

What's in a word?

Re-tooling experimental methods and indices of abstractness

Zachary Stein

Final paper for H154: seminar on ERP research methods. Fall, 2008.
Harvard Graduate School of Education

Through the discussion of a set of Event Related Potential (ERP) studies I argue that we should re-build a certain set of experimental paradigms in psychology. I'm concerned with those that gather data through the use of linguistic stimuli, such as when a subject is asked to respond to sentences in an fMRI scanner or with electrodes attached to their scalp. The words populating an experimental psychologist's linguistic stimuli typically come from a limited number of psycholinguistic databases. These psycholinguistic databases consist of inventories of words—lexical inventories—organized and classified in terms of certain properties. They are valuable because psychologists need to choose words of specific types to control for variables they are seeking to isolate, such as word length or part of speech. I argue below that many existing lexical inventories were built in an under-theorized and unsystematic way; and they have come to, rather secretly, have a detrimental effect on our methods.

In particular, I am interested in *abstractness* as a property of linguistic constructions. Many standard psycholinguistic databases issue an intuitively determined degree of abstraction for each word in the lexicon, providing experimentalists with an abstractness-ranking for each word, ranging from relatively abstract to relatively concrete. The ERP studies I discuss address a phenomenon known as "the concreteness effect," which refers to the reliable differential responsiveness of individuals to linguistic constructions classified as relatively abstract or concrete. That is, results from a variety of studies show that "concrete" words are processed differently than "abstract" words. Importantly, the abstractness rankings used in this tradition of research come almost entirely from the standard psycholinguistic databases mentioned above, which thus ties the construct under investigation to a specific lexical inventory, one that I argue leaves much to be desired.

I begin with a discussion of the concreteness effect and its relation to the psycholinguistic databases that are most widely used for experimental paradigm construction. Then I review a set of EPR studies about the concreteness effect, discussing which aspects of their methods are wedded to these psycholinguistic databases and which are not. Finally, I suggest a viable alternative to existing lexical inventories organized around abstractness-rankings. I describe the research and theory behind a *developmentally organized lexical inventory* that is based on over a century of research in cognitive developmental psychology. I argue that it offers a reasonable, preferable, and readily available alternative for experimentalists looking for indices of linguistic abstraction. Along the way I offer a broad criticism of non-developmental approaches to psychological research, echoing the sentiments of James, Baldwin, Piaget, and Fischer, who suggest that the fragmentation of knowledge production in psychology is a symptom of our views of the mind as being stationary in its synchronic and modular organization. Moreover, I argue that re-tooling specific experimental paradigms as I recommend would allow the concreteness effect to be re-framed in terms of the tradition of cognitive developmental psychology.

The concreteness effect and the psycholinguistic databases that created it

The concreteness effect has been the focus of psycholinguistic and psychophysiological research for decades. It was first brought to prominence by the Canadian psychologist, Allan Paivio (1969; 1991). His work surrounding "dual coding theory" was the first systematic documentation and explanation of the effect, and remains the most widely cited and endorsed position on the issue. Paivio demonstrated, by using reaction-time studies and other behavioral measures, that "concrete" words function differently than "abstract" words in a variety of experimental conditions. In particular, concrete words (words with referents that can be literally be seen or imaged, such as *chair*) are encoded, retrieved, and comprehended faster than abstract words (words with referents that are general or complex, such as *honesty*). In light of comparable findings from various research groups Paivio suggested an explanation by posting

the existence of two separate but related language processing systems, one tied to visual and sensorimotor systems, the other being strictly linguistic and symbolic. On this view the concreteness effect is the result the "cognitive processing advantages" of the imagistic and concrete language processing system, which is robustly geared into the world of simple and actual referents.

As the ERP studies discussed below indicate, there have been experimental follow-ups on various fronts. While these have clearly confirmed the existence of the concreteness effect they have shed only minimal light on the feasibility Paivio's proposed explanation. While goal of this paper is not to offer a better explanation of the concreteness effect—but rather to take up some relevant methodological issues that bear on the nature of the findings in question—I will say that the dual coding theory proposed by Paivio is typical of non-developmental and non-dynamic approaches to psychological research.

As Piaget (1970) suggested in some of his late meta-theoretical reflections, the most basic (in the sense of least advanced) form of psychological explanation is the *post hoc* positing of synchronically interlocking mechanisms to account for a set of more or less arbitrarily gathered experimental findings; i.e. we measure what we can and explain it as simply as possible. This leads to a psychological science in which each set of findings has its own explanation, without any overarching comprehensive view of things. Importantly, Piaget saw this type of psychology as a phase following in the wake of advances in measurements or methods, a phase followed upon by attempts at synthesis and a search for broader explanatory constructs, typically of a developmental nature. This was the kind of groping fragmented science that James (1890) took the psychological of his time to be, as basic advances in physiology and psychophysiology ushered in a new era. And it is the kind of cognitive science we see today, as our theories strain to keep up with advances in brain science methods.

This aside is pertinent because Paivio's work on the concreteness effect followed hard upon certain methodological advances, specifically the early stages of the codification and

standardization of the Medical Research Council Psycholinguistic Database (MRCPD; see: Coltheart, 1981). In fact, Paivio was one of a handful researchers directly responsible for the construction of the database, which set the bar for comparable attempts at standardizing lexicons for—among other things—language based experimental stimuli. A network of less than half a dozen comparable databases exists, none of which have moved far beyond the research behind the MRCPD. These have been compiled into the Oxford Text Archives and made available as computerized lexical inventories.

These databases serve the function of cataloguing words in terms of their key properties (Wilson, 1988). Some word-classifications are uncontroversial, such as: *number of letters in the word*, or *number of syllables in the word*. Other classifications are much less straightforward, such as: *meaningfulness*, *familiarity*, and *abstractness*. It is these later types of word-classifications—specifically abstractness—that I am taking issue with. The issue is not whether words have properties such as meaningfulness, familiarity, or degree of abstraction. My plan is, in fact, to suggest a better way to classify words in terms of their degree of abstraction. But as it stands, these not uncontroversial categories of the database leave much to be desired.

The lexicons organized around the three contentious word-properties listed above are mainly built from the findings of two seminal psycholinguistic studies: Paivio's (1968) monograph supplement to the *Journal of Experimental Psychology* and the *Handbook of Semantic Word Norms* by Toggia and Battig (1978). Both studies used the same methodology: (literally) sit a large number of college sophomores down, provide them a large list of words and a short list of questions about each word, run some inter-rater reliability numbers on the results, and re-organize the list of words in terms of the classifications generated by the participants. Thus, and to get to the heart of what this paper is about, the abstractness-ranking of any given word in the MRCPD simply reflects the collective intuition of a lecture hall full of college sophomores in the 1960s. Now, I would be willing to argue that *any* widely used standardized

psycholinguistic database is better than *none* at all. But clearly, this way of determining which words should populate which categories is sub-optimal.

Once again, it seems that taking a static and strictly synchronic view of the mind has led to a certain type of undesirable simplicity in the adoption of a methodological approach. Since James Mark Baldwin (1906) first wrote on the development of cognition, language, social interaction, and knowledge, the insight that the *same word*, event, or task can mean *different things* at different stages of development has been continually reconfirmed (e.g., Piaget, 1932; Case, 1992). Moreover, *abstraction* in particular has been identified as one key dimension of the developmental process itself, i.e., the course of development can be seen as a hierarchy of increasingly abstract conceptual abilities and action-orientations (Baldwin, 1906; Piaget, 1972; Fischer & Bidell, 2006).

Thus, from the perspective of a developmentalist, the strategy adopted by Paivio, Toglia, and Battig for empirically grounding the organization of a lexicon in terms of abstractness-rankings is wrongheaded. On the one hand, a developmentally homogeneous group of raters will drown-out developmentally variable aspects of polysemy that bear crucially on abstractness-rankings. As Piaget (1926) demonstrated in his first book, vocabulary (both in terms of the words you use and what they mean) varies radically from one developmental level to the next; and these variations correlate with variations in the abstraction of conceptual abilities and action orientations. What's in a word—what it means and how much abstraction it contains—varies as a function of development. Below I argue that an empirically grounded lexicon organized by abstractness-rankings should be (can be, and has been) generated from cognitive developmental data sets. However, before I explain my alternative and its implications, I want to tie up this section and move on to the next.

In light of what has just been discussed one could claim that Poivio's work on the concreteness effect and of most the subsequent follow-ups are *merely* artifacts of Poivio's own lexical inventories. I am not looking to make that kind of simple argument. What I am suggesting

is that while certain aspects of the concreteness effect are best understood as artifacts of Poivio's under-theorized approach to determining word-abstractness rankings, there is nevertheless a moment of truth in this family of findings, which has been obscured by certain non-developmental theoretical commitments and perpetrated by the limitations of existing lexical inventories. The concreteness effect is too simple, in part, because of the simplicity of the available lexical inventories cataloguing abstractness-rankings for use in experimental paradigm design.

I argue below that an appropriate use of lexical inventories with abstractness-rankings garnered from cognitive developmental research could allow us to refine our findings concerning the nature of the concreteness effect and re-frame them in terms of the broader explanatory principles of the cognitive developmental tradition. Briefly, cognitive developmental psychologists have demonstrated for decades that there are at least three (with a probable fourth) major re-organizations of action and thought (Fischer & Bidell, 2006). While these tiers, and the levels within them, can be characterized as ranging from concrete to abstract, detecting their manifestation in linguistic performances requires more than a gross differentiation between concrete and abstract word-types. The development of thought and language is best characterized as a *spectrum* of abstraction and complexity, not as a simple distinction between abstract and concrete. In the next section I discuss an exemplary set of ERP studies concerning the concreteness effect to show how these paradigms could be (and should be) re-tooled with a new lexical inventory that would effectively refine and re-frame the findings, bringing them into the fold of a century-old research tradition.

ERP indexes of the concreteness effect, and the words that evoke them

In 1990's a research group lead by Tufts University's Phillip Holcomb produced a trio of studies on the concreteness effect (Holcomb, Kounios, Anderson, & West, 1999; Kounios, & Holcomb, 1994; West & Holcomb, 2000), which are exemplary in their rigor and elegance. While other good contemporary ERP studies of the concreteness effect exist (e.g. Koenig, Kochi, &

Lehmann, 1998; Lee & Federmeier, 2008), they will not be discussed; my goal here is not to present a literature review. I am addressing only the exemplary trio in order to make broader theoretical and methodological points. However, importantly, my behind-the-scenes literature review leaves me convinced that the studies discussed here are, in fact, exemplary, i.e., the points I raise concerning them apply to a large set of ERP studies concerning related issues.

There are a limited number of ways to investigate the concreteness effect using ERP methods. In all cases subjects are presented stimuli consisting of time-locked, systematically varying linguistic constructions, about which they are asked to make judgments. The key dimension of variability in the stimuli is, of course, the abstractness or concreteness of the words presented. For example, in the earliest of the three studies under discussion (Kounios, & Holcomb, 1994), two experimental conditions were designed, one involving a lexical decision task, the other involving a concrete-abstract word classification task. The former task involves the presentation of letter-strings (non-words) and actual words (varying in terms of abstractness and concreteness) with subjects asked to indicate which stimuli are words and which are non-words via button pressing. The second experimental condition involves only the presentation of actual words, again varying in terms of abstractness and concreteness, with subjects asked to classify the presented words as either concrete or abstract via button pressing. The results were similar across conditions, displaying the classic reaction-time "cognitive processing advantages" for concrete words, and differential ERP wave forms around N400 for responses to concrete and abstract respectively. Importantly, in both cases the determination of which words were abstract and which were concrete was made via reference to the lexical inventories cited above, specifically the Toggia, and Battig (1978).

This simple paradigm shows the general pattern of how existing psycholinguistic databases interface with experimental paradigm construction. What is so striking—and thus entails the writing of this paper—is that in studies of the concreteness effect the experimental stimuli are meant to be geared into key properties of the construct under investigation, yet they

are simply lifted *ad hoc* from existing lexical inventories. Nowhere is there anything but a superficial discussion of how distinctions between abstract and concrete words should be characterized, nor is there a discussion about how such a distinction could be operationalized to create an empirically grounded lexicon organized in terms of it. This is the kind of basic (in the sense of not sophisticated) psychology I lamented above, echoing Piaget and James. An under-theorized approach to the determination of linguistic abstraction leads to a simplistic experimental paradigm and to results that reconfirm the *post hoc* stipulation of a simple synchronic cognitive architecture, in this case one with intermeshing abstract and concrete language processing modules.

Unfortunately, the ostensibly more sophisticated paradigms from the other two papers in the trio fair no better in this respect. For example, West and Holcomb (2000) build a complex paradigm in order to control for various factors and more carefully test certain aspects of Poivio's dual coding theory. They randomly assign subjects to three conditions wherein they are asked to respond to the truthfulness of sentences. In different conditions subjects get exposed to different sentence-stem-types, with abstract or concrete final words:

Imagery type (respond true or false): *It is easy to form a mental image of tables/bravery.*

Semantic type (respond true or false): *It is common for people to have tables/bravery.*

Surface type (respond true or false): *There is an "r" in the word table/bravery.*

Their thought is that these are distinct tasks that pull for different facets of the concreteness effect. The first sentence-stem-type engages imagery, the second engages semantics, while the third engages only the surface features of the words. Their hypothesis is (and their findings confirm) that the ERPs for the imagery and semantic tasks will show significant differences around N400 for concrete and abstract words while the ERPs for the surface task will show no significant differences. Furthermore, they suggest (and find) that there will be significantly unique scalp distributions for the ERPs from the imagery and semantic tasks.

This paradigm is elegant and greatly improves upon the simpler lexical decision and classification tasks used in the first study. But like the first study, the heart of the paradigm is taken *ad hoc* from the Oxford Text Archive. Again, the construct under investigation is handled far too simply and nonchalantly, and the lineage of research remains wedded to a-theoretical lexical inventories generated in the 1960s.

The problem with these paradigms is not their internal validity; they are rigorous and justifiable in and of themselves. The problem is construct validity. *Abstractness* remains relatively undefined. It is used in an everyday sense and with nowhere near the specificity needed from a psychologically operationalizable construct. Thus the methods function under the assumption that abstractness-rankings for specific words can be, and have been, unproblematically generated. The final results are experimental paradigms built around a specious construct. These paradigms in turn generate findings that support certain simplistic and untenable theoretical ideas about how the mind works. Re-tooling these paradigms requires building both a more rigorous construct and a new set of lexical inventories based on this revision. Below I suggest that we can look to the tradition of cognitive developmental psychology for both, which upon delivery may just resuscitate the relevance and importance of something like the concreteness effect.

Re-framing the concreteness effect and the paradigms that created it

I have shown that the concreteness effect should be understood as a group of experimental findings wedded to a specific set of lexical inventories, which themselves are based on an inadequate operational definition of *abstraction*. Moreover, putting on the hat of a staunch developmentalist, I have suggested that current explanations of the concreteness effect (e.g., Paivio's dual coding theory) are inadequate because they merely describe an existing structure without accounting for its genesis. Here I will argue that we can remedy both of these problems by radically changing the nature of the lexical inventories used in designing experiments about the concreteness effect.

My task is greatly eased by the fact that the type of lexical inventory I am recommending already exists. Theo Dawson and colleagues (Dawson & Wilson, 2004; Dawson-Tunik, Commons, Wilson, & Fischer, 2005) generated a unique lexical inventory during the course of conducting basic research in cognitive developmental psychology. This lexical inventory consists of words classified in terms of the *developmental level* at which they appear. The inventory was produced during efforts at refining the psychometric precision with which linguistic performances can be assigned a developmental level score. Below I explain the background needed to understand how this lexical inventory was produced and why it is valuable. Then I offer suggestions on how it should be used in experiments like the ones discussed above and speculate about the implications of its use.

Developmental assessment systems (often referred to as scoring systems) have been around since Piaget first pioneered the use of qualitative interview methods in developmental psychology. In a series of now classic studies Piaget (1926; 1929) demonstrated how to look for and discover the structural properties of linguistic performances that are indicative of development. The developmental analysis of qualitative interviews and linguistic performances continued with Kohlberg's work in moral development and is alive and well today.

As Piaget's example shows, these types of analyses entail looking for the properties of linguistic performances that are indicative of their development. Different researchers take different approaches to this problem, taking different properties as indexes of development. Kohlberg and many others focus on the types of *conceptual content* that appear at different levels, and determine the level of a performance by identifying the presence of specific concepts (such as "the value of the law" or "the golden rule"). Dawson and colleagues set out differently, and focus on the *deep structures* of the performances. Thus the criteria by which they determine the level of a performance is not tied to the conceptual content of the domain or skill in question. Their developmental analyzes focus on universal properties indicative of development, which have been distilled in the wake empirical research and model building.

In 1980, Kurt Fischer first explicitly modeled the basic developmental process of *hierarchical integration*, which had been an important but relatively implicit construct in developmental theorizing for nearly a century. This clarification cleared the way for new analytic techniques for assessing development by focusing on the differentiation, integration, abstraction, and complexity of performances, as opposed only their conceptual content. In the 1990s Dawson employed a variety of psychometric techniques to refine the analytical techniques that unfolded in the wake of Fischer's model, and the *Lectical Assessment System* (the LAS) was devised (Dawson, 2008; Dawson-Tunik, Commons, Wilson, Fischer, 2005)

The LAS is the first and only psychometrically calibrated domain general developmental assessment system. In essence, the LAS provides strict criteria for determining the developmental level of a linguistic performance by focusing on the core structure of abstraction and complexity. Framed by Fischer's model, the LAS identifies at least 7 developmental levels in which language play a key role: 3 levels of basic action-orienting concrete concepts, known as *representations*, and three levels reflective intangible ideas, known as *abstractions*, up to one level (with two projected) of discourse-regulative overarching constructs, known as *principles*. This general skill-scale is a hierarchy of increasing complexity, abstraction, and capability, which has been used in dozens of studies on various fronts (Fischer & Bidell. 2006).

Two recent outgrowths of this tradition of research and theory culminated in the generation of the *developmentally organized lexical inventory* (the DOLI). Both hinge on the fact that Dawson has compiled the largest database of qualitative interviews ever analyzed for developmental purposes. This massive database contains literally thousands of pages of linguistic performances representing the full spectrum of capabilities from *representations* through *abstractions* to *principles*. While conducting psychometric reliability studies and working to build a computerized and automated developmental assessment, Dawson distilled the DOLI by running her entire database through what is essentially an algorithmic sieve that correlates—among other things—word frequency and level score.

Thus, in sum, a long tradition of cognitive developmental research and theory has resulted in a sophisticated form of developmental assessment—Lexical Assessment—that can be used to analyze linguistic performances and assign them a level score. This method was used to code an unprecedented number of qualitative interviews, from which the DOLI was distilled. This makes the DOLI a uniquely sophisticated lexical inventory, and importantly for my purposes here, it is one built around a sophisticated operational definition of *abstraction*.

Recall that *abstractness* is more or less the topic of one key family of explanatory and descriptive constructs in cognitive developmental theory. This property of cognitive developmental processes has been referred to in a number of ways: e.g., *reflecting abstraction* (Piaget, 1977), *hierarchical integration* (Fischer, 1980), and *order of abstraction* (Dawson, 2008). Importantly, as explained above, the LAS specifically operationalizes this family of constructs about abstractness, offering analytical techniques for in engaging with the developmental properties of linguistic performances. My claim here is that because the LAS was used to generate the DOLI, it (the DOLI) should be considered the most sophisticated, theoretically robust, and empirically grounded lexical inventory organized in terms of abstractness-rankings for specific words. In fact, in the DOLI each word is placed in an empirically specified position along a psychometrically valid scale for measuring abstraction. There are 7 levels that the LAS can specify, and thus the DOLI is organized around 7 valid and reliably distinguished degrees of abstraction in terms of which to rank words.

Briefly, and to come full circle, compare this method for generating an abstractness-rankings with the methods used by Poivio and colleagues in generating the MRCPD. Two concerns that have been noted before resurface in this comparison: (1) the MRCPD was generated relatively a-theoretically (see page 4 above), whereas the DOLI is the result of a century of research and theory, and; (2) the MRCPD downplays the developmental issues implicit in key categories in the lexical inventory (see page 5 above), whereas the DOLI is explicitly framed in terms of these developmental issues. Both of these points suggest to me

that the DOLI is preferable; or at least they lead me to assert that the DOLI was built using a more valid operational definition of *abstraction* (see page 9 above). Thus my motivating thought for this whole paper: the DOLI should replace the lexical inventories currently being used to research the concreteness effect.

This move would have relatively simple methodological consequences, e.g., the ERP paradigms discussed above could be outfitted with new word-abstractness rankings. But the theoretical implications would be very complex and, from where I sit, the concreteness effect could be re-framed and flushed out terms of developmental explanatory constructs. For example, we could re-tool the kind of sentence-stem paradigm discussed above with words given three distinct abstractness rankings (as opposed to the two currently in use). While the details of stimuli presentation order, item frequency, and number of trials would need to be augmented due to the addition of a third construct-relevant stimuli, the basic principles would remain the same. The patterns of reliable differential responsiveness indexed by ERP waveforms would be likely to vary also, now displaying three distinct patterns of activation.

That last sentence represents the kind of hypotheses we could now test with the DOLI in hand. It is an open question as to whether we can, in fact, detect finer grained responsiveness differentials that correlate with finer grained abstractness-rankings. Looking into this kind of question is one research area—and there are many—that would bring questions from cognitive developmental psychology into the reach of ERP methods via the adoption of the DOLI in experimental paradigm design. This is a second reason for seeking to replace existing lexical inventories with the DOLI: it would facilitate the incorporation of more diverse experimental methods into the fold of cognitive developmental theory and research. The important directions for and implications of this research are beyond the scope of this paper.

What's in a word: conclusions about re-tooling our methods

I have argued that a new kind of lexical inventory should be adopted for use in certain types of experimental paradigm construction. I have two reasons: the preferable inventory is

more valid (it was generated by better methods and theory) and the preferable inventory is more generative (it allows for the interface of cognitive development psychology and a variety of methods). This second reason is the one I find most compelling. Throughout the paper I have roundly criticized non-developmental views of the mind. I suggested that Poivio's "dual coding theory" should be understood as symptomatic of general trends in psychological theory building, e.g., the *post hoc* positing of synchronically interlocking mechanisms to account for a set of more or less arbitrarily gathered experimental findings. This paradigmatically parochial form of psychological theorizing—wherein each new set of findings gets a new explanation—is often the result of limitations of methods that force us to dissect the mind piecemeal.

One way of thinking about what I have suggested here is as a methodological solution to this problem of intra-disciplinary fragmentation. I am suggesting that we insert the cumulative findings of an entire research tradition into the next generation of experimental apparatuses used by that tradition. This is exactly the kind of cumulative, iterative, and integrative methodological innovation, taken by Piaget (1970) to be the *modus operandi* of a mature psychological science.

Sources:

- Baldwin, J. M. (1906). *Thought and things: A study in the development of meaning and thought or genetic logic* (Vol. 1-3). New York: Macmillan Co.
- Case, R. (1992). *The minds staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Lawrence Erlbaum.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly journal of experimental psychology*, 33a(4), 497-505.
- Dawson, T. L. (2008, 11/1/04). The Lectical™ Assessment System. 1. Retrieved September, 2008, from <http://www.lectica.info>
- Dawson, T. L., & Wilson, M. (2004). The LAAS: A computerized developmental scoring system for small- and large-scale assessments. *Educational Assessment*, 9, 153-191.

- Dawson-Tunik, T. L., Commons, M. L., Wilson, M., & Fischer, K. W. (2005). The shape of development. *The International Journal of Cognitive Development, 2*(163-196).
- Fischer, K., & Bidell, T. (2006). Dynamic development of psychological structures in action and thought. In W. Damon & L. R.M. (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. One, pp. 1-62). New York: John Wiley & Sons.
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual coding, context availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of experimental psychology: learning, memory, and cognition, 25*, 721-742.
- James, W. ([1890] 1950). *The principles of psychology* (Vol. 1-2): Dover Publications, Inc.
- Koenig, T., Kochi, K., & Lehmann, D. (1998). Event-related electrical microstates of the brain differ between words with visual and abstract meaning. *Electroencephalography and clinical neurophysiology, 106*, 535-546.
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of experimental psychology: learning, memory, and cognition, 20*(804-823).
- Lee, C.I., & Federmeier, K. D. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language, 104*(2), 145-158.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological review, 76*, 241-263.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255-287.
- Paivio, A., Yuille, J. C., & Madigan, A. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology monograph supplement, 76*(1).
- Piaget, J. (1926). *The language and thought of the child*. London: Lund Humphries.
- Piaget, J. (1929). *The child's conception of the world*. London: Routledge and Kegan Paul.
- Piaget, J. (1932). *The moral judgment of the child*. London: Routledge and Kegan Paul.
- Piaget, J. (1970). *Main trends in psychology*. New York: Harper & Row.
- Piaget, J. (1972). *The principles of genetic epistemology* (W. Mays, Trans.). London: Routledge & Kegan Paul.
- Piaget, J. (2001 (1977)). *Studies in reflecting abstraction* (R. L. Campbell, Trans.). Philadelphia, PA: Psychology Press.
- Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Somerset, NJ: Lawrence Erlbaum.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, 12(6), 1024-1037.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioral Research Methods, Instruments and Computers*, 20(1), 6-11.