theory of cognition, language, or language processing. However, some convincing empirical evidence supports the claim that the language access system is a complex dynamic system.

One of the hallmarks of a dynamic, self-organizing system is the presence of pink noise. Pink noise shows up as an inverse (1/f) relationship between the frequency and amplitude of composite waves in a signal on log scales. Van Orden, Holden & Turvey (2003) consider a series of word naming tasks where reaction time is recorded on each trial. They hypothesized that if this behavior was underlain by a self-organizing dynamic system, then reaction times should follow a pink noise pattern instead of the naively assumed white (random) noise. They demonstrated quite clearly that this is the case. The variations in word naming times are not random (normal). Rather, they vary precisely as we would expect them to if they were produced by a self-organizing system.

If this sort of self organizing, reciprocally causal hypothesis is a hallmark of language processing systems, then nonlinear analyses must be a fundamental tool for understanding the pattern of interactivity within such systems. It is not going to be convenient to break down all of the massively interdependent components as linear effects. It is simply not tractable. Nonlinear analysis and description is required if the cognitive sciences are to progress.

Another line of research implicating nonlinearity in linguistic processing is John Holden's (2002) work on spelling and pronunciation (see also Van Orden, Pennington, and Stone, 2001). Holden asserts and provides evidence that ambiguity in orthography and phonology is best described as a *fractal pattern*. Fractals are a collection of self-similar, nested forms that have no *characteristic scale*. An example of the British coastline is often given to elucidate what is meant by this concept. The British coastline has many bays and peninsulas. If you were told to measure the length of that coastline, how would you go about doing this? You could use a meter stick to get a rough estimate of the length. But what about all the smaller peninsulas and bays embedded within the ones that your meter stick can easily measure? You could move down to a smaller stick – perhaps a centimeter stick – for finer resolution of these smaller jags. But what about the ones embedded in these finer features you can now pick up?

True mathematical fractal patterns have infinitely many layers of self-similarity, like these nested bays and peninsula on the British coastline. They also have no characteristic scale of measurement, since any scale we choose to use – whether it be meters, centimeters, millimeters, or whatever – provides us with a different estimate of length.

Holden claims that, just like the bays and peninsulas embedded within each other along the British coastline, there are scales of ambiguity embedded within the sound/spelling of words. Consider the word form, *lead*. At the whole word (semantic) scale, it is ambiguous with one version of the word rhyming with *bead* and the other rhyming with *head*. At the grapheme scale, there is also ambiguity. Consider the _ *e a* _ portion of our word form. This combination of letters maps to multiple phonemic forms – like the two different ones found in the words *mead* and *bread*. This sets the stage for multiple scales of ambiguity in word forms.

If a psychological phenomenon is best characterized as a fractal pattern, no linear descriptions suffice to capture it – it will take far too many parameters to characterize the layers upon layers of self-similarity within the phenomenon. We must necessarily turn to nonlinear methods of description and analysis. Fractals are fundamentally mathematical entities, so we know they can be described mathematically – however, doing so is beyond tractable with linear methods.

Baayen (2004) took steps towards encouraging nonlinear analyses in language research. He used restricted cubic splines to look at the nonlinear relationships between 13 predictors and lexical decision/word naming reaction times. Six of his predictors enter into nonlinear relationships with reaction times. Furthermore, these six predictors are quite varied conceptually: frequency, morphological family size, inflectional entropy, number of simplex synonym sets, word length, and neighbourhood size. Nine of the predictors he looked at had a nonlinear relationship with naming times. Some of these relationships are also *non-monotonic*: that is to say, continual increase of cause intensity does not lead to continual increase of effect intensity.

For instance, Baayen found that word length is best described as having a u-shaped relationship with word naming times, with a minimum in the bend at a word length of 4 (he suggests this may reflect response optimization of our cognitive systems to the structure of language. The median word length for words in his data happened to be 4). Such u-shaped relationships are impossible to get at with conventional analytical techniques, such as two-way ANOVAs, that are most often used by psychologists. When a factor is looked at, it is typically sampled from the two extremes of its range. This completely overlooks the possibility of finding a nonlinear relationship.

Related work closer to our own work that we will discuss in the next section was carried out by Westbury, Buchanan, Sanderson, Rhemtulla & Phillips (2003). Westbury et al. used evolutionary search techniques (genetic programming, described in more detail below; see Koza, 1992) to capture mathematically the nonlinear nature of various psychological phenomena.

In one experiment, they mathematically modeled the effect of the interaction between orthographic neighbourhood size and orthographic frequency on lexical decision reaction times (LDRTs). These are two rigorously studied constructs in the psycholinguistic literature, whose relationship is characterized by a frequency-modulated orthographic neighborhood effect. LDRTs are faster for words with more neighbours, but only for low-frequency words. Westbury et al. were able to mathematically characterize this relationship, providing a more in-depth description of how the frequency-mediated neighbourhood effect works (see Figure 1; see also Hollis, & Westbury, in press) Specifically, convergence of the neighbourhood effect appears to happen around a frequency of 16 occurrences/million. The simple non-linear equation they found to describe the relationship (that was used to generate Figure 1) account for 23.2% of the variance in LDRTs, as opposed to just 4.8% that was accountable for with a linear regression.

The effect is obviously highly nonlinear, and therefore not something which could be characterized by standard linear statistical tools.

Though relatively little work has been done on characterizing the nonlinear aspects of linguistic phenomena in psychology, the work reviewed above convincingly shows that such nonlinear description is valuable. The use of non-linear tools helps us more accurately characterize psychological phenomena, and in some cases it can provide characterizations of phenomena that simply cannot be made with our normal analytic tools (e.g. in the case of non-monotonic relations).

Putting Humpty together again: Mathematical synthesis in psycholinguistics

Linear thinking imposes stifling restrictions on our scientific imagination and understanding. Since there is no limit to the number of descriptions we are allowed to ponder, why should we ponder only the single one that satisfies the assumptions of linearity? Why not pick the best one we can find among the infinite number of nonlinear descriptions? If we can describe an algorithmic way of recognizing a good description, then we can go and drink beer and argue about theoretical questions (because that is what we humans are good at) while our computers search through a very large portion of the infinite space of intervening variables for us (because that is what computers are good at). This is how we propose to put Humpty together again.

Infinity is problematic. Searching a space that is infinitely large will take infinitely long, no matter how fast our computers search it. When we are faced with a very large space that we must describe or search, one method we often use is random sampling. We could ask our computers to make up random mathematical descriptions, to assess each one for value, and to present the best one it has found to us whenever we ask it to. This method might be mildly helpful: occasionally the computer might come up with a better description of a dataset than we had. However, since the space of *possible* descriptions (as well as the subspace of frankly *stupid* descriptions) is very much more massive than the subspace of *good* descriptions, the odds are naturally stacked against this method.

Fortunately, nature has come up with a much more intelligent way of searching infinitely large spaces than randomly: natural selection. Nature does not randomly create new organisms and see if they are viable. It only creates small variants of organisms it already knows are viable, and ‘keeps’ the best variants. Natural selection thereby searches a very small portion of the infinitely large ‘possible organism space’. There are infinitely many creatures that will never evolve on earth that *could* have evolved here. However, the ones that did evolve here are (by definition) guaranteed to be good at the one thing they were selected for: producing viable offspring.

Natural selection is in fact an algorithm for searching a space of possibilities in a clever way (for mathematical analysis, see Holland, 1992; for an extended non-mathematical discussion of this idea, see Dennett, 1995). Natural selection searches only those possibilities that are similar to possibilities that have previously been identified as being good ones. Instead of trying to randomly come up with a good idea, natural selection comes up with good ideas by tweaking good ideas it has already had.

We have made use of this algorithm to get our computers to search the space of intervening variables, using a computational technique known as genetic programming. The algorithm of natural selection needs three things: a way of producing variants, a way of choosing the good ones, and a way of tweaking those good ones to produce more variants.

The production of variants is easy when we consider the space of mathematically-defined models. Each variant is an equation that relates some measured set of independent variables to some dependent variable of interest. For example, if we wanted to have a mathematical model of how ON and orthographic frequency interact in terms of their affect on LDRT, we can easily tell the computer to make up equations that set some function of ON and frequency equal to RT.

How do we know we tell a better predictor equation from a worse one? This is the problem of defining a fitness function. In genetic programming, fitness can be anything that we wish to maximize and that we can describe to a computer. In our work to date we have used a well-defined and well-motivated fitness function: the amount of variance that we can explain in some dependent measure (the square of the linear correlation). When we compare two mathematical models of a phenomenon to each other, we usually think that the one that fits the phenomenon most closely is the better one. The amount of variance explained is one way (though not the only way) of measuring how closely any mathematical model matches the data. This measure has the advantage of being both mathematically well-defined (so we can tell the computer how to use it) and bounded (convenient for avoiding over-flow errors such as may occur if one used summed squared error as a fitness function, since summed squared error can be arbitrarily large).

We have a way producing random variants and a way of assessing their goodness as predictors of some dependent variable. The third and final element we need for genetic programming is a way of tweaking the good variants so that we look for variants in the same general area of search space as the good ones we have already found. Since our variants are equations, this is very easy to do: we can simply combine random elements of good equations to make up new equations. We represent our equations as trees (using reverse Polish notation, which always places the operator before its arguments), randomly choose two equations identified as good, and swap one random substring chosen from each of those equations. The 'offspring' are similar to their parents in the sense that each one only contains elements that are also contained in its parents. However, they are different from their parents because they contain at least one element that is not contained in either parent.[2]

The new population is populated with these offspring, and the process of assessing fitness and allowing the fit to create offspring is repeated. Over time, equations evolve that are better and better at the task for which they have been selected: predicting variance in the dependent measure.

This process is purely stochastic and there is no guarantee that any solution reached is the best one. We use many software techniques to increase the probability of finding a good solution. We will not discuss them in detail here (they are discussed in Westbury et al, 2005; Hollis & Westbury, 2006; Hollis, Westbury, & Peterson, in press). Most of them amount in one way or another to one of two methods. The first is checking to make sure that evolved equations are good at predicting variance in datasets other than the one the one on which they were evolved, to decrease the chance of over-fitting, finding a locally good but universally bad solution. The second method is repeating searches many times, to increase the probability of covering a sufficiently large portion of the search space. Hollis & Westbury (2006) demonstrated that solutions evolved independently are usually very highly correlated, which increases our confidence that these may indeed be the best solutions. They also conducted simulations showing that when there is a known perfect or very good solution to a prediction problem, GP usually

[2] It is of course possible that two the randomly chosen substrings might by chance swap an identical substring, resulting in an offspring that is clone of one of its parents. However, our software disallows this possibility, as well as disallowing duplicates in the population.

finds it very quickly.

We have released free platform-independent software that allows users to harness the power of GP for themselves very easily. The software is called Naturalistic University of Alberta Correlation Explorer (NUANCE) and is available from the Psychonomic Society website at http://psychonomic.org/archive (see Hollis & Westbury, 2006; Hollis, Westbury, & Peterson, in press). In the remainder of this section, we discuss one application of that software, in which we used it to study the effects of 15 predictors of LDRTs and their pairwise interactions[3]. LDRTs are currently a particularly suitable dependent measure for synthetic work on the type we are advocating here, because a large collection of average LDRTs has already been compiled for the English Lexicon Project (Balota, Cortese, Hutchison, Neely, Nelson, Simpson, & Treiman, 2002).

One advantage of the synthetic approach over factorially-designed experiments is that it is possible to study the effects of predictors over a very large set of stimuli if a large set of dependent measures is available. In the work we report here, we looked at LDRTs for 4,778 words. This was every word that was between three and six letters long for which were able to obtain information about all 16 of our predictors,

The sixteen predictors are defined in Table 1. They consisted of five measures computed across both orthographic and phonological word representations: length (letter or phoneme count), frequency, neighbourhood size, neighbourhood frequency, and position-controlled and position-uncontrolled bigram/biphone frequencies. These were calculated using the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). We used the same database to calculate initial and final controlled-trigram frequencies for all stimuli, as an estimate of orthographic head

[3] This work has been previously discussed in Hollis, Westbury, and Peterson, in press.

and body frequencies. The final two predictors were measures taken from a co-occurrence model of semantic organization (Shaoul & Westbury, in press). These models depend up measures of the Euclidean distance between long vectors representing how often each word in the dictionary occurred within a small window (10 words) of each word. One measure, NN, was the number of co-occurrence neighbours (from our 150,000 word dictionary) that fell within a specified distance threshold from the target word. The other measure, Average Radius of Context or ARC, is the average distance from the target word of those co-occurrence neighbours. Each of these has been shown to be predictive of LDRTs (Shaoul & Westbury, in press).

Method

Our method was to use NUANCE to evolve equations using one or two predictors that accounted for the most variance in the LDRTs. We randomly divided our 4778 words into two sets. One set, the training set, was used to evolve the equations. The other set, the validation set, was set aside so that we could make sure those equations generalized to a new dataset. We ran every predictor singleton or pair through NUANCE twenty times in order to maximize the chance that we would find the best predictor equation.

An important point to stress is that we report results for the best predictor equation across all of those runs, *as it performed on the validation set*. Very few users of linear regression ever set aside a data subset to use as a validation set. Failing to do so can almost guarantee that the reported r values will be over-estimates. Any regression- linear or otherwise- computed across a particular dataset is able to account for both dataset-specific and dataset-general variance. Dataset-general variance is of scientific interest. However, dataset-specific variance is, by definition, a source of error when a regression equation is applied to a different set. Increasing error will decrease the amount of variance that can be accounted for in a new dataset. Many studies compound this error by computing their regression equations over a small dataset, which makes it easier to fit. It is safe to say that almost every reported correlation in the psycholinguistic literature is an over-estimate (see Westbury et al, 2003, for an empirical demonstration).

Results

We had several goals in undertaking this study.

One was to find mathematical models of variable-predictor relations that would allow us to better understand those relations.

We were particularly interested in understanding how much explained variance is 'lost' under the assumption of linearity. The amount of variance in LDRTs accounted for in the validation set by each individual predictor in displayed in Table 2[4]. In that table, we compare the amount of variance in LDRTs that is accounted for by the best evolved equation to that accounted for by a linear regression of the same predictor, and by a linear regression of the same predictor after log transformation. Fifteen of the sixteen variables (all but ONFREQ, the average frequency of the target word's orthographic neighbours) were reliably linearly correlated with LDRTs ($p < 0.05$). All of them were reliably correlated after log transformation, which improved the correlation for every variable except UNBP (the length-uncontrolled, place-uncontrolled summed biphone frequency of the target word) and LETTERS and PHONEMES (the number of letters and phonemes, small numbers for which log transformation makes little sense).

We statistically tested the difference in the correlations between the NUANCE transformations, and their untransformed and logged counterparts (see Blalock, 1972, for methodological details). Eight of the sixteen NUANCE-transformed values correlated reliably

[4] Note that the summed variance accounted for is greater than 1. This reflects the fact that many of the variables are highly correlated, and account individually for the same variance.

better ($p < 0.05$) with LDRTs than either their untransformed or logged values. Some of the cases (e.g. ONFREQ, PNFREQ, PFREQ, and CONBG) are of particular interest because the transformation changed a correlation that was very near 0 to a correlation that was much higher. The average of the untransformed r^2 values of ONFREQ, PNFREQ, PFREQ, and CONBG is a negligible 0.002. The average of their NUANCE-transformed r^2 values is over 41 times larger, 0.07. Although log transformation of the four variables reduces this difference substantially, the average r^2 value of the NUANCE-transformed variables is still 1.4 times larger than the average of the log-transformed variables (0.05).

This comparison of r^2 values underscores one benefit of looking at correlations across a range, as opposed to factorial manipulations, of predictors: one is able to get an accurate idea of the effect size one is studying across that entire range. Almost everyone understands that a reliable effect in a factorial experiment is no guarantee that the independent variable accounts for a compelling amount of variance in the dependent variable. Although it is possible to calculate an effect size for any experimental effect, this can only give the size of the effect at the values where the factors were blocked, usually their extremes. None of the untransformed variables accounted for more than 5% of the variance in LDRTs. The average was 2.3%. When the predictors were logged, the average was 6.9%. The NUANCE-transformed variables accounted for an average of 8.3%.

Simply averaging the r^2 values is problematic, for the reason we have given above. When we average r^2 values, some of the variance is being counted more than once because some of the variables are accounting for over-lapping variance. The average therefore over-estimates how much variance is being accounted for. One of the benefits of the NUANCE-transformed variables is that they were evolved to have as close a linear relationship to the dependent variable as possible. This makes it appropriate to run a *linear* stepwise backwards regression on LDRTs of those transformed variables, despite our general suspicion of linear modeling. In this case we know that the relation of he transformed values to the LDRTs is indeed roughly linear, because it was designed to be so.

When we did this regression, five of the variables did not enter into the model: PFREQ, CONBP, PN, PNFREQ, and ONFREQ. The first three have closely orthographic analogues in the model. The eleven remaining predictors together accounted for 41% of the variance in LDRTs. Just four of those eleven variables account for 96% of this variance (39% of total variance): OFREQ (orthographic word frequency), LETTERS (word length), ON (orthographic neighbourhood), and LASTRI (the frequency of the last three letters of the word, an approximate measure of body frequency). All of these variables have been well-studied and are well-known in the study of lexical access.

Interactions

Accounting for as much variance as possible in a relevant behavioral measure is one method of judging the goodness of psychological models. The use of a hard quantitative measure serves as an objective arbiter between models that provide such a measure. However, another desideratum of models of language is that they capture something about *how* different elements in the model are organized and interact. Information about which predictors interact with each other can potentially be used to constrain language models that are more fully specified, such as box-and-arrow models or neurological models. If we know that two or more predictors interact, then a good model should be able to give an account for why this interaction occurs.

Because we have so many variables, we confined ourselves in this study to using

NUANCE to look at all pair-wise interactions between our sixteen predictors, in the manner described above.

There is a price to be paid for working with evolved nonlinear regression equations. Linear regression equations are designed to break up the variance into that portion attributable to each variable, and that attributable to interactions. The equations evolved by NUANCE mix these two sources of variance up in arbitrarily complex ways, leaving no obvious way to decompose each function into its component main and interaction effects.

We have addressed this problem using a somewhat imperfect approach that should nevertheless allow us to get a principled estimate of how much variance we can attribute to any interaction effect. Our method takes advantage, as we did above, of the fact that the output from NUANCE has been transformed to optimize its *linear* relationship to the dependent variable. We conducted two multiple linear regressions, one containing only terms for each variables alone, and the other containing an additional interaction term. By subtracting the variance accounted for by the first equation from the variance accounted for by the second, we obtain an estimate of the size of the interaction term in which we are interested. The approach is imperfect since we cannot be sure that the interaction function contains the same transformation of each predictor as the one that we obtained when we ran that predictor alone: that is, we cannot be sure that two independently-evolved equations are equivalent in terms of how they account for variance in LDRTs. However, we believe that as a rough heuristic this method is a reasonable way to estimate the size of an interaction effect when we cannot decompose the equations by hand.

12/120 (10%) of all possible interaction effects were large enough to be reliable after Bonferroni correction. An additional 14 (12%) were reliable ($p < 0.05$) before the correction. These 26 interactions are presented in Table 3. Although most of the effects are small in themselves, together they account for 12.9% of the variance in LDRTs. Altogether our 16 variables together therefore can account for as much as 54% of the total variance in LDRTs (or just 3.5% less if we exclude the interactions that did not meet the very conservative Bonferroni correction). This is probably a small overestimation because some of the interactions may be accounting for common variance, since they may share a variable or contain a predictor that is correlated with other predictors,

A few observations may serve to guide models of lexical access. One is that several variables that seem to account for almost no variance on their own enter into reliable interactions. For example, ONFREQ (the frequency of the target word's orthographic neighbours) did not enter into the backwards regression on LDRTs, but it was second only to PFREQ (phonological frequency) in entering into reliable interactions. Similarly, UNBP (the length-uncontrolled, place-uncontrolled summed biphone frequency of the target word) accounted for very little (1.8%) of the variance in LDRTs on its own but was the third most interactive variable. These results raise the possibility that some variables that have little direct effect on lexical access may nevertheless modulate the effects of other variables that do have direct effects on lexical access.

Another interesting result from studying interactivity is that it can provide information about how similar or different two variables may be. Across all the words in the validation set, the NUANCE transformation of ON (orthographic neighbourhood size) and ONFREQ (the average frequency of the target word's orthographic neighbours) are highly correlated ($r = 0.74$; $p < 0.0001$), suggesting that these two variables may bear a similar relation to LDRT and therefore perhaps to each other. This would explain why ONFREQ did not enter into the stepwise backward regression: because ON did. However, their pattern of interactivity is quite dissimilar. They both interact with just a single predictor, PFREQ (phonological frequency). ONFREQ interacts with four other variables (UNBP, PN, PHONEMES, LETTERS) while ON interacts with just one other, OFREQ (orthographic frequency). These variables also interact ($r = 0.06$; $p < 0.005$). All this suggests that the two variables do indeed probably play distinct roles in lexical access.

A somewhat related analysis can be done on the PFREQ and OFREQ, phonological and orthographic frequency. These variables are highly correlated ($r = 0.51$ across all 14,582 words of length 4, 5, and 6 in the CELEX database; $p < 0.001$). Nevertheless, interactions with PFREQ account for about three times as much variance as interactions with OFREQ. This finding invites further study through more traditional studies, which we are presently undertaking. We believe that other of the interactions in Table 3 may also reward further investigation, but we leave their identification as an exercise for the reader.

The Shape of Things

Another benefit of the synthetic approach is that it allows researchers to put on firmer ground certain practical decisions that have to be made in the course of psycholinguistic research. Here we consider as an example the convention of taking the logarithm of frequency-influenced variables when they are used as predictors. This transformation is driven by the fact that many psycholinguistic predictors have a non-linear distribution that does not map easily onto the linear range of behavioral measures. The convention of logging flattens the curve. As is clear from Table 2, this makes most log-transformed variables better predictors of behavioral measures than their untransformed counterparts.

We are not aware of anyone who has ever considered that there may be a better transformation than logging, but Table 2 shows that NUANCE can almost always find a better transformation. We have examined these transformations and found a general pattern. All but two (UNBP and UNBG) of the NUANCE transformations in Table 2 are reciprocal transformations, which transform N into some function of 1/N. On average, those transformations account for 34.6% more variance in LDRTs than the logged variables do. This is 'money for nothing', in the sense that merely looking at the same data set in a different way allowed us to make substantial and well-validated gains in our ability to predict LDRTs.

We believe that the reciprocal transformation is a better transformation because its properties are more biologically realistic than the log function. As illustrated in Figure 2, the log function is continually increasing as its argument increases, while the reciprocal function flattens out. Since there is a floor effect in human reaction times, a function that does not continually increasing as its argument increases is generally likely to be better fit to the human data.

Conclusion

We have had several related goals in this chapter.

The first goal has been to champion a synthetic approach towards studies of lexical access. Although we do not believe that a synthetic approach should be the only approach, we do believe the approach is a necessary adjunct to standard factorial experimental manipulations. We provide several reasons for this.

One is philosophical: only the synthetic approach can help us escape from the reification of predictors that so dangerously attractive in factorial studies. Just because a manipulation of a factor has a measurable behavioral effect does not necessarily mean that the factor actually enters into the neural computations underlying language. A synthetic approach opens our minds to the idea that alien combinations of well-defined variables that may ultimately prove to be more ontologically realistic than the more neatly defined variables of human imagination. We need to get used to the idea that the variables we define are almost certainly descriptive proxies for what is really happening in our brains when we use language. They are shorthand descriptions for something much more complex.

A second reason for our enthusiasm for synthetic approaches is that they allow us to consider the effects of variables across their entire range, rather than limiting us to snapshots of their effect at the extremes of that range as factorial experiments almost always do. When we look at variables of interest as continuous rather than categorical, we get an understanding of their effects that is better in many ways that what we get from factorial experiments: better because we can more easily estimate their effect size; we can see how their effects may change across their range; we can more easily understand how they interact with other variables; and we can more easily consider different ways of defining them that may be more scientifically helpful.

A third reason for championing a synthetic approach, to which we have only alluded in this chapter, is purely pragmatic. As the number of well-defined variables that impact on language access increases, it becomes increasingly impractical to conduct all the experiments that would be necessary to understand how they interact with each other. The sixteen variables we considered here by no means exhaust the space of variables that might be worth considering, but they have 120 pair-wise interactions. Although some of those might be ruled out a priori as uninteresting, we are doubtful that researchers' intuitions have a sufficiently comprehensive grounding to be trustworthy to make such rulings. However, to conduct 120 factorial experiments to determine which interactions might be worthy of consideration in our models of language is totally impractical: the work would take years and most of it would be dedicated to uncovering null effects. We need more powerful tools for searching variable space than experiments. In this chapter we have not even considered higher-order interactions such as the 560 possible three-way interactions among our variables. However, the synthetic program, by virtue of its principled skepticism about the ontological status of human-defined variables, gives reasons to believe that some of these interactions may be worth knowing about.

A related goal to this main goal is to encourage the use of the free tool we have developed for undertaking synthetic studies without making assumptions about linearity: our platform-independent computer program, NUANCE. The last section of this chapter was dedicated to giving some idea of the kinds of information that can be gleaned using NUANCE. It has been designed to be as user-friendly as possible. More detailed information about NUANCE can be obtained from Hollis & Westbury (2006) and Hollis, Westbury, and Peterson (in press).

Another goal of this paper that has been to motivate certain considerations for building models of language processing. One criterion more implicit than explicit is to encourage the development of algorithmically well-defined measures for comparing models of language processing. Box-and-arrow models of language processing are common, but are almost certain by their very nature to encourage reification and are almost impossible to compare empirically. In this paper we have concentrated on considering mathematical models that predict a behavioral measure. Although such models will never be the final endpoint for understanding language (leaving aside as they do a great many important issues, such as neurological and algorithmic instantiation of the processes suggested by the model) we believe that they are an important and under-used stepping stone towards that such final models. Abstract mathematical models can provide a great many constraints on building models that connect in more ways to empirical findings such as lesion data, imaging data, error data, and developmental data.

The second consideration motivated by synthetic approaches is to take variable interactions more seriously. Interactions may alert us to computational/neurological co-computation may be occurring or to aspects of our understanding where human-defined variables may be too neat.

Most of these goals would not be possible if they were not nested in the meta-goal of taking non-linearity seriously. We believe that the use of only linear methods in psychology conceals a great deal of the evidence we need in order to understand the processes we are studying. We may labor in vain for years to understand research puzzles using linear tools that might be trivial if we awoke ourselves from our dogmatic slumbers and allowed ourselves to use non-linear methods. Now that the most under-funded researcher is able to access computational resources that would have been unthinkable a few decades ago, it has become very easy to harness the power of nonlinearity.

We titled our chapter 'Putting Humpty Together Again'. Although they could not put Humpty together again, we can make some good guesses about what all the king's men were able to do. Examining the pieces of poor Humpty, they were certainly able to obtain a great deal of locally accurate information about each piece of him. They would have been able to document the weight, shape, and curvature of every piece. What they were not able to do, as we are told, was put those pieces back together. Traditional experimental psycholinguistics is in a similar situation to the king's men. It has taken us to a stage where we can say that this variable affects that behavior but not this one, whereas another variable has opposite effects, and a third affects both. This is local information about the pieces that may be parts of the much larger process of lexical access. What we cannot yet do is to put these pieces together to come up with an unambiguous, uncontroversial big picture of that process. In part this is because, unlike all the king's men, we do not even know if we have all the pieces, or the right pieces, or even if the pieces we have are all relevant to our larger goal. We believe that we will only be able to know this if we try to work backwards at the same time that we work forwards, synthesizing back into language the pieces that we have first so carefully analyzed out.

Baayen, R. (2005). Data mining at the intersection of psychology and linguistics. . In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones*. p. 69-83. Hillsdale, NJ: Lawrence Erlbaum Press.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995*). The Celex lexical database*. *release 2* (CD-aROM). Philadelphia, Pennsylvania: Linguistic Data Consortium, University of Pennsylvania.

Balota, D.A., Cortese, M.J., Hutchison, K.A., Neely, J.H., Nelson, D., Simpson, G.B., Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. http://elexicon.wustl.edu/, Washington University.

Blalock, H. (1972). *Social statistics*. NY: McGraw-Hill.

Carnap, R. (1936). Testability & Meaning, Parts I-III. *Philosophy of Science*, 3, 419-71.

Churchland, P.M. (1981) Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy* 78(2): 67-90.

Dawson, M. (2004). *Minds and machines: Connectionism and Psychological Modeling*. Malden, MA: Blackwell Publishing.

Holden, J. G. (2002). Fractal Characteristics of Response Time Variability. *Ecological Psychology*, 14(1&2): 53-86.

Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York, NY: Simon & Schuster.

Holland, J. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA: MIT press.

Hollis. G. & Westbury, C. (2006). NUANCE: Naturalistic University of Alberta Nonlinear Correlation Explorer. *Behavior Research Methods*. 8(1): 8-23.

Hollis, G., Westbury, C., & Peterson, J. (In press). NUANCE 3.0: Using Genetic Programming to Model Variable Relationships. *Behavioral Research Methods*.

Hull, C.L. (1943). *Principles of Behavior*. New York, NY: Appleton-Century.

Koza, J. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

Meehl, P. (1977). Construct validity in psychological tests. In P. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 3-31). New York, NY: W.W. Norton & Co.

Meehl, P. & MacCorquodale, K. (1991). On a distinction between hypothetical constructs and intervening variables. In C. Anderson & K. Gunderson (Eds.) *Paul E. Meehl: Selected Philosophical & Methodological Papers* (pp.249-263). Minneapolis, MN: University of Minnesota Press.

Minsky, M. and Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. MIT Press. Cambridge, Mass.

Quine, W.V.O. (1960) *Word and Object*. Cambridge, MA: MIT Press.

Reichenbach, H. (1938). *Experience and prediction*. Chicago, IL: University of Chicago Press.

Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review*, 45, 1-41.

Rorty, R. (1970). In Defense of Eliminative Materialism. *The Review of Metaphysics* XXIV.

Shaoul, C. & Westbury, C. (In press). Word frequency effects in high-dimensional co-occurrence models: A new approach. *Behavioral Research Methods.*

Vaihinger, H. (1935). *The philosophy of 'As if': A system of the theoretical, practical, and religious fictions of mankind* (2nd ed.). London, UK: Lund Humphries.

Van Orden, G., & Paap, K. (1997). Functional neuroimaging fails to discover pieces of mind in the parts of the brain. *Philosophy of Science*, 64, 85–94.

Van Orden, G., Pennington, B.F. & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*. 25(1): 111-172

Van Orden, G., Holden, J., & Turvey, M. (2003). Self-organization of Cognitive Performance. *Journal of Experimental Psychology: General*. 132:3, 331–350.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and 1/f scaling. *Journal of Experimental Psychology: General*. 134, 117-123.

Westbury, C. & Dennett, D. (2000). Mining the past to construct the future: Memory and belief as forms of knowledge. In: Schacter, D.L. and Scarry, E. (Eds.) *Memory, Brain, and Belief*. Cambridge, MA: Harvard University Press.

Westbury, C., Buchanan, L., Anderson, M., Rhemtulla, M., & Phillips, L. (2003). Using genetic programming to discover nonlinear variable interactions. *Behavioral Research Methods, Instruments, and Computers*, 28, 202–216.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford, England: Blackwell.

Figure 1: Figure 2. Graph across a range of variation of the two best-evolved functions relating orthographic neighborhood (ON) to frequency. Equation1 was sensitive to changes across the range of ON; Equation2 distinguished only between ON ≤ 2 and ON > 2. Reprinted from Westbury et al, 2003.

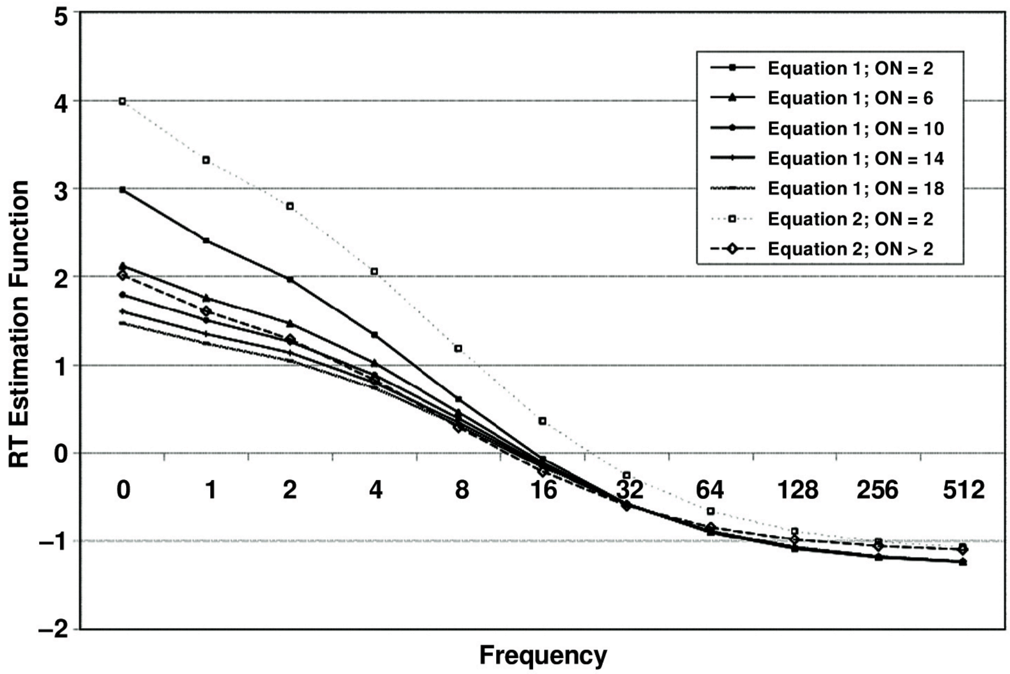
Figure 2: A comparison of the shape of the logarithmic and inverse transformations. Values have been normalized and the inverse curve flipped vertically in order to make the differences in their shape clear.

Table 1: The sixteen measures used to predict LDRTs.

Table 2: Table 2 Variance in LDRTs accounted for *on the validation set* by the 16 predictors, their log transformations, and their best-fit NUANCE transformation. All log and NUANCE-transformed effects significant at $p < 0.001$. For untransformed variables, p-values of 0.05, 0.01 and 0.001 denoted by *, **, and ***, respectively. Differences in predictive power between NUANCE-derived fits and best maximum of the other two fits are marked: p-values of 0.05, 0.01, and 0.001 denoted by †, ††, and †††, respectively (for the methodology used to determine significance values for correlational differences, see Blalock, 1972). Reprinted from Hollis & Westbury, in press.

Table 3: Reliable pair-wise interactions *in the validation set* among the 16 predictors of LDRT. P-values of $p < 0.05/120$, $p < 0.01$, and $p < 0.05$ denoted by ***, **, and *, respectively. Reprinted from Hollis & Westbury, in press.

Figure 1


5
4
3
2
1
0
−1
−2
RT Estimation Function
0
1
2
4
8
16
32
64
128
256
512
Frequency
Equation 1; ON = 2
Equation 1; ON = 6
Equation 1; ON = 10
Equation 1; ON = 14
Equation 1; ON = 18
Equation 2; ON = 2
Equation 2; ON > 2

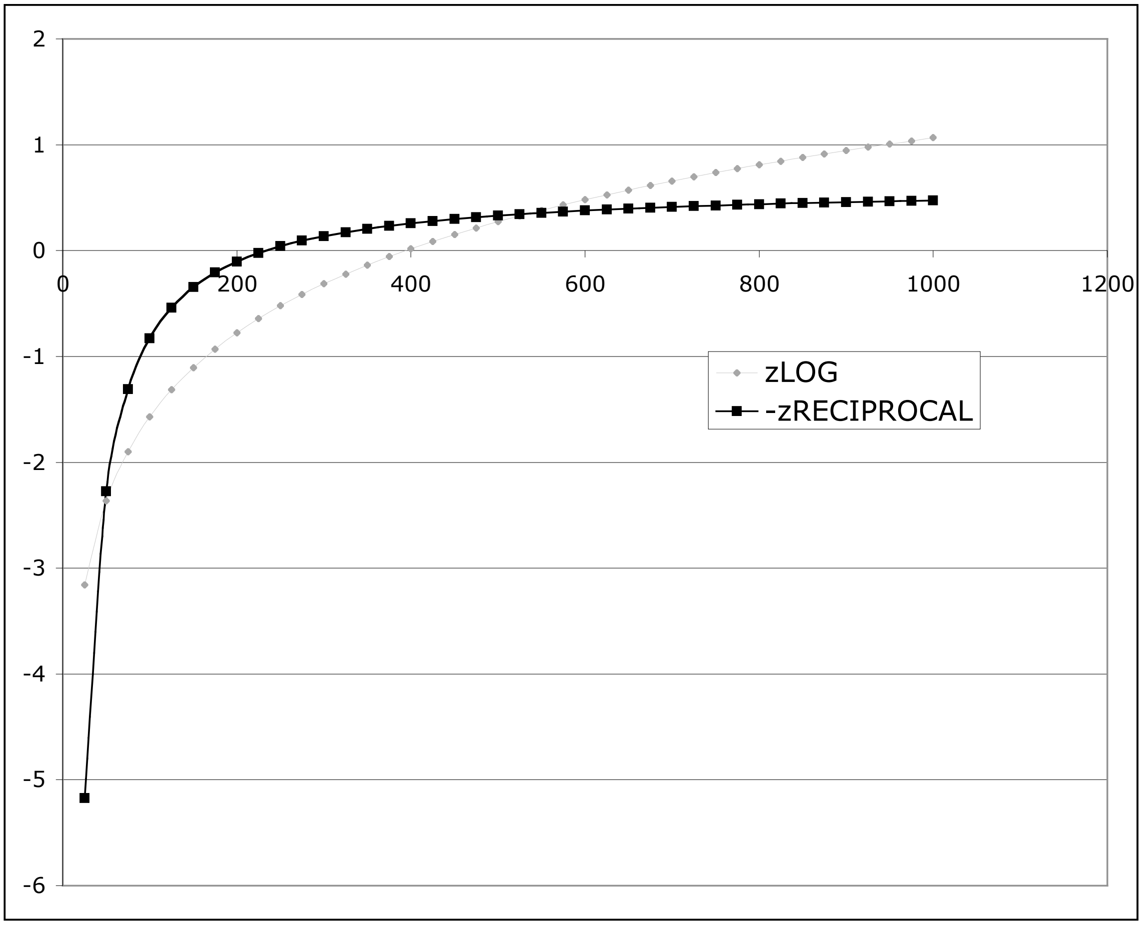

2
1
0
-1
-2
-3
-4
-5
-6
0
200
400
600
800
1000
1200
zLOG
-zRECIPROCAL

TABLE 1

Variable	Description
LETTERS	The word’s length, in letters
PHONEMES	The word’s length, in phonemes
OFREQ	The orthographic frequency (per million) of the word
ON	The number of orthographic neighbours of the word
ONFREQ	The average OFREQ of the word’s orthographic neighbours
PFREQ	The phonological frequency (per million) of the word
PN	The number of phonological neighbours of the word
PNFREQ	The PFREQ of the word’s phonological neighbours
CONBG	The summed frequency that any two letter-pairs in the word occur together in the place they are in for the current word. Only counted across words of the same length.
UNBG	The summed frequency that any two letter-pairs in the word occur. Position in the word and word length do not matter.
CONBP	The summed frequency that all two phoneme-pairs in the word occur together in the place they are in for the current word, only in words with an equal number of phonemes.
UNBP	The summed frequency that any two phoneme-pair in the word occur. Position in the word and phoneme count do not matter.
FIRSTTRI	The frequency with which the first three letters of the word occur as the first three letters for all words of the same length.
LASTTRI	The frequency with which the last three letters of the word occur as the last three letters for all words of the same length.
ARC	The average distance between a word and all of its semantic neighbours.
NN	The number of semantic neighbours the word has.

TABLE 2

Variable	Untransformed	Log Transform	NUANCE
OFREQ	0.015***	0.331	0.363††
PFREQ	0.002**	0.121	0.141†
LASTTRI	0.003***	0.072	0.131†††
FIRSTTRI	0.004***	0.092	0.115††
ON	0.078***	0.093	0.093
NN	0.065***	0.096	0.085
ONFREQ	0.000	0.045	0.076†††
PN	0.054***	0.072	0.066
ARC	0.039***	0.047	0.059
LETTERS	0.053***	0.048	0.059
PHONEMES	0.042***	0.039	0.048
PNFREQ	0.001*	0.027	0.048†††
CONBG	0.004***	0.008	0.025†††
UNBP	0.011***	0.011	0.018†
UNBG	0.006***	0.007	0.006
CONBP	0.001*	0.005	0.003

TABLE 3

Var. 1	Var. 2	R^2
LETTERS	PFREQ	0.015***
FIRSTTRI	LASTTRI	0.010***
ON	PFREQ	0.009***
CONBP	UNBP	0.008***
LETTERS	OFREQ	0.007***
CONBG	UNBG	0.006***
ONFREQ	PFREQ	0.006***
PHONEMES	ONFREQ	0.006***
ONFREQ	UNBP	0.005***
ONFREQ	PN	0.005***
LETTERS	ONFREQ	0.005***
OFREQ	ON	0.005***
PFREQ	NN	0.004**
ONFREQ	UNBG	0.004**

Var. 1	Var. 2	R^2
PHONEMES	PFREQ	0.004**
ON	ONFREQ	0.004**
OFREQ	PN	0.003**
PN	PNFREQ	0.003*
UNBG	UNBP	0.003*
PHONEMES	ON	0.003*
LETTERS	CONBP	0.002*
PN	UNBP	0.002*
PN	UNBG	0.002*
ON	NN	0.002*
PFREQ	UNBP	0.002*
PNFREQ	UNBP	0.002*
PFREQ	PNFREQ	0.002*