DEVELOPING A COMPUTER CODING SCHEME FOR

THE IMPLICIT ACHIEVEMENT MOTIVE

A Dissertation presented to the Faculty of the Graduate School

University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

MARC HALUSIC

Dr. Kennon M. Sheldon, Dissertation Supervisor

JULY 2015

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

DEVELOPING A COMPUTER CODING SCHEME FOR

THE IMPLICIT ACHIEVEMENT MOTIVE

presented by Marc Halusic,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Kennon M. Sheldon

Professor Laura A. King

Professor Laura D. Scherer

Professor Todd R. Schachtman

Professor Philip Robbins

ACKNOWLEDGEMENTS

I would like to thank my advisor, Kennon M. Sheldon, my lab mate, Mike Prentice, and my committee members, Laura A. King, Laura D. Scherer, Todd R. Schachtman, and Philip Robbins for their guidance and support in the process of conceptualizing and implementing this research. Additionally, I am indebted to both David G. Winter and Ted Pedersen for providing me with more help than I could have hoped for when I first introduced myself to them over email. To David G. Winter, I am particularly thankful for his help in learning about the history of motive coding and for challenging me to think more carefully about the intricacies of automating motive coding. To Ted Pedersen, I am particularly thankful for his help in planning how to adapt the WordNet::Similarity computer program to compute the statistics that were crucial to this project.

Finally, this research was largely based on the analysis of archival data, which was generously shared with me by the following researchers: David G. Winter, Oliver C. Schultheiss, and Steven J. Stanton. Further archival data was generously shared with my by the Henry A. Murray Research Archive.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figure

Table

LIST OF ABBREVIATIONS

LIWC	Linguistic Inquiry and Word Count
MSSR	Maximum Synset to Sentence Relatedness
NAch	Need for Achievement
PSE	Picture Story Exercise
SR-AW	WordNet::SenseRelate::AllWords
TAT	Thematic Apperception Test

Abstract

Implicit motives are measured using a projective assessment, the Picture Story Exercise (PSE), involving labor-intensive coding of participant-generated writing. The present research uses insights from previous attempts to automate coding, as well as advances in natural language processing and machine learning, to create a new method of automated coding for the achievement motive (NAch). In part 1, I collected coded PSE sentences from implicit motive researchers. Two models were generated using multilayer perceptron neural networks to predict achievement motive imagery, one using the Linguistic Inquiry and Word Count (LIWC; Pennebaker, 2001) software, and one using a novel text processing system, called Maximum Synset-to-Sentence Relatedness (MSSR). Part 2 sought to experimentally manipulate NAch, and produce 2 more neural network models similar to those of part 1, except that the models in this case predicted experimental condition. Further, human generated NAch scores from the PSEs collected in this part were compared against computer generated NAch scores produced by the models from part 1, to provide another test of the magnitude of the relation between human and computer generated NAch scores. Part 3 tested all 4 models to predict achievement motive imagery in archival data collected by Ratliff (1979). Because these data were coded using a different NAch coding scheme, and also included other variables theoretically related to NAch, these tests were used to search for evidence of convergent and predictive validity. Findings were promising for both models developed in part 1, but further improvements will be necessary before they can replace human coders.

Introduction

Overview

Daily experience can motivate people in a number of qualitatively distinct ways. It is common that people will seek out and master challenges, engage in "warm" social interactions, or take pleasure in the act of influencing others. Just as situations can differ in the quality of motivation they foster, people can also differ in the forms of motivational experiences that attract their attention, energize their behavior, and lead to emotional satisfaction. These individual differences are perhaps best measured using a carefully developed form of projective assessment, called the Picture Story Exercise (PSE), which involves a standardized procedure for people to write a set of creative stories inspired by a set of images, which can then be coded for motive imagery. Unfortunately, the PSE is a particularly onerous measures for researchers to use, with motive imagery coding requiring that each story be carefully read and coded by an extensively trained team of research assistants. Due to the onerous nature of coding the PSE, some researchers (Litwin, 1965; Schultheiss, 2013; Smith, 1968) have made efforts to use computers to automate the process. Those efforts have been promising, but none has approached a level of performance that could justify replacing human coders with computer algorithms.

In the present paper, I argue that advances in the computer sciences of natural language processing and machine learning can be used to improve upon those efforts, and I present original research attempting to do just that. I first introduce the constructs that are measured by the PSE, called implicit motives, and explain why they are psychologically interesting and important. I then explain how the PSE coding schemes were developed, which is necessary in part to consider ways that an automated coding scheme might also be developed, and in part to shed additional light on what exactly is measured by the PSE. I then outline the history of attempts to automate motive coding using computers, followed by an in depth introduction of several areas of modern natural language processing techniques that I contend should improve the accuracy of computer motive coding algorithms. Next, I introduce a method of generating a prediction algorithm using machine learning that should further be an improvement on previous methods. Finally, I outline a 3-part research program that I used to create and validate several algorithms that use computers to assign Achievement motive imagery to PSE stories.

The Implicit Motive Construct

Implicit motives are social motives, such as affiliation and achievement, that uniquely predict spontaneous behaviors, long-term behavioral trends, and other outcomes related to the non-deliberative components of motivated behavior. For example, the implicit achievement motive (called NAch) has been found to predict entrepreneurial activity over time in both the United States (McClelland, 1965) and India (McClelland, 1987), the intimacy motive measured at age 30 has been found to predict marital satisfaction 17 years later (McAdams & Vaillant, 1982), and the affiliation motive has been found to predict the likelihood that participants were engaged in conversation when randomly beeped (McClelland, 1985).

It is perhaps surprising given the relative sophistication of computer-based reaction time measures for other forms of implicit measures (e.g. Greenwald, Nosek,

& Banaji, 2003; Stacy & Weirs, 2010) that social motives would still be measured in a manner that is reminiscent of Freudian psychoanalysis. In order to understand why this tradition of motive assessment has produced such fruitful measures, it is necessary to understand the techniques involved in forming them.

The Creation of Implicit Motive Measures

This research tradition was born from David McClelland's focus on human motivation involving questions derived from psychodynamic and personological theory, but involving methods informed by Hullian animal research (Winter, 1998). Using Hullian animal research as a model meant that the first step of studying a motive was to experimentally arouse it, and compare the effects of that arousal against a control group. The exact nature of the effects of motive arousal were unknown, so the next step in this process was to use a large subsection of the data for qualitative analysis, looking for patterns that differentiated the experimental group from the control group. Once those patterns in the data were identified, they were used to try to discriminate between participants whose data had not been used for qualitative analysis.

From that starting point, McClelland tested candidate measures of motive arousal starting with one very simple and easily manipulated motive: hunger. Reasoning that hunger is undoubtedly a motive, and further that it is a motive that is easy to manipulate, McClelland sought to identify assessment techniques that could differentiate between hungry and sated participants.

The earliest attempt at creating such a measure involved a study of perception, asking participants to identify the objects presented in subliminally projected slides (many of which were blank). Few differences emerged between groups, so McClelland moved from perception to apperception, using a variant of the Thematic Apperception Test (TAT) which differed from the original TAT in that the images involved were meant to represent mundane experiences rather than complex psychodynamic constructs (e.g. the Oedipus complex). The images in this version of the TAT were chosen to be evocative of food without being directly about food, such as a picture of four homeless people sitting together. As such, stories written about these images could be related to food and hunger, but could be about other topics as well.

By comparing the stories written by people at varying levels of food deprivation, several important points emerged. First, different themes were present in the hungry vs. sated groups, with hungry participants writing more about themes such as food deprivation and instrumental activities to overcome obstacles to food, but also involved less discussion of consumption of food. This suggested that, unlike the earlier perception study, the combination of the motive manipulation procedure and the TAT was a viable model for producing experimentally derived motive measures. Second, it demonstrated that simple, a-priori coding schemes are unlikely to be successful. For example, simple counts of food-related words were insufficient to differentiate the hungry and sated groups. Rather, the elements in a story that indicate the arousal of a motive must be identified rather than guessed. Third, it showed that early conceptions of projective assessment—that imagery functions to satisfy unmet needs—was unlikely to be entirely true because hungry participants wrote less about consuming food.

Having demonstrated the potential of the motive arousal technique (sometimes called motive induction) for producing empirically derived TAT coding (later called Picture Story Exercise to differentiate it from the earlier version of the TAT due to the different set of pictures and experimentally derived coding schemes), McClelland, Atkinson and colleagues produced motive scoring systems for several social motives, most notably achievement, affiliation and power. The achievement coding technique, for example, was derived from comparison of PSE stories written by students who completed a set of anagrams under one of several sets of task instructions. Several sets of instructions were meant to be ego-involving, emphasizing the anagram task as a measure of ability. The ego-involving conditions differed from each other in terms of the task norms cited for participants, with one condition including norms that were so high as to ensure that participants felt that they were performing poorly (called the failure condition), another set with norms so low as to ensure that participants were performing well (called the success condition) and a third with norms that varied so as to produce a sense of initial success followed by failure. The control group completed the same anagrams, but were told that the experimenters were simply "trying out" some new measures, suggesting that no meaningful norms existed. Participants then completed the PSE. In a finding that was unexpected to the experimenters, the failure condition did not lead to specifically deprivation-focused imagery and success did not lead to specifically satisfaction-focused imagery. Rather, all ego-involved conditions produced similar imagery, and that imagery differed from that produced by the participants in the control group.

Development of a More Modern NAch Coding Technique

In the original coding technique for NAch developed by McClelland et al. (1949), each story was coded using a checklist procedure. For example, if a story contained statements about a unique achievement, it would accrue one point for that type of imagery, regardless of the number of times such imagery occurred. Other types of imagery, such as positive anticipatory goal state, or an overall sense that achievement forms the central plot of the story (called "thema") would similarly each contribute a single point to a story's imagery score. David Winter (1991) sought to simplify this and other motive coding systems, making changes that would make motive coding procedures more simple and clear. One major departure from McClelland's original coding system is that the Winter system does away with the checklist coding procedure, instead allowing a given type of motive imagery to be scored once every other sentence,[1] although other major changes exist as well. Winter (1991; 1994) demonstrated that this coding scheme provides convergence with coding from the original NAch coding scheme, and even is able to discriminate between the experimental and control groups in McClelland et al.'s (1949) original motive induction study. It has since become a popular method of coding NAch.

Interpretations of the Implicit Motive Measure

The decades following the development of implicit motive coding schemes produced a large body of consistent findings, suggesting that implicit motives

[1] This rule can only be broken if a different type of imagery exists between two instances of the same type of imagery. For example, if a sentence has an instance of achievement imagery, followed by affiliation imagery, a subsequent instance of achievement imagery may be scored.

uniquely predict psychologically relevant outcomes such as long-term behavior trends, hedonic impact of motive congruent goals, the content and structure of memory, and health outcomes such as cardiovascular disease (Brunstein, Schultheiss, & Grässman, 1998; McClelland, Koestner, & Weinberger, 1989; Spangler, 1992; Woike & Pollo, 2001). Such empirical findings beg the question of why the PSE story content would hold information rich enough to predict this broad array of outcomes.

According to the psychodynamic perspective, projective assessments should reveal nonconscious processes because they allow for the expression of repressed impulse in a way that is not personally threatening. The idea of expressing repressed impulses is rooted in Freudian dream analysis (Freud, 1999) and might be taken to be the inspiration of modern projective assessments such as the Multi-Motive Grid (Sokolowski, Schmalt, Langens, & Puca, 2000). At least two lines of reasoning suggest that this defensive mechanism is likely untrue. One line of reasoning concerns McClelland's failed attempt to allow participants to project their needs onto subliminally presented images. That methodology should have provided participants with an ideally unthreatening situation to express suppressed needs, but nonetheless proved ineffective. A second line of reasoning concerns the nature of the drives studied by Freud relative to the motives studied by McClelland. Freud postulated drives that were repressed due to being socially unacceptable, particularly related to sexuality and aggression. Those drives could plausibly be so threatening as to cause strong anxiety which might then plausibly lead to repression. McClelland and colleagues, by contrast, studied motives to achieve, to influence others, and to be concerned with social relationships. These motives are so comparatively tepid that the possibility that they could cause a similar form of anxiety and self-denial strains credulity. An alternate proposed mechanism by which the PSE might function was motive satisfaction through fantasy, which assumes that that writing about a motive served to satisfy it. Early work by Atkinson and McClelland argued against this mechanism by pointing out that hungry participants wrote more about food pursuit, but no more about food consumption (in fact less) than their sated counterparts (Atkinson & McClelland, 1948). Nonetheless, Atkinson later used this notion in his dynamics of action theory to explain poor correlations across stories in the TAT, assuming that motive arousal can lead to fantasy in one story, leading to motive satiation and reduction in subsequent stories. Pennebaker has offered a simpler explanation for a lack of similarity in content across ostensibly similar writing exercises in his own work. Pennebaker argues that ordinary phenomena such as the norm to avoid redundancy could explain them (Tausczik & Pennebaker, 2010).

A modern understanding of implicit motives and details of the PSE procedure lend clues to why PSE stories appear to reveal spontaneous motives. PSE images have been carefully chosen to be ambiguous, and are only presented for 8-10 seconds, meaning that the elements that will be stored in memory during the writing exercise will be colored by a person's presently activated schemata. In fact, images that are more concrete have been found to produce less meaningful PSE scores (Ramsay & Pang, 2013). Given that no experimental motive induction has taken place, these are assumed to be chronically activated schemata, reflecting frequently activated concerns. Such personal influences affecting perceptions of ambiguous stimuli have their roots in the New Look Movement of psychology (Bruner & Postman, 1948), and are most famously demonstrated in the Donald paradigm (Higgins, Rholes, & Jones, 1977). Most directly relevant to motives, though, research demonstrates that athletes that have been reminded of an athletic failure are particularly fast at recognizing words related to their sport (Moskowitz, 2002). Further, Schultheiss has described implicit motives as indicating those stimuli for which a person experiences incentive salience (Berridge, 1996; Schultheiss, 2008), suggesting that if participants are having fun with the writing exercise, their idea of fun in a spontaneous situation will guide the imagery in that activity.

Disadvantages of the PSE

Despite such advantages to these measures, the onerous nature of producing reliable motive scores from PSE stories has been used as a reason to eschew the technique in favor of less time-consuming self-report measures. For example, Blankenship (2010) notes that mastering a single coding technique requires an average of 20 hours of work for a research assistant, and Schultheiss (2013) similarly points out that once coders are trained, the coding process often takes a substantial amount of time. Even after investing such a substantial amount of time, coders sometimes produce work with low inter-coder reliability, meaning that their work is generally unpublishable. Further, coders tend to be undergraduate research assistants, who are generally only involved in a given lab for a year or two, leaving the ratio of time spent training to the time spent coding unacceptably high. These and similar problems have created a circumstance where using the PSE is simply impractical for modern psychologists. The time intensive nature of PSE coding is problematic in part because it leaves important questions about human motivation unexplored, but is further detrimental because it prevents researchers from engaging in best-practices for producing reliable findings, such as studies with large sample sizes and studies that directly replicate previous findings.

McClelland defended against such criticisms, noting that biologists often spend many hours carefully producing a single data-point for their research (Winter, 1998). This biologist model is inapplicable to modern psychology, however, due to systematic pressure to produce large volumes of studies that all achieve statistically significant results. In an attempt to make the method easier for researchers, and therefore more likely to be employed, several researchers have attempted to automate the process.

Automation of PSE Coding

Attempts at using computers for PSE automation began in the mid 1960's (Litwin, 1965; Williamson, 1965), and were refined by Smith (1968) using the General Inquirer computer system. Each of the three major motives was coded in the General Inquirer using roughly 20 dictionaries and a complex system of conditional rules that worked on a single sentence at a time. For example, in order to score a sentence as satisfying one of the ten conditions for having power motive imagery, the sentence must have a word/phrase from each of the following: authority role, subservient role, and influence role. Therefore, the following sentence from Smith (1968) "[t]he director (authority role) wants to use pressure (influence role) on the junior executive (subservient role)" would be scored as having an instance of power imagery due to the constellation of words it possesses, rather than due to the presence of any one of those words in particular. If such a sentence is identified, the computer marks it as having an instance of motive imagery. The three motive dictionaries suggested various levels of promise, but also severe limitations. Each was tested against expert coders, using both archived coded stories found in example coding materials used to train human coders (Atkinson, 1958) as well as new stories to novel PSE pictures. The old stories were arranged into four independent groupings, and the new stories were arranged into three groupings. The achievement motive coding system provided adequate reliability for both the old (100-78% agreement across all sets) and new (94-84% agreement across all sets) stories. The need for affiliation coding system provided better reliability for old (96-90%) than new (90-80%) stories. The need for power was unreliable (95-57%) for old stories, and new stories were not coded due to a lack of trained coders on the project. These findings suggest promise in complex computer coding for achievement and affiliation. Unfortunately, subsequent testing of the NAch automated coding scheme showed that it failed to replicate its earlier successes when applied to new picture sets or stories written by participants from India (D. Winter, personal communication, October 29, 2014). Thus, although this coding scheme might not be complete in its original form, its successes indicate some possibly important characteristics of highly successful attempts to replicate existent content coding procedures: analyzing stories at the level of sentences, considering combinations of word-types that are necessary to code an instance of motive imagery, and the possibility that a given dictionary will be more appropriate for stories written to some pictures than others.

All subsequent attempts to create computer-based motive coding have ignored those insights in favor of word-level analysis (Blankenship, 2010). These word-level analyses rely on the word-marker hypothesis, which is the idea that motive coding can be simplified to counts of motive relevant words and phrases proportional to overall word-count. Several areas of research have suggested that written or spoken content can indicate social motives in ways that predict actual behavior or human coding of PSE stories. Originally, research in this area showed that simple motive dictionaries could be made to produce findings that would be expected of implicit motive PSE coding, but did not compare their motive measures to implicit motives derived from the PSE.

In a first step towards demonstrating that computer coding of TAT content could produce meaningful results, Schnurr, Rosenberg, and Oxman (1992) analyzed TAT stories and free speech content using the Dartmouth Adaptations of the General Inquirer system (Oxman, Rosenberg, Schnurr, & Tucker, 1988). The version of the General Inquirer system used 55 first order dictionaries, such as self, roles, and feelings, as well as 28 second order dictionaries. For example, the word magnificent would be coded both in the first order category of "good" and the second order category of "overstate". Using these 83 categories, the authors completed a 2-step process for each of their predicted variables. For each variable, they first calculated bivariate correlation coefficients to identify categories that correlated at $p<.05$. They then computed stepwise regression to generate prediction equations. For the

TAT, the number of significant variables was considerably above the 4 (5% of 83) that would be expected by chance. When comparing the predictive validity of the two types of text samples, Schnurr and colleagues found that data obtained from analyzing TAT (relative to free speech data) content consistently predicted a greater proportion of variance for predicted variables such as gender, depression, and various subscales of the California Psychological Inventory. This research, nonetheless, showed the familiar phenomenon of low between-story reliability, suggesting a low signal-to-noise ratio in the data.

Subsequent research by Hogenraad (2003; 2005) has used the word-marker hypothesis to test McClelland's (1975) claim about the ebb and flow of power and affiliation motives in response to changing political winds. Specifically, his work tests the claim that the relative occurrence of power to affiliation motives expressed in political speech will increase at historical times when people are building towards going to war. In this research, Hogenraad (e.g. 2005) developed dictionaries for power and affiliation-related words in his new Motive Dictionary, and used them to calculate a discrepancy score. Most strikingly, he supported this hypothesis in analyses of speeches by both British Prime Minister Tony Blair and U.S. President George W. Bush in the time elapsing between September 11th, 2001, and the invasion of Iraq in late March, 2003. This hypothesis has also held when analyzing documents related to the first World War, and the Cuban Missile Crisis (Hogenraad, 2003), and histories and works of fiction reflecting periods of varying levels of conflict (Hogenraad, 2005).

The first test of the idea that the word-level analyses could act as a quick-and-dirty alternative to hand coded PSE stories was conducted by Pennebaker and King (1999) in their test of the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001) computer program. LIWC is a text analysis program that is designed to capture the stylistic elements of a text document that transcend its subject matter. Those stylistic elements, which Pennebaker likens to verbal mannerisms (Tausczik & Pennebaker, 2010), include categories that range from very broad to more specific, such as pronouns, negative emotions, causation, sexuality, and achievement. The LIWC categories were created by human raters and are not mutually exclusive, meaning that a given word may exist in multiple categories. In a first study, Pennebaker & King (1999) identified a factor structure to those categories, simplifying their data into four factors. They then used those factors in analyzing PSE stories that had also been coded for achievement, affiliation, and power motives. These analyses were exploratory, designed to learn the contours of LIWC, and the authors expressed skepticism that their analyses would relate to PSE coding. They note in particular that the conditional rules and categories involved in PSE coding are markedly different from those used by LIWC. Despite their skepticism, the researchers found two of their four factors significantly related to the implicit achievement motive, and a third was marginally related. The use of factors occludes the exact psychological meaning of the analyses, but educated guesswork suggests that people who scored high on achievement used more words that were: 6 or more letters in length (perhaps demonstrating a greater use of vocabulary), in the past tense, and social, and fewer words that were: in the first-person or that expressed positive emotions.

More recently, Schultheiss (2013) attempted in earnest to use LIWC as a replacement for human PSE coding. This replacement required a two-step process: after obtaining 8 PSE stories that had already been coded by an expert coder from each of 146 English speaking participants, Schultheiss obtained bivariate correlations between each LIWC category and the PSE the achievement, affiliation, and power motives. Then, for each implicit motive, he used all LIWC categories that had obtained a statistically significant bivariate correlation to generate regression equations by adding them as predictors in a single step. He then excluded all variables that failed to predict unique variance, and generated a final linear regression equation. The result was a set of three multiple regression equations designed to predict hand-coded implicit motives from LIWC analysis of PSE stories. Those multiple regression equations predicted achievement (R=.57), affiliation (R=.60), and power (R=.41) to degrees indicating a relation between the two measures that is meaningful, but far short of what would be considered adequate reliability between human coders. A motive manipulation study appears to support Schultheiss' regression models. In this study, participants filled out PSE stories, then underwent a motive manipulation involving watching a 30 minute film clip involving either power- or affiliation-arousing imagery (Godfather II for power or Bridges of Madison County for affiliation), then completed more PSE stories. Original PSE hand-coding demonstrated that the induction led to expected changes in the manipulated motive, and use of Schultheiss' regression equations produced an identical pattern of results. This data involved an extremely small sample, however, with approximately 10 participants per condition.

Modern Advances in Natural Language Processing

Major technical advances in the computer science of natural language processing have occurred in recent years that can be leveraged to update the fundamental idea of word-searches employed by adherents to the word-marker hypothesis. In all cases, the implementations of the word-marker hypothesis have relied on complete dictionaries that categorize raw text that has, at most, been modified by correcting participant spelling errors. This dictionary-based technique requires that the researcher compiles extensive lists of all words that could potentially be of interest in the form of computer dictionaries, and that words with multiple meanings be treated identically regardless of the sense in which they were meant. The problem of forming complete dictionaries is mainly one of tedium, requiring researchers to identify all related words of interest, as well as every possible way those words can be written, notably, including every possible conjugation. The problem of working with words with multiple meanings, also known as polysemous words, is more insidious because it means that a word that is a true marker of motive imagery when meant in one sense, can be totally devoid of motive imagery when meant in another sense. We can address both of these problems by taking advantage of research and freely available computer programs produced and used by the natural language processing community, but as yet infrequently used within the psychological community.

How to Address the Problem of Polysemy

To address the problem of polysemy, a common route is to try to transform a raw word into its precise sense meaning. Such a transformation essentially boils down to translation of spoken English into a different language, if there existed a perfectly logical language in which all words had only one meaning, and no two words were redundant.

As fanciful as the translation analogy sounds, such a perfectly logical language does in fact exist in the form of an ambitious project from Princeton University to map out all sense meanings of the English language, called WordNet (Miller, 1995). WordNet is an extensive, network-based mapping of English that has several noteworthy features. In WordNet, words are replaced by symbols that represent clusters of synonyms. Those symbols are called synonym sets, or synsets[2]. Two words are said to belong to the same synset if it is true that they can generally replace one another in a typical sentence without changing the meaning of the sentence. For example, the words achievement and accomplishment can generally be used interchangeably and so are both contained within the same synset. Further, each synset contains only a single sense in which a word is meant, so a given word can be contained in a variety of synsets. The word *fly*, for example, can refer to the verb that describes what birds do, the noun that describes an insect,

---

[2] synsets take the form: word#part-of-speech#number, with part-of-speech including v=verb, n=noun, a=adjective, and r=adverb, and number indicating the sense in which it is meant, with lower numbers indicating more common senses. So win#v#1 indicates that the word win is a verb, and is meant in its most common sense.

or the noun that describes the opening of a garment. As such, the word fly is contained in multiple synsets.

To automatically disambiguate words into their most likely synsets, a number of steps can be taken as outlined by Kolhatkar (2009) and others (Joshi, Dad, Bagul, & Haribhakta, 2014; Manning & Schütze, 1999). A first step involves the recognition that colloquial uses of language often use phrases to suggest meanings that are not present in any of the component words. If we read about an *up and coming* musician, racing horses that are *neck and neck*, or an employee who is getting ready to change careers and *move up*, we may find achievement imagery in phrases where none exists in any individual word. Thus, for an automated coding algorithm to have a chance at identifying all instances of motive imagery, it is necessary to analyze text at the level of phrases when appropriate, rather than exclusively at the level of words.

Within WordNet 3.0, these phrases are called compound words, and also include proper names such as George_Washington. Compounds make up over 40% (64,331) of all words on WordNet, and generally only have a single sense meaning. Thus, once compounds are identified, they can be disambiguated extremely easily into their only valid sense. By contrast, failing to identify compounds assures that all words that should be in compounds will receive the wrong sense meaning. Fortunately, the freely available Perl module, WordNet::Tools includes the *Compoundify* function, which converts phrases that it recognizes as WordNet compounds into the compound form, such as neck_and_neck.

A second step towards disambiguating raw text into synsets involves the recognition that different word senses often cross parts of speech, meaning that if a word's part of speech could be identified, the number of senses to which it could refer will often be reduced. Part of speech tagging (POS-tagging) is a sophisticated and fairly accurate area of natural language processing that uses predictive models to determine the part of speech of each word in a text using sentence structure to make educated guesses. The current state of the art baseline accuracy is 97.24% (Toutanova, Klein, Manning & Singer, 2003), though that includes some classes of words such as conjunctions that are not of interest for our present purposes, with a more conservative estimate being around 92% accuracy for nouns, verbs, adjectives, and adverbs (Kolhatkar, 2009). Using POS-tagging, ambiguous words become less so, or may be entirely disambiguated. In the case of the word *fly* discussed previously, if it were found to be a verb, the word could be disambiguated into the synset that refers to what birds and airplanes do. If it were found to be a noun, the number of possible synsets to which it could refer would at least be narrowed.

Once a set of words have been grouped as compounds, and tagged for part of speech, context-based word sense-disambiguation programs can perform the final step, transforming those words into their most likely synsets. Context-based word sense disambiguation is a method by which machine translation and information retrieval programs function. The only form of such word sense-disambiguation that does not require extremely resource-intensive model building relies on the logic that related words tend to co-occur in the same sentence. They therefore use the relatedness measures that will be described later to compare each potential meaning of a word against the possible meanings of the other words in a sentence. For example, the word bank can refer to a river bank, a financial institution, or a way to angle the wings of an airplane. After POS-tagging identifies bank as a noun, thus eliminating the airplane possibility, disambiguation software then computes relatedness scores between each remaining synset that the word could mean and the possible synsets that the surrounding words could mean. Then the presence of words such as cashier, money, or businessman would tilt the disambiguation software towards identifying the word as a financial institution, and words like water, river, and boat would tilt it towards identifying the word as a river's edge.

How to Address the Problem of Dictionary Generation

As mentioned previously, the problem of dictionary generation is not intractable, but it is labor intensive and tedious. Much of the appeal to the LIWC software for psychologists resides in the extensive dictionaries produced by Pennebaker and colleagues, so much so that those dictionaries are copyrighted and kept unpublished. By converting raw text into synsets, that problem is partly solved because various synonymous words will be converted into the same synset. Most words that are closely related to each other, however, are not precise synonyms, so even after conversion to synsets, dictionary generation would be extremely time-consuming. Fortunately, researchers have produced considerable work outlining ways to quantify the extent to which synsets that are not identical are related to one another. By identifying closely related synsets, researchers can ask questions such as "is the synset 'win#v#1' in this text, and if not, is there another synset that is closely related to it?" By asking such a question, it becomes possible to side-step the process of generating exhaustive computer dictionaries, instead searching for a limited number of synsets and those synsets that are closely related to them.

Relatedness is calculated using WordNet's definitions for each synset, called glosses. The prototypical form of relatedness is the Lesk relatedness score, which compares the words in one gloss definition with the words in another gloss definition, and counts the number of overlapping words. To compare the WordNet definition of hammer ("a hand tool with a heavy rigid head and a handle; used to deliver an impulsive force by striking") and nail ("a thin pointed piece of metal that is hammered into materials as a fastener"), we can see that even very related words can happen to have a low relatedness score according to the Lesk relatedness score. To extend the gloss definitions, Pedersen (Banerjee & Pedersen, 2003) proposed an Extended Lesk relatedness score, which used the gloss definitions of each synset, combined with the definitions of the synsets to which they are directly connected in the WordNet network, thus producing "superglosses". A supergloss for the word "hammer", for example, might include the gloss definitions of hammer, the more general category of hand-tools, and the more specific categories of mallet, ball-pein hammer, and so on. As a final extension of relatedness scores, Pedersen (2012) also produced a relatedness measure which he calls Vector relatedness. Vector relatedness takes the words in the supergloss definition of each synset, and compares their meaning rather than their exact words, using a complicated and powerful technique called latent semantic analysis[3], which allows it to identify gloss

[3] The reader may wonder why, if such a powerful technique for identifying similarity of meaning exists, it is not elaborated upon in this paper or used in the present research. Latent semantic analysis, although powerful in many contexts, is

definitions that overlap in meaning, even if they do not share many of the same words. Although the Vector measure has yet to produce the body of findings amassed by Lesk and Extended Lesk, it appears to mostly function as well or better than other relatedness scores (Kolhatkar, 2009).

Use of Natural Language Processing for Motive Imagery Coding

In light of the preceding discussion, I propose the following formula for developing an automated motive coding system.

Disambiguate Text Into Synsets. The first step of automated motive coding is to determine the sense meaning of each word. As such, PSE stories are first processed using the *Compoundify* function of the WordNet::Tools Perl module, turning common phrases and compound words into single strings. Next, words are tagged using a state of the art POS-tagger to differentiate verbs, nouns, adjectives, and adverbs. Finally, words are disambiguated using a context-based disambiguation program that uses Vector relatedness to determine word sense.

Calculate Maximum Synset-to-Sentence Relatedness using Synsets that Likely Indicate Motive Imagery. The next step involves breaking each PSE story into its component sentences, and then comparing each synset in a given sentence against a synset of interest, say “flair#n#1”, and then identifying the greatest relatedness score across all synsets in the sentence. For example, the gloss definition of flair is “a natural talent “, so when computing relatedness scores of that synset with each synset in the following sentence “his#CL father#n#3 know#v#11

---

inappropriate for the present research for several reasons including: 1. it does not differentiate between polysemous words; 2. it requires a much larger training set than are available for the PSE; 3. it compares meanings of entire blocks of text, rather than individual words an phrases within sentences.

that#IT Joe#n#ND could#IT use#v#01 and#IT exercise#v#1 his#IT talent#n#1 in#IT more#IT productive#a#1 way#n#2 .IT", it returns the relatedness score between flair#n#1 and talent#n#1, which has a gloss definition of "natural abilities or qualities" leading to a relatedness score of 0.52[4].

This process, which I will call the Maximum Synset-to-Sentence Relatedness (MSSR), creates a series of variables with scores that range from 0 to 1, with each variable indicating the extent to which any word in a disambiguated sentence is closely related to a synset of interest. By combining variables, it becomes possible to look for phrases, such as by simultaneously considering whether any word in a given sentence is highly related to "want#v#1" and whether any word in that same sentence is highly related to "win#v#1".

I propose that the MSSR, which has thus far only been used to calculate relationships among full sentences (Mihalcea, Corley & Strapparava, 2006; see also Arnulf, Larsen, Martinsen, & Bong, 2014), has the potential to be a more targeted alternative to LIWC. Because it relies on extensive text disambiguation, it can distinguish different meanings of the same word, which LIWC cannot, and which will only improve as the associated computer programs are refined. Because it allows the researcher to determine the synsets of interest, it can focus in a more nuanced way on certain classes of words, such as various classes of achievement words.

Detect and Model Non-Linear Relationships Among Words. Nearly all models of motive imagery have relied on assumptions of linear relationships

[4] This gloss example is a simplification. The Vector relatedness score uses superglosses, rather than simple glosses.

between the quantity of certain classes of words and an individual's motive scores. Pennebaker and King (1999) used multiple regression to track main effects of LIWC scores, after those scores had been reduced to composite variables by using factor analysis. Schultheiss (2013) similarly tracked main effects of LIWC variables on motive scores, using multiple regression. Interestingly, the earlier models generated by researchers such as Litwin (1965) and Smith (1968) used the more complex assumption that constellations of words could indicate motive imagery in a non-linear fashion. An assumption of non-linear relationships allows for the possibility of accurately coding sentences such as "they would be rewarded for all their hard work after all." Such sentences require the presence of two words, in this case "hard" and "work", neither of which imply achievement imagery by themselves, but which tend to indicate achievement imagery when together.

A likely reason that the assumption of non-linear relationships has been abandoned is that they generally require the specifications of all such relationships beforehand. In the cases of both Pennebaker and King (1999) and Schultheiss (2013), regression prediction models were used, meaning that all non-linear relationships would need to be pre-specified in the form of additional interaction variables. Although such statistics would theoretically be capable of detecting complex interactions if all possible interactions were modeled and exploratory methods, such as step-wise regression, such an ambitious project would likely be rife with findings that track random fluctuations in their data (Simmons, Nelson, & Simonsohn, 2011). By tracking random fluctuations, a model developed using such exploratory techniques would be unlikely to generalize to a different data set.

Alternate methods of developing prediction equations have been developed that are capable of detecting interactions in an exploratory fashion, while simultaneously minimizing the problem of over-fitting to idiosyncrasies in the data. One such method is called a multilayer perceptron neural network (which will subsequently be shortened to "neural network"; Richard & Lippmann, 1991). Neural networks function in a way meant to be analogous to the functioning of groups of neurons involved in learning. Two major components of neural networks are the input layer and the output layer, each of which contain one or more nodes. For simplicity sake, consider a person learning to skip a rock on a lake. The nodes in the input layer each correspond to a variable in the model, such as proprioceptive information about the current position of one's hand, the weight of the rock, shape of the rock, and distance to the water. The output layer generates decisions, such as force and angle of the throw (each represented by a different node in the output layer). If only linear relationships existed, the input layer could send each piece of information directly to the output layer along connections that adjust the influence of each piece of information by the weights assigned to each connection in a manner very similar to linear regression. In such a simplified model, the system would learn as humans do, making an attempt to throw a rock, observing success or failure, making incremental changes to connection weights in response to failures, and making no changes when encountering successes. Such learning in response to success and failure is called back-propagation. In a more complex model, recognizing that input variables often interact, it is necessary to include one or more hidden layers. Hidden layers serve as intermediate steps in data processing, accepting information from the weighted connections of the input layer, and passing it along to the output layer using an activation function and further weighted connections. Based on the pattern of connections to and from the hidden layer, as well as the number of nodes in it, neural networks that contain a hidden layer are capable of modeling complex interactions in data. The functioning of these models are so complex that they are generally recommended for situations when quality of prediction is more important than interpretability of the model.

Unlike regression, neural networks have been designed to simultaneously model a data set's complexities without becoming overly susceptible to modeling its idiosyncrasies. Neural networks accomplish this task with data partitioning. Data partitioning is the act of randomly carving a set of data into subsamples, namely a training sample and a testing sample. Neural networks converge towards optimal connection weights and numbers of nodes in the hidden layer by making small changes to them in directions that it calculates should reduce it's R^2 error in reference to the training sample. Left unchecked, this method would ensure a model that is over-fit to the training data. However, after each modification to the model, the neural network also notes whether it has reduced R^2 error in reference to the testing data, which was not used to make the modifications. If the model fit has also improved in reference to the testing sample, another iteration of learning from the training sample and making small modifications is performed. If model fit is not improved, model construction is either immediately terminated, or allowed to continue for a limited number of additional iterations to try to improve model fit relative to the testing sample before it is terminated.

An Updated Word-Marker Hypothesis

The above discussion suggests that the word-marker hypothesis is a valuable idea but could be updated by techniques developed by natural language processing and machine learning researchers. First, it could be updated in part by recognizing that raw words can be processed into disambiguated forms using a multi-step process. Second, it could be updated by replacing computer dictionaries with relatedness scores produced by the MSSR procedure. Finally, it could be updated by searching for constellations of words that exist within individual sentences. One means of finding and modeling such constellations of words involves splitting sense-disambiguated PSE stories into their component sentences, and then searching for non-linear relationships among variables from the MSSR process using machine-learning-based prediction algorithms, namely multilayer perceptron neural networks (e.g. Russo, Vempala, & Sondstrom, 2013).

The Present Research

The present research was conducted in three parts. Although it is more conventional to refer to these parts as studies, the reliance of parts 1 and 3 on archival data means that the use of the word "study" is not entirely accurate. In the first part, I collected PSE stories that had been coded for NAch imagery at the sentence level using the Winter coding system for NAch, converted the stories into their component sentences, producing a data set containing 2,371 sentences, 855 of which contained achievement imagery. I then manually searched for words and phrases that tended to indicate achievement imagery (in the data set, in Litwin's dictionaries, and as made sense according to the Winter scoring system for NAch), converted those into a much larger list of synsets, and used those synsets to produce an initial set of nearly 600 potential MSSR variables. I then analyzed the sentences using both LIWC[5] and MSSR, producing numeric variables associated with each analysis type.

Next, in order to reduce the number of MSSR variables to a more manageable number, I ran a series of exploratory neural network models predicting NAch imagery using the MSSR variables, recording the model performance, and the influence of each variable. The model was trained on 70% of the data, and the remaining 30% of the data was reserved as a testing sample. I then re-ran the model excluding the 10 least influential variables, and noting whether model performance decreased. I repeated the process until I had reduced a large number of extraneous variables from the model without appreciably reducing the resultant model's performance.

The remaining MSSR variables and all of the LIWC variables were then entered into 2 neural network prediction algorithms (one for each text analysis type) to train a model to predict human coded NAch imagery using the same settings as the exploratory neural networks. I recorded the performance of those models in terms of the training and testing samples respectively, and saved them for later use.

In the second part, I designed and implemented a motive induction study for the achievement motive, modeled heavily on the procedures described by

[5] LIWC was applied to the raw, not disambiguated text.

McClelland and colleagues (1949; McClelland & Atkinson, 1955). Consistent with all other motive induction studies, following motive arousal, participants completed the PSE. The resultant PSE stories were broken down by sentence, and converted to numeric variables by LIWC and MSSR using the reduced set of MSSR variables identified in part 1. Similar to part 1, I then produced 2 neural-network prediction models, this time predicting condition from story content. Additionally, I had trained coders code all stories for NAch. The human coding of NAch served as a manipulation check, and as a new data set to use to test the models developed in part 1.

In part 3, I tested the models generated in parts 1 and 2 on an additional data set collected and coded for NAch (using the system developed by McClelland and colleagues) by Ratliff (1979) from the Henry A. Murray Research Archive. That data included sets of PSE stories from 114 research participants along with the overall NAch score of each participant. After breaking the stories into sentences, I used both models to code the stories, which I then aggregated into an overall person-level score. I tested the scores produced by the two models from part 1, and the two models from part 2 against those produced by human raters. Agreement between the model predictions and the McClelland coding scheme would be taken as evidence of convergent validity. I further used all 5 achievement scores to predict additional variables measured by Ratliff, including other implicit motives and academic self esteem. Correlations between model predictions and these additional variables would be taken as evidence of predictive validity.

Part 1

Overview

For Part 1, I contacted various prominent researchers associated with the MDT literature, asking for PSE data that had been coded for NAch, with sentences containing motive imagery indicated. I received data meeting those criteria from David G. Winter, Oliver C. Schultheiss, and Steven J. Stanton. In addition, I incorporated sentences from a data set collected by Kennon M. Sheldon[6]. The resultant data set contained 2,371 sentences, 858 of which contained achievement imagery.

Next, I compiled a list of words and phrases that appeared to be associated with NAch (See appendix I) based on my knowledge of the coding criteria, an informal review of the data, and the list compiled by Litwin (1965). I then converted those words and phrases to all possible synsets to which they could refer, and retained those that were relevant for achievement imagery. These synsets formed my initial variables for the MSSR procedure.

Next, I processed the raw text of the PSEs into synsets by first running compoundify in the WordNet::Tools Perl module, then running POS-tagging, then running a context-based word-sense disambiguation program using Vector relatedness. I then used original computer programs programmed in Python 3, in tandem with a freely available computer program called WordNet::Similarity, to calculate MSSR. Next, I produced a series of neural network prediction models,

[6] All of these data sets were coded using the Winter (1991, 1994) coding system, which influenced my decision to use that system in the present research. However, I would like to recognize Andrew J. Elliot's generous contribution, consisting of a set of PSEs coded for NAch using the McClelland coding system.

predicting NAch from MSSR variables, to reduce the number of input variables that were least influential (see Appendix IV for the final list of MSSR variables). Finally, I produced 2 final neural-network based prediction algorithms for NAch, using LIWC and the remaining MSSR variables respectively.

Method

Data Sets. Data set #1 was a training set used to train coders in the Winter Running-Text coding system (Winter, 1991; 1994). It contains 180 PSE stories with 901 sentences and 89 instances of motive imagery. Coded text that were not from PSE stories, such as excerpts from political speeches, were not included because motive imagery in such formats tend to be expressed qualitatively differently, involving more nuance of language and less direct statements (Winter, 1994). Data set #2 was a calibration set used to re-train coders that had previously mastered the Winters coding system. It contains 30 stories with 218 sentences and 26 instances of motive imagery. Data set #3 was obtained from Oliver Schultheiss. It contains 48 stories with 320 sentences and 47 instances of achievement imagery. Data set #4 was obtained from Steven Stanton and are from unpublished research. The stories were recorded in scanned, hand-written PDFs with instances of motive imagery underlined in pen. Due to time limitations, it was not feasible to transcribe these stories, and so only sentences containing achievement imagery were transcribed and included for analysis. There were 331 such sentences containing achievement imagery included for analysis. Data set #5 was from 2 data sets collected by Sheldon and colleagues, one of which was used in published research (Sheldon, Prentice, Halusic, & Schüler, 2015) and one of which is unpublished.

Because the intercoder reliability was not excellent in the published sample, (ICC = 0.68), and the data from the unpublished sample was not structured in a way that readily allowed for the computation of intercoder reliability, only sentences containing achievement imagery were included, and those were only included after I confirmed that they contained unambiguous instances of NAch imagery. Three hundred sixty two such sentences were included.

Computer Programs.

Linguistic Inquiry and Word Count 2001. The Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis & Booth, 2001), pronounced "Luke", is a widely used form of data analysis within the psychological community, available for purchase at *www.liwc.net*. LIWC functions by categorizing the words in a text into potentially psychologically meaningful categories, such as achievement, anger, and prepositions, producing variables for each text that—except for words per sentence and word count—indicate the proportion of words in the text that fall into each category. These categories themselves fall into different groupings, such as Standard Linguistic Dimensions (e.g. pronouns, negations, prepositions), Affective or Emotional Processes (e.g. positive emotions, negative emotions) and Personal Concerns (e.g. achievement, school, sports).

Pennebaker and colleagues (Pennebaker, 2011; Tausczik & Pennebaker, 2010), have argued that standard linguistic dimensions make up much of the stylistic portion of speech, rather than other categories that are more tied to the content of the information being conveyed. As such, they argue that those categories are particularly revealing of underlying psychological and social dimensions of the speaker. For example, people who experience more physical or emotional pain are more self-focused, and so use more first-person singular pronouns (Rude, Gortner, & Pennebaker, 2004). By contrast, men who report lower relationship satisfaction use marginally more second-person pronouns when speaking to their significant others (Slatcher, Vazire, & Pennebaker, 2008). LIWC has been used by Schultheiss (2013) to create a regression-based model of implicit motives, using a composite of all stories written by an individual as the level of analysis (rather than each sentence, as is the focus here).

Stanford Part-of-Speech Tagger. The Stanford Part-of-Speech Tagger (Toutanova & Manning, 2000; Toutanova, Klein, Manning & Singer, 2003) is a state-of-the-art maximum entropy POS-tagger produced by the Stanford Natural Language Processing group. The term maximum entropy refers to a class of machine learning prediction algorithms that generate the model that simultaneously does the best job of describing the training data while assuming the smallest defensible number of interdependence among predictors. It is freely available at *nlp.stanford.edu/software/tagger.shtml* and has been programed to work using Java, meaning that it is platform independent. It has been tested to produce 96.86% accuracy when tested on known words, and 86.91% accuracy when tested on new words (Toutanova & Manning, 2000). POS-Taggers often function by assigning the probability that a given word belongs to each possible parts of speech, and then assigning a POS-tag that is consistent with the highest probability. Probabilities rely on such metrics as the general likelihood that a given word is a particular part of speech (frequency information), the specific likelihood that a word of a given part of speech follows a word with the part of speech of the preceding word, and the likelihood that a word of a given part of speech follows the sets of parts of speech of the preceding two words. Because the Stanford Part-of-Speech Tagger includes a bidirectional model, it is able to use contexts in both directions of the word, rather than just that preceding the words (Toutanova, Klein, Manning & Singer, 2003). Because it is a maximum entropy tagger, it is capable of performing much more sophisticated calculations with the contextual information than those just described.

WordNet::Tools Compoundify Function. The compoundify function of WordNet::Tools replaces groupings of words with their compound equivalent in WordNet. It employs no artificial intelligence nor complex decision rules in determining whether to convert words into compounds, rather, it merely recognizes clusters of words that could be compounds. It uses a "greedy" search method, meaning that when 2 overlapping compounds are possible, such "the United States of America" containing both United_States and United_States_of_America, it chooses the larger of the two compounds. The advantage of using compoundify is that it makes it possible to correctly identify concepts that are expressed as compounds, such as "best_friend", but the disadvantage of using it is that it has no mechanism of preventing false-positives, which will necessarily occur when words appear together but do not imply a unitary concept.

WordNet::Similarity. WordNet::Similarity is a program designed to calculate all major similarity and relatedness scores that use WordNet, and is freely available from *wn-similarity.sourceforge.net*. It is the only freely available computer program capable of computing Extended Lesk, and Vector relatedness. A major limitation of Lesk, and Extended Lesk relatedness measures is that they require exact matches across glosses, meaning synonymous words across glosses ("beat" vs. "defeat") and even different conjugations of the same word ("win" vs. "won") are treated as entirely unrelated. Vector relatedness deals with this problem by using latent semantic analysis to replace each word in the supergloss with a vector that summarizes all the words to which that word is related, based on co-occurrence data identified by analyzing all gloss definitions and synset examples in WordNet. The vectors of each word in the supergloss are then combined into a single vector, and compared against the vector that summarizes the supergloss of a second synset. The cosine of the angle between the two vectors is computed as the relatedness between the two synsets (Pedersen, 2012).

Other free programs that can calculate WordNet-based relationships among synsets are also available, with Python's Natural Language Tool Kit (NLTK; Loper & Bird, 2002) being the most prominent. NLTK, however, can only calculate structure-based similarity, rather than gloss-based relatedness, which have a variety of important limitations that make them inappropriate for the current purposes.

WordNet::SenseRelate::AllWords. WordNet::SenseRelate::AllWords (SR-AW; Pedersen & Kolhatkar, 2009) is a word sense disambiguation software program that is freely available from *senserelate.sourceforge.net.* It leverages the similarity and relatedness metrics available from WordNet::Similarity to disambiguate text without requiring researchers to manually disambiguate a large corpus for training a machine learning program. It works by moving word by word in a sentence from left to right, comparing each word to all words within a user-defined window. Each word is expanded into all possible synsets to which it could possibly refer, as are all of the context words within the window. For example, if a window were set to 8, each word of interest would be compared to 4 words to its right, and 4 words to its left. Then, using any of the 10 possible similarity or relatedness measures available in WordNet::Similarity, it calculates overall similarity/relatedness of each possible synset of the word to the words in its context, and returns the synset that is most similar or related to its neighbors within the context window.

Only one systematic review has assessed the accuracy of SR-AW, leading to several conclusion, though subsequent changes to the program mean that not all of its functions have been extensively studied. Although SR-AW can function using raw text, its accuracy is improved by using text that has already been processed using compoundify and POS-tagging. Comparisons of similarity versus relatedness measures show that Adapted Lesk and Vector relatedness outperform all similarity measures, and that those two relatedness measures generally perform similarly to one another. Overall accuracy increases somewhat as window size increases from 4 to 15. A new addition that has not been thoroughly tested is the "backoff" option, which assigns the most common word sense if no best sense can be determined. In light of the finding that simply assigning the most common word sense outperforms all disambiguation algorithms (Kolhatkar, 2009), the use of the "backoff" option may be a step towards exceeding that benchmark.

Original Programs. I commissioned two original programs to manipulate data in spreadsheets to augment the function of WordNet::Similarity. The programs were commissioned using Odesk.com (Groysberg, Thomas, & Tydlaska, 2011), which is similar to Amazon's Mechanical Turk system (Buhrmester, Kwang, & Gosling, 2011), with the exception that the workers are skilled professionals, including computer programmers. WordNet::Similarity is designed to return the similarity or relatedness scores of any pair of synsets, provided the pair is by itself on a line of a text file. Therefore, to compute relatedness scores of all words in a series of sentences against all comparison words of interest, it is necessary to use the first program to reformat a spreadsheet that looks something like Figure 1, into a space-delimited text file, excluding all words that are not tagged as nouns, verbs, adverbs, or adjectives, that looks something like Figure 2, as well as a similar spreadsheet that also keeps track of which words and sentences are being compared that looks like Figure 3. The space-delimited file can then used as an input file for WordNet::Similarity to produce a list of relatedness scores that, with minor editing, can be added to the more complete spreadsheet to create a column of relatedness scores. The resultant file should then look like Figure 4, which can then be loaded into the second original Python 3 program to find the largest similarity score between each target word and all words in a sentence, which looks like Figure 5.

Even though this analysis occurs at the word-level, combinations of words may be found in a neural network using multiple nodes in the hidden layer, leading to something more like multi-word phrases.

One point that may be apparent to the reader is that some words appear particularly related to the synset be#v#1. That synset is the disambiguated form of "be", "is", "was", "been", and " 's". Because that relationship is never predicted to be of interest, I removed all forms of "be#v#1" at the disambiguation phase, providing the program with a list of stop-words. Stop-word lists are often used in natural language processing, and are lists of words that are very common and hold little value for the researcher. By including a stop-word list, a researcher may tell a program to ignore certain words. By providing a stop-word list to WordNet::SenseRelate::AllWords, the words on that list were not disambiguated, and instead were tagged as being ignorable for all subsequent steps in the MSSR process.

Procedure

The first step of processing the raw text was to spell-check all stories because subsequent stages of text processing work under the assumption that words are spelled correctly. PSE stories were then converted into sentences and restructured into a spreadsheet such that each row contained only one sentence. These sentences were then processed using LIWC. Variables produced by LIWC, other than word count, were converted to binary variables, with all numbers greater than 0 being converted to 1. Words per sentence were not included in any models because it was completely redundant with word-count. The decision to convert these variables from continuous to binary was due to the observation that with such small text samples (a single sentence) the meaning of the variables as proportions of text that fit into a given category would not be meaningful, with fluctuations above 0 mostly indicating the total number of words in the sentence.

Creating a List of Synsets for MSSR

To create a list of synsets to use as variables in the MSSR procedure, I first looked through the data for words and phrases associated with achievement to begin compiling a list of potential achievement-related words and phrases (see Appendix I). Rather than simply data mining, I confined my search to words and phrases that made sense within the context of the Winter scoring system. Some of the words identified could be expected to lead to achievement imagery on their own (e.g. the word "victorious") and others were expected to be useful in combination with other words (e.g. the word "scientist" in phrases like "she is the world's leading scientist"). I then augmented that list with words and phrases from Litwin's automatic coding system that were absent from the list so far (e.g. "Gutenberg").

Next, I used the "synsets" command in the WordNet module of NLTK (Loper & Bird, 2002) for Python. The "synsets" command produces all synsets to which a particular word might refer. This produced a much larger list of synsets than the original list of words[7]. I then used the "definition" and "examples" commands in the WordNet module of NLTK to determine which of the synsets produced in the previous step were potentially relevant for NAch by viewing their gloss definitions and examples of their usage in sentences. This process culled the list of synsets to 589 synsets that might be relevant to Nach, either alone, or in combination (see Appendix II).

[7] I broke phrases into component words at this point out of computational necessity. Those words can be recombined after individual relatedness scores are produced, in the form hidden layers in the neural network

To generate the MSSR variables from the list of 589 synsets, it was next necessary to convert the PSE sentences into synsets. Using the compoundify function of WordNet::Tools, I created compound words that could be recognized by WordNet. After doing so, I noted that compoundify often inappropriately created the phrase “good_and” whenever those words appeared together and in that order. Such a pairing is potentially problematic because the word “good” is almost always a marker of achievement imagery. By contrast “good_and”, as in “good and tired”, is defined as “completely or thoroughly” and so would not be a marker of achievement imagery. Based on the assumption that a large proportion of PSE stories that contain the phrase “good and” do not mean it in the compound sense, I replaced instances of “good_and” with “good and” using a simple search and replace.

Next, I used the Stanford Part-of-Speech Tagger to assign POS tags to each word. I used the bidirectional tagger, which is more accurate than the alternative model, which only uses words to the left of the to-be-tagged word for context.

Next, I used WordNet::SenseRelate::AllWords to disambiguate the PSE stories into synsets using Vector relatedness as the metric to determine the extent of relatedness between each possible synset and its context. In recognition of the fact that the “pick the first sense” heuristic outperforms all other disambiguation procedures, I used the “backoff” setting, which assigns the most common sense of a word if no best sense can otherwise be determined. The stopword list (see Appendix III) was composed of 242 common stopwords, with special care taken to ensure that all forms of be#v#1 were included. By including these words in the stopword list, those words were not disambiguated by SR-AW, instead listed in the form “word#IT”, which prevented them from influencing disambiguation or being involved in the MSSR variables at later stages of text processing.

Next, I used the original Python 3 programs in conjunction with WordNet::Similarity to produce the maximum relatedness score of each target synset compared to all synsets in a given disambiguated sentence; the maximum synset-to-sentence relatedness.

Reducing the Number of Synsets of Interest

The next step was to cull the number of variables involved in MSSR by reducing the number of synsets of interest. This was accomplished by generating a number of neural network models for exploratory purposes. These models, although only used to determine the importance of each variable, were carried out in a manner identical to that used to create the final prediction algorithms (settings therefore will be discussed in the subsequent section on model generation). An initial model was generated using all MSSR variables in the input layer of the neural network. The resultant model provided a list of relative variable importance ranging from 100% (most important) to nearly 0% (most unimportant). The 10 least important variables were removed, and the model was re-run using the same settings. The process was repeated multiple times, but terminated before model accuracy appreciably decreased.

Model Generation

In the cases of both the LIWC and MSSR, models were created using multilayer perceptron neural networks using the respective data sources as input variables, and achievement scores as the dependent variable.

Data were analyzed using the Neural Networks module of SPSS 21. The form of network chosen was a multilayer perceptron model with back-propagation, allowing the model to learn experientially (IBM, 2012) and iteratively in a manner inspired by the functioning of networks of neurons (Maroco & Bártolo-Ribeiro, 2013). Model creation in this context involves the determination of the proper network architecture in terms of both the number of nodes in the hidden layer(s), and the weights of the connections from the nodes in the input layer (the predictor variables) to the nodes in the hidden layer, and from the nodes in the hidden layer to the nodes in the output layer.

The model was designed to produce 2 independent pseudo-probabilities, the probability that NAch is absent in a given sentence, and the probability that it is present. Although these probabilities would be expected to be the inverse of one another, they are pseudo-probabilities because they are independently calculated, meaning that they do not necessarily sum to 1. These two probabilities are then compared to each other, and the larger probability "wins" producing an output variable that takes the form of either 0 (no imagery present) or 1 (imagery present). The training method chosen was "batch" training, meaning that each time the synaptic weights are updated, the model first passes through all cases in the training partition of the data set. Batch training is the recommended method of training a multilayer perceptron model when the number of cases to be analyzed is not extremely large.

The network structure was pre-specified to have a single hidden layer containing between 1 and 50 nodes (plus a bias node). A model containing a single node in the hidden layer would suggest that only main effects are being modeled. Greater numbers of nodes in the hidden layer allow for increasingly complex non-linear relationships between the predictor variables and the dependent variable. Greater numbers of nodes concurrently carry the risk of leading to a model which is over-fit to its training sample. Data partitioning attempts to mitigate that risk. If extremely complex non-linear relationships between the predictor variables and the dependent variable were hypothesized, a structure with two hidden layers might be appropriate. However, such a model would require a particularly large training sample to identify such complex interactions, and as such, a single hidden layer was deemed more appropriate.

When building a neural network, data are partitioned into training and testing samples. In the present study, 70% of the data were randomly partitioned into a training sample, and the remaining 30% were partitioned into a testing sample. The training sample is used to generate a model by incrementally adjusting connection weights and numbers of nodes in the hidden layer towards the goal of minimizing the R^2 prediction error of the training sample. The testing sample is used after each iteration to determine whether the model has actually improved relative to data not used to determine the changes in that process. When 5 iterations have passed that have all failed to lead to improvements in the model as tested against the testing sample, modifications to the model are terminated, and model building is complete.

All input variables were standardized, set to have a mean of 0 and a standard deviation of 1. The SPSS random number generator was used to select cases in each partition, to select initial synaptic weights, and to determine initial model structure. As a result, identical instances of model building can lead to slightly different models, as initial values can influence where the model eventually converges.

Results

Reducing the List of MSSR Variables. The first step towards MSSR model generation was to determine which of the 589 synsets to include as variables. Determining the synsets to include was performed by first generating a neural-network prediction model, including all MSSR variables in the input-layer. In general, including too many variables in a neural-network model is not problematic, because the use of a testing sample makes over-fitting to the data unlikely, and variables that are not useful can simply have their synaptic weights reduced to near zero. For elegance and computational ease, however, it is desirable to calculate as few MSSR variables as possible.

After generating the first model, I recorded the overall accuracy of predicting variables in the testing sample (73.72%), the accuracy of correctly identifying instances of NAch imagery, also known as the model's sensitivity (72.24%) and 10 least influential variables as measured by the degree to which a change in a given variable leads to a change in the output of the model. I then generated a second model containing all but those 10 least influential variables to estimate the degree to which predictive accuracy decreased in response to the omission of those variables (Maroco & Bártolo-Ribeiro, 2013). The overall accuracy of that model was 70.66%, and the sensitivity was 44.61% which is an expected degree of variability given that neural-network models randomly select cases to include in each partition, initial values of synaptic weights, and initial network architecture. I again removed the 10 least influential variables and generated a new model, this time with overall accuracy of 76.1% and a sensitivity of 64.6%. At this point, I began removing the 30 least influential variables at a time, leading to a number of models with overall accuracies that ranged from 79.53% to 74.15%, and sensitivities ranging from 72.51% to 51.99%. This process was repeated until the variables had been reduced from 589 to 189, with an overall accuracy of 80.66% and a sensitivity of 65.94%. Because of the degree of random fluctuation present between models, it was difficult to create a stopping rule to this procedure, but 189 seemed to be a manageable number of variables, and there were no longer any variables with importance values at or below 0.001 (with the range now being between .0013 and .012, M=.005).

Maximum Synset-to-Sentence Relatedness Neural Network. I then produced one final MSSR neural network, using the reduced set of inputs in the previous step. The neural network based on inputs from the MSSR variables had 70% of cases partitioned into the training sample and 30% of cases partitioned into the testing sample. It had 5 nodes in the hidden layer plus the mandatory bias node. The training sample had a sensitivity of 73.50% and a specificity (the percent of accurately identifying instances when NAch was absent) of 89.6%, for an overall accuracy of 83.80%. In the more stringent testing sample, the model had a sensitivity of 67.41% and a specificity of 86.81%, for an overall accuracy of 79.64%. Table 1 shows a list of the 10 variables most strongly influenced the model's output.

The reliability between the predicted NAch scores and actual NAch scores is *ICC*=.62 ($p<.001$).

LIWC Neural Network. The LIWC-based model was trained using all variables in the input layer, with settings identical to those used to train the MSSR model. The architecture of the neural-network had 8 nodes in the hidden layer plus the bias node. In the training sample, the model had a sensitivity of 56.05%, and a specificity of 89.35%, with an overall accuracy of 77.49%. In the testing sample, the model had a sensitivity of 54.48%, a specificity of 86.72%, and an overall accuracy of 74.72%. The 10 most influential categories were occupation (e.g. work, class, boss), positive emotion (e.g. happy, pretty, good), word count, job (e.g. employ, boss, career), metaphysical concerns (e.g. God, heaven, coffin), hearing (e.g. heard, listen, sound), achievement (e.g. try, goal, win), body (e.g. ache, heart, cough), friends (e.g. pal, buddy, coworker), and optimism (e.g. certainty, pride, win). The reliability between the predicted NAch scores and actual NAch scores is *ICC*=.47 ($p<.001$).

Discussion

In part 1, I generated a list of synsets to use with the MSSR measure, and created 2 neural network models of NAch. The models were built using PSE stories that had been coded by expert human coders, one using LIWC scores as inputs, and the other using MSSR variables as inputs. The MSSR model had roughly 80% accuracy, but investigations of reliability between predicted values and human-coded values indicated that the model did not reach a level sufficient to consider the automated coding to be interchangeable with the human coding. The LIWC model had roughly 75% accuracy, and reliability with human coders that would be considered very poor.

Several noteworthy points emerge from investigation of the 10 most influential MSSR variables. First, because synsets are organized by part of speech, words that we think of as synonymous (e.g. "win" vs. "winner") must be identified as being related, because they do not disambiguate into the same synset. Second, not every influential word is a positive predictor of NAch. Highly#r#3, for example appears to indicate the absence of NAch imagery. Third, importance of a synset is calculated by the extent to which a change in that variable leads to a change in the model's prediction. As such, very rare words, like valet#n#1, whose presence dovetails with the presence or absence of NAch imagery, will be marked as highly influential even if they rarely have high values.

The LIWC model was also interesting in that it replicated some of the findings of Schultheiss's regression model (2013) namely that optimism, achievement and metaphysical concerns were all strong predictors of achievement. Although optimism, with words like "win" and achievement make intuitive sense, the idea that metaphysical concerns would be related to achievement, and that such a finding would be a replication of previous research, does not make intuitive sense. An investigation of instances when the presence of metaphysical concerns coincides with achievement imagery makes clear that these stories are not more philosophical, but simply contain examples such as "she *prays* that she will succeed," or "this could be a breakthrough that will stop all of the *death* and destruction" or "no one had *faith* in him, but he kept working hard." The importance of word count may be related to the finding by Pennebaker and King that people high on NAch use more 6+ letter words, as in each case this might suggest a behavioral manifestation of NAch, or it could simply indicate that sentences with few words have fewer opportunities to demonstrate any motive imagery.

Part 2

Overview

Part 2 was a motive induction study for the Need for Achievement modeled closely after the original motive induction study performed by McClelland and colleagues (1949). In that study, an experimenter asks participants to perform a series of arbitrary tasks that could ostensibly be interpreted as measures of mental ability. The original tasks chosen by McClelland et al. included an anagram task, a word scramble[8] and copying words in different orientations, such as backwards. In the present study, we retained the anagram task but replaced the other tasks with a Hidden Figures task that has proven capable of evoking the sort of ego-involvement necessary to arouse NAch. After performing the two mental ability tasks with instructions designed to lead to high or low arousal of the achievement motive, participants completed 8 PSE stories using a new picture set designed to be particularly relevant for NAch. These PSEs were then used to produce 2 more neural network models, using the shortened list of MSSR variables from part 1, and LIWC. Further, the coded PSEs were used to validate the prediction algorithms produced in part 1.

---

[8] It is unclear how anagrams differ from word scrambles. It is possible that the term refers to scrambled sentences, in which complete words are presented out of order and must be rearranged.

Method

Participants. Participants were 180 introductory psychology students as a large Midwestern research institution. Gender was distributed fairly evenly (44.17% female). Mean age was 18.55 (*SD* = 0.87), and participants were mostly Caucasian (83.44%) and African American (12.27%) with smaller numbers of participants identifying as Hispanic (3%), Asian (1.84%) and other (1.84%).

Materials.

Mental Ability Tasks. The mental ability task consisted of a series of anagrams and the Hidden Figures task presented in a paper packet. The anagrams were 50 4-letter words, each of which could have its letters rearranged to form a different word. For example, the word "sink" can be rearranged to form the word "skin". Although the individual words were not very challenging, the 5 minute time limit was sufficient to ensure that no participant was able to complete the entire task (see Appendix V for the complete list of anagrams used). The other ability task was the hidden figures task, which is sometimes called a filed dependence task. The hidden figures task is a series of geometric puzzles that require participants to find a simple shape, such as a cross or diamond, embedded within a more complex shape. The hidden figures task has historically been used as an indication of intelligence, and has also been found to elicit ego-involved behavior when it is accurately identified as a measure of intelligence (Ryan, 1982).

Picture Story Exercise (PSE). The PSE consisted of a series of 8 writing exercises in which participants were instructed to view a picture of an ambiguous scene, and write a story based on that picture. Each image was be presented for 8 seconds, and the writing task will require participants to write for a minimum of 3 minutes and a maximum of 4 minutes. The instructions were taken from the best-practices suggestions of Schultheiss and Pang (2007), and the images were taken from the new Achievement picture set compiled by Ramsay and Pang (2013).

As a manipulation check, all PSE stories were hand coded for NAch by two coders trained on the Achievement portion of the Winter coding system. Motive coding was independently produced for each story by 2 coders who had been trained with standardized training materials, and were regularly tested against the work of an expert coder until their work reached the 85% criterion for agreement with the expert coder to be considered expert coders themselves (Winter, 1994). They also met after coding the first 10 sets of PSEs from this study to discuss points of disagreement. Consistent with best practices, coders organized their coding by PSE image rather than by participant, which prevents the motive content of one story written by a given participant from coloring the coding of her subsequent stories. An achievement motive imagery score was produced by averaging the scores from both coders.

Procedure

A female, undergraduate experimenter greeted groups of participants in a campus computer lab in groups ranging in size from 14 to 22 participants[9] (M=19.38, SD=5.62). All participants were then asked to silence any cellular phones, log into a computer survey that contained the PSE and demographic questions, and

[9] Except for the first 2 groups, which consisted of 7 and 8 participants, respectively, because those groups also served as a pilot study.

provide written consent. Depending on condition, groups of participants were next provided with one of two sets of instructions for the next task.

The arousal condition was modeled after McClelland et al.'s "success" arousal condition. Included in the instructions were the following statements:

> In the first part, you will be working on a couple of activities that measure different components of your mental ability, or IQ. Psychological research shows that all sorts of mental abilities are related to each other. Because of that, a lot of psychologists have come to the conclusion that some people are just smarter than others. IQ predicts all sorts of positive life outcomes, like doing well in school, being successful at whatever job you end up in, and being a good leader. The assessments of mental ability that we have chosen, in addition to being related to your overall IQ, measure your capacity to organize material, your ability to evaluate situations quickly and accurately, and so are particularly relevant to seeing whether you have what it takes to be a leader.
>
> Our previous testing shows that University of Missouri students generally do well on these tests, but that's just on average, and plenty of students perform well below average. After each ability measure, I'll let you know how well MU students typically do on that task so you can see how well you are doing compared to other MU students.

To mimic the frequent feedback from the many small tasks devised by McClelland and colleagues, the experimenter interrupted the participants once every minute to ask them to record their progress up until that point on the bottom of their page.

The instructions in the control condition emphasized low pressure. Whereas the experimenter had been instructed to behave in a professional and slightly "cold" manner in the experimental condition, in the control condition, she was instructed to behave in a more relaxed manner, including making a joke that was written into the script. Included in the instructions were the following statements:

> In our lab, one of the grad students is thinking about using some new measures in his research someday. Right now, though, we don't know how people usually do on them, which questions are easier than others, or anything like that. We can guess, but it's better to just try things out and see. So in this part of the study, all we want is for you guys to work on a couple different tasks so we can see what typical responses on them look like, and also so we can see what the range of responses tends to be. The activities are different kinds of paper and pencil puzzles, and he's pretty sure you should be able to have fun with them.

After the "ability" tasks were over, participants were introduced to the PSE as a test of creative imagination. Participants then completed 8 PSE stories using Qualtrics online survey software (Qualtrics, Provo, UT), timed to display each image for 8 seconds, and then allow 4 minutes to write each story before allowing the participant to advance. Finally, participants provided basic demographic information.

Results

Reliability between the two coders, as measured by the interclass-correlation coefficient was adequate (*ICC*=.74).

Attempting to predict condition from human coded implicit motive imagery using binary logistic regression yielded very poor model fit (β=.07, *n.s.)*.

Maximum Synset-to-Sentence Relatedness Neural Network. A neural network was generated based on inputs from the MSSR variables, with 70% of cases partitioned into the training sample and 30% of cases partitioned into the testing sample. It had 7 nodes plus the mandatory bias node in the hidden layer. In the training sample, the model had a sensitivity of 44.66% and a specificity of 62.87%, for an overall accuracy of 54.99%. In the testing sample, the model had a sensitivity of 42.79% and a specificity of 64.06%, for an overall accuracy of 53.68%. Kappa was statistically significant due to the large number of sentences under investigation, but was not practically significant (Kappa=.07, *p*<.001).

LIWC Neural Network. The LIWC-based model was trained using all variables in the input layer, with settings identical to those used to train the MSSR model. The architecture of the neural-network had 11 nodes in the hidden layer plus the bias node. In the training sample, the model had a sensitivity of 32.81% and a specificity of 69.37%, for an overall accuracy of 51.42%. In the testing sample, the model had a sensitivity of 31.74% and a specificity of 72.51%, for an overall accuracy of 52.93%. Kappa was again statistically significant, but was not practically significant (Kappa=.03, *p*<.01).

Due to poor model fit, neither model was investigated in greater detail.

Predicting NAch Using Networks from Part 1. A final step was to check for agreement between ratings of NAch imagery as determined by the human coders and as determined by each of the two prediction algorithms generated in part 1. Human coded NAch was significantly correlated with predictions from both the MSSR-based neural network (r=.63, p<.001) and the LIWC-based neural network (r=.66, p<.001).

Discussion

Despite modeling the induction study after that of McClelland et al. (1949), part 2 appears to be a failure in that the motive induction did not produce the expected increase in NAch imagery, and similarly did not produce differences in writing style that were detectable using model construction based in MSSR or LIWC. It is possible to propose reasons for those failures in hindsight, though such reasons are, of course, highly speculative.

One reason for the failure of the experimental manipulation to influence motive imagery could be the use of an undergraduate research assistant as the experimenter. McClelland et al. used graduate students to deliver their instructions. The use of an experimenter who is not a peer, is a little older, and might be seen as an authority figure, might make the experimental condition more impactful. In support of this assumption, experimenters who are not figures of authority are considered ideal for measuring natural levels of implicit motives, while professor experimenters generally lead to unstable NAch levels, presumably because they lead to heightened levels for all participants (Lundy, 1988). It is also possible that only highly scholastically motivated students are motivated by assessments of cognitive ability. In describing his use of the Embedded Figures task to generate ego-involvement, Ryan (1982) noted that the undergraduates at his institution (University of Rochester) were highly motivated, and so his manipulation might not extend to student populations that are less ego-involved in their high ability. Given that McClelland et al.'s original motive induction study was performed at Wesleyan University, Ryan's logic might extend to that study as well.

Nonetheless, part 2 had considerable value in that it was able to provide evidence in support of the viability of the two neural network prediction equations developed in part 1. Both neural networks correlated strongly with human coded NAch, though not quite strongly enough to replace human coding. While not perfect, the finding that the neural network models did a reasonably good job of predicting achievement imagery despite that imagery being generated from new picture cues suggests that the neural networks are reasonably robust in a way that was not true of the computer coding schemes developed in the 1960's by Litwin and Smith (D. Winter, personal communication, October 29, 2014).

Part 3

Overview

In part 3, I obtained further coded PSE data, this time from the Henry A. Murray Research Archive, to provide an additional test the neural network prediction models developed in parts 1 and 2. The minimum criteria necessary to use data from the Murray Library for part 3 were that the data include electronic copies of PSE stories (rather than paper, which were only accessible on site), and a spreadsheet documenting the NAch score of those stories. Only a single data set, collected by Ratliff (1979), met those criteria, and so only those data were included in Part 3. A flaw in those data for the present purposes is that the PSE data was coded using the earlier NAch coding scheme created by McClelland and Atkinson, rather than the Winter system. The data is still of interest, however, in part because the McClelland NAch coding system is correlated with the Winter system at r=.54, (Winter, 1991), so a correlation between predicted NAch scores trained on the Winter system and hand-coded NAch scores produced by the McClelland system would demonstrate convergent validity. The data set also contains variables, such as Self Esteem, and Fear of Success, which could demonstrate predictive validity.

Method

Participants. The study on which this archival analysis was based is an unpublished follow-up study to Romer (1973), which sought to study the development of achievement motivation in school children. Data were collected by Romer before 1972 from children of different grade levels, ranging from 5th to 11th grade. The original study consisted of 386 participants, but due to attrition, only 114 participants participated in this follow up. The participants were Caucasian, middle-class high-school or college students, almost all of which had been in either the 5th grade or 7th grade samples in the original study.

Materials.

Verbal PSE Story Cues. Participants were given brief descriptions of a scene, rather than being shown pictures, and asked to write a story about each one using the standard cues recommended by McClelland et al. (1958; see also Schultheiss & Pang, 2007). Participants were provided with the following 5 cues (Ratliff, 1979)

1. A young woman (man) is resting with her (his) head in her (his) hands.
2. A young woman (man) is walking home from school with some books under her (his) arm.
3. A young woman (man) sitting at a desk.
4. An older person is talking to a younger person about something important.
5. At the end of the school year, Ann (John) finds herself (himself) at the top of her (his) college class. OR; After first term finals Ann (John) finds herself (himself) at the top of her (his) medical school class. Most participants received the medical school cue, 40 participants received the college cue.

In addition to being coded for NAch using McClelland's original coding system, they were also coded for fear of success, the power motive (using both the Winter and the Veroff coding scheme), and the affiliation motive.

MSSR and LIWC Prediction Models. NAch was predicted using the models created in Parts 1 and 2, applied to the archival PSE data. These predictions were made on a sentence-by-sentence basis, and then summed per participant, such that a participant who, for example, had 6 instances of NAch imagery across all stories would receive a score of 6. This calculation was designed to be as similar as possible to the motive imagery composites produced by Ratliff (1979).

Debilitating Anxiety Questionnaire (DAQ). The DAQ is a modified form of the debilitating anxiety subscale from the Achievement Anxiety Test (Alpert & Harber, 1960). This scale measures the extent to which achievement-relevant anxiety prevents achievement.

Self-Esteem Inventory (SEI). The SEI (Coopersmith, 1959) is a measure of self esteem that was developed to be appropriate for use with children. It contains short descriptions of people, who the respondent indicates is either “like me” or “not like me”. It contains 4 subscales and a “lie” scale. Of those scales, the General Self, and School Academic scales are the only ones that are potentially relevant to the achievement motive, and so only those subscales were computed.

Results

Both models from Study 1 were significantly related to human NAch coding to about the same degree (see table 2), with the MSSR model correlation at r=.52 (p<.001) and the LIWC model correlating at r=.57 (p<.001). They also correlated with Need for Power motive imagery as measured by the updated, Winter scoring system, but not as measured by the original Veroff system. The models produced in Part 2 were unrelated to NAch.

Because people who write more sentences have the opportunity to demonstrate more achievement imagery, I calculated the number of sentences written by each participant, and included it as a control variable in a partial correlation (see table 2). Major findings were unaffected, though the positive correlations between the models from Parts 1 and 2 disappeared.

Discussion

In Part 3, convergent validity was established between computer coded NAch generated by the two models trained in part 1, and human coded NAch involving PSE stories written to different cues, in a different modality (verbal) and coded using an alternate form of NAch coding. Such a test is the most challenging that has yet been applied to an automated coding method, and the magnitude of the effect is on par with the previously established .54 correlation between NAch as coded by the Winter and McClelland systems (Winter, 1991). Other variables that might have demonstrated predictive validity failed to correlate even with human coded NAch, and so the lack of correlation between those variables and the NAch scores produced by the neural network models neither affirmed nor challenged the validity of those models.

General Discussion

Measuring individual differences in motive dispositions using projective assessments have proven uniquely capable of predicting certain classes of motivated behavior. Those classes of behavior are behaviors that are spontaneous, emotional, or relating to long-term trends in behavior (McClelland, Koestner, & Weinberger, 1989). The labor-intensive nature of human coding of projective assessments have been identified as one reason that those assessments are not used more frequently. In the present research, I tested four neural network models of automated coding for NAch, two trained using LIWC variables as inputs and two trained using variables from the new MSSR procedure. Drawing on inspiration from earlier attempts to create similar automated coding schemes (Litwin, 1965; Smith, 1968) I analyzed the texts at the sentence level, rather than aggregating all PSEs to be analyzed as a whole. The two models trained to predict human coding of NAch using archival coded PSEs performed reasonably well in part 1, with Kappas ranging from .62 (MSSR) to .48 (LIWC), and in part 2, with correlations ranging from .63 (MSSR) to .66 (LIWC). The two models trained to predict condition in a motive arousal study performed nearly at chance, and investigation of human-coded NAch strongly suggested that the achievement motive arousal condition had not led to the anticipated motive arousal. Consistent with these findings, the two models trained to predict human coding performed well, predicting NAch imagery again in part 3, despite the fact that that data set involved a different achievement coding scheme, different images, and even a different format for conveying the PSE image (verbally).

It is interesting and unexpected that the LIWC-based model produced in part 1 performed as well as the MSSR-based model. This is unexpected because the greater ability of the MSSR procedure to differentiate different senses of the same word should theoretically make it a more accurate means of text analysis. One possible explanation is that word sense disambiguation technology is not completely accurate, so where LIWC must accept the possibility of incorrectly accepting a word as, say, achievement relevant (e.g. identifying "best" in "best friend"), MSSR must accept the possibility of incorrectly assigning a word that is meant in a motive relevant sense to a motive-irrelevant synset. A second possible explanation is that some categories of language are relevant to achievement in non-obvious ways. Such a possibility would give LIWC and advantage, because it categorizes texts along a wide variety of categories, whereas MSSR focused on words and phrases that were obviously related to achievement. The idea that non-obvious categories of text are important is supported by the relatively large number of nodes in the hidden layer of the LIWC model from part 1, and the finding that the metaphysical concerns category was highly influential in the model.

Limitations

There exist a variety of areas where there present research was not ideal. One clear failure was the failure of the motive induction study to provide any evidence that it actually aroused any NAch. Therefore, the fact that the models produced in Part 2 did not correlate with NAch in Part 3 should be interpreted as being uninformative about such a method, not as being inconsistent with the proposition that a motive induction study could yield a workable automated coding model.

A related limitation was that the variables used in the MSSR procedure in Part 2 were those found to be related to NAch in Part 1. A truly inductively generated model from a motive arousal study would make few assumptions about the types of words that such motive arousal would create, rather than constraining the types of words that the researcher assumes would differ between conditions.

A further limitation is that the POS-tagger used in this research had not been trained on words processed with WordNet::Tools Compoundify function, meaning that its error rate with those words was particularly high, and so certain important compounds might have been improperly tagged, leading to failures to convert those compounds into synsets. The Stanford Part of Speech Tagger is capable of being trained on new sets of data, though the creation of such a data set, including all WordNet compound words used in a variety of ways, is beyond the scope and resources of the current project. Nonetheless, it is one route to improving the accuracy of the MSSR procedure.

A final limitation was the lack of behavioral variables to provide a measure of predictive validity. The importance of establishing predictive validity in terms of spontaneous behavior, emotional reactions, or long-term behavioral trends cannot be overstated, as those are the ultimate reasons to find the PSE of interest for the study of human behavior.

Future Research

In light of the research limitations, several future directions are obvious, including a more well-designed and extensively pilot tested motive induction study and validation data containing behavioral variables that also correlate with NAch to establish predictive validity.

The other problem identified as a limitation of the present research, that of producing MSSR variables that are learned inductively from a motive arousal study, is actually quite possible. Text classifiers, with an excellent example being the Stanford Classifier (available at *http://nlp.stanford.edu/software/classifier.shtml*), inductively learns to discriminate among examples of pre-coded text during a training phase, and can later be used to categorize uncoded texts. Text classifiers require and possess no pre-existing knowledge of the words or language in the text because they can simply search for regularities of words and phrases that are predictors of categorization. By first processing text into synsets, and then using a text classifier to identify synsets and synset-phrases that distinguish sentences produced by participants in the experimental versus control conditions, it is possible to derive MSSR variables in a more data-driven fashion.

Conclusions

This research is far from producing a finished replacement for human motive coding. It nonetheless provides a novel set of tactics, such as splitting PSEs by sentence, processing raw text with MSSR, and building models using neural networks, that can be useful in generating even more accurate models. MSSR appeared to produce superior predictions in part 1, but when applied to new data sets, it performed on par with the LIWC-based model. It is therefore unclear whether MSSR in it’s current state is the better of the two text processing systems for NAch coding. However, word sense disambiguation and POS-tagging are active areas of research, and as such, MSSR can be expected to become more accurate as those areas advance.

References

Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., & Bong, C. H. (2014). Predicting survey responses: How and why semantics shape survey statistics on organizational behaviour. *PLoS ONE, 9*(9), e106361.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal Of Personality And Social Psychology*, *83*(2), 380-393.

Atkinson, J. W., Heyns, R. W., & Veroff, J. (1954). The effect of experimental arousal of the affiliation motive on thematic apperception. *The Journal of Abnormal and Social Psychology*, *49*(3), 405.

Atkinson, J. W., & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *The journal of abnormal and social psychology*, *60*(1), 52.

Atkinson, J. W., & McClelland, D. C. (1948). The projective expression of needs. II. The effect of different intensities of the hunger drive on Thematic Apperception. *Journal Of Experimental Psychology*, *38*(6), 643-658.

Banerjee, S., & Pedersen, T. (2003). Extended gloss overlaps as a measure of semantic relatedness. In *IJCAI* (Vol. 3, pp. 805-810).

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: some questions and answers. *Trends in cognitive sciences*, *9*(4), 159-162.

Berridge, K. C. (1996). Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews*, *20*(1), 1-25.

Blankenship, V. (2010). Computer-based modeling, assessment, and coding of implicit motives. *In* O.C. Schultheiss & J.C. Brunstein (Eds.), Implicit Motives. New York, NY: Oxford University Press, pp. 186-208.

Brickenkamp, R., & Zillmer, E. (1998). *The d2 test of attention*. Hogrefe & Huber Pub.

Bruner, J. S., & Postman, L. (1948). Symbolic value as an organizing factor in perception. *The Journal of social psychology*, *27*(2), 203-208.

Brunstein, J. C., Schultheiss, O. C., & Grässman, R. (1998). Personal goals and emotional well-being: The moderating role of motive dispositions. *Journal Of Personality And Social Psychology*, *75*(2), 494-508.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, *6*(1), 3-5.

Cox, W., Fadardi, J., & Pothos, E. M. (2006). The Addiction-Stroop Test: Theoretical Considerations and Procedural Recommendations. *Psychological Bulletin*, *132*(3), 443-476.

deCharms, R. & Plimpton, F. (1992). The origin scoring system. In C. R. Smith (Ed.), *Motivation and Personality: Handbook of thematic content analysis* (pp. 334 – 375). Cambridge: Cambridge University Press.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, *82*(1), 62.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of experimental social psychology*, *33*(5), 510-540.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal Of Personality And Social Psychology*, *83*(6), 1441-1455.

Field, M., Mogg, K., & Bradley, B. P. (2006). Attention to Drug-Related Cues in Drug Abuse and Addiction: Component Processes. In R. W. Wiers, A. W. Stacy (Eds.) , *Handbook of implicit cognition and addiction* (pp. 151-163). Thousand Oaks, CA, US: Sage Publications, Inc.

Freud, S. (*1999*). The interpretation of dreams (J.Crick Trans.). New York, NY: Oxford University Press. (Original work published 1909)

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology*, *85*(2), 197.

Groysberg, B., Thomas, D. A., & Tydlaska, J. (2011). oDesk: Changing How the World Works. *Harvard Business School Organizational Behavior Unit Case*, (411-078).

Heckhausen, H. (1963). *Hoffnung und Furcht in der Leistungsmotivation* (Vol. 6). Anton Hain.

Higgins, T.E., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*(2), 141-154.

Hogenraad, R. (2003). The words that predict the outbreak of wars. *Empirical studies of the Arts*, *21*(1), 5-20.

Hogenraad, R. (2005). What the Words of War Can Tell Us About the Risk of War. *Peace and Conflict: Journal of Peace Psychology*, *11*(2), 137.

Johnson, B. W., Redfield, D. L., Miller, R. L., & Simpson, R. E. (1983). The Coopersmith self-esteem inventory: A construct validation study. *Educational and Psychological Measurement*, *43*(3), 907-913.

Joshi, V. V., Dad, M. R., Bagul, Y. S., & Haribhakta, Y. V. (2014) Literature Survey on Disambiguating Polysemous Queries for Document Retrieval. *International Journal of Engineering Science and Innovative Technology, 3*(3), 361-369.

Kolhatkar, V. (2009). An extended analysis of a method of all words sense disambiguation (Doctoral dissertation, University of Minnesota).

Litwin, G. H. (1965). *The language of achievement: An analysis of achievement related themes in fantasy using mechanical methods*. Unpublished doctoral dissertation, Harvard University.

Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1* (pp. 63-70). Association for Computational Linguistics.

Maio, G. R. and Haddock, G. (2007). Attitude change. In: Kruglanski, A. and Higgins, E. eds. *Social Psychology: Handbook of Basic Principles (2nd ed.)*. New York, NY: Guilford Press, pp. 565-586.

Manning, C. & Schütze, H. *Foundations of Statistical Natural Language Processing*.

The MIT Press, Cambridge, MA, 1999.

Maroco, J., & Bártolo-Ribeiro, R. (2013). Selection of Air Force pilot candidates: A case study on the predictive accuracy of discriminant analysis, logistic regression, and four neural network types. *The International Journal of Aviation Psychology*, *23*(2), 130-152.

McAdams, D. P. (1980). A thematic coding system for the intimacy motive. *Journal of research in personality*, *14*(4), 413-432.

McAdams, D. P., & Vaillant, G. E. (1982). Intimacy Motivation and Psychosocial Adjustment: A Longitudinal Study. *Journal Of Personality Assessment*, *46*(6), 586.

McClelland, D. C. (1965). N achievement and entrepreneurship: A longitudinal study. *Journal Of Personality And Social Psychology*, *1*(4), 389-392.

McClelland, D. C. (1975). *Power: The inner experience*. New York: Irvington Publishers.

McClelland, D. C. (1985). How motives, skills, and values determine what people do. *American Psychologist*, *40*(7), 812-825.

McClelland, D. C. (1987). Characteristics of successful entrepreneurs. *The Journal Of Creative Behavior*, *21*(3), 219-233.

McClelland, D. C., Clark, R. A., Roby, T. B., & Atkinson, J. W. (1949). The projective expression of needs. IV. The effect of the need for achievement on thematic apperception. *Journal of Experimental Psychology*, *39*(2), 242.

McClelland, D. C., Koestner, R., & Weinberger, J. (1992). How do self-attributed and implicit motives differ?. In C. P. Smith, J. W. Atkinson, D. C. McClelland, J.

Veroff (Eds.) , *Motivation and personality: Handbook of thematic content analysis* (pp. 49-72). New York, NY, US: Cambridge University Press.

Mihalcea R, Corley C, Strapparava C (2006) Corpus-based and knowledge-based measures of text semantic similarity. AAAI 6: 775–780.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39-41.

Moskowitz, G. B. (2002). Preconscious effects of temporary goals on attention. *Journal of Experimental Social Psychology, 38*(4), 397-404.

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychological bulletin*, *124*(1), 54.

Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., & Tucker, G. J. (1988). Diagnostic classification through content analysis of patients' speech. *American Journal of Psychiatry*, *145*(4), 464-468.

Pang, J. S. (2006). A revised content-coding measure for hope of success (HS) and fear of failure (FF). Unpublished dissertation, University of Michigan, Ann Arbor.

Pang, J.S. (2010). The Achievement Motive: A Review of Theory and Assessment of N Achievement, Hope of Success, and Fear of Failure. *In* O.C. Schultheiss & J.C. Brunstein (Eds.), Implicit Motives. New York, NY: Oxford University Press, pp. 30-70.

Pedersen, T. (2012, June). Duluth: Measuring degrees of relational similarity with the gloss vector measure of semantic relatedness. In *Proceedings of the First*

Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (pp. 497-501). Association for Computational Linguistics.

Pedersen, T., & Kolhatkar, V. (2009). WordNet:: SenseRelate:: AllWords: a broad coverage word sense tagger that maximizes semantic relatedness. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics, companion volume: Demonstration session* (pp. 17-20). Association for Computational Linguistics.

Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, *211*(2828), 42-45.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*, 2001.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, *77*(6), 1296-1312.

Perkins, J. (2014). *Python 3 Text Processing with NLTK 3 Cookbook*. Packt Publishing Ltd.

Ramsay, J. E., & Pang, J. S. (2013). Set ambiguity: A key determinant of reliability and validity in the picture story exercise. *Motivation and Emotion*, *37*(4), 661-674.

Ratliff, Elyse. (1979). Data collection on a follow-up study of achievement-related motives. Unpublished final report, submitted to the Henry A. Murray

Research Center.

Richard, M. D., & Lippmann, R. P. (1991). Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural computation*, *3*(4), 461-483.

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal Of Behavior Therapy And Experimental Psychiatry*, *38*(2), 105-120.

Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121-1133.

Russo FA, Vempala NN, Sandstrom GM. (2013). Predicting musically induced emotions from physiological inputs: linear and neural network models. *Frontiers in Psychology*. 4:468. doi:10.3389/fpsyg.2013.00468.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, *43*(3), 450.

Schnurr, P. P., Rosenberg, S. D., & Oxman, T. E. (1992). Comparison of TAT and free speech techniques for eliciting source material in computerized content analysis. *Journal of personality assessment*, *58*(2), 311-325.

Schnurr, P. P., Rosenberg, S. D., & Oxman, T. E. (1993). Issues in the comparison of techniques for eliciting source material in computerized content analysis. *Journal of personality assessment*, *61*(2), 237-242.

Schultheiss, O. C. (2008). Implicit motives. In O. P. John, R. W. Robins & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (3 ed., pp. 603-633).

New York: Guilford.

Schultheiss, O. C. (2013). Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Frontiers in psychology*, *4,* 1-20.

Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. *Handbook of research methods in personality psychology*, 322-344.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, *8*(9), e73791.

Sheldon, K. M. (2011). Integrating behavioral-motive and experiential-requirement perspectives on psychological needs: A two process model. *Psychological review*, *118*(4), 552.

Sheldon, K. M., Prentice, M., Halusic, M., & Schüler, J. (2015). Matches between assigned goal-types and both implicit and explicit motive dispositions predict goal self-concordance. *Motivation and Emotion*, *39*(3), 335-343.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Slabbinck, H., De Houwer, J., & Van Kenhove, P. (2011). A pictorial attitude IAT as a measure of implicit motives. *European Journal of Personality*, *25*(1), 76-86.

Slabbinck, H., De Houwer, J., & Van Kenhove, P. (2013). Convergent, discriminant, and incremental validity of the pictorial attitude implicit association test and the picture story exercise as measures of the implicit power motive.

European Journal of Personality, *27*(1), 30-38.

Slatcher, R. B., Vazire, S., & Pennebaker, J. W. (2008). Am “I” more important than “we”? Couples’ word use in instant messages. *Personal Relationships*, *15*(4), 407-424.

Smith, M. S. (1968). The computer and the TAT. Journal of School Psychol. 6, 206–214.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin*, *112*(1), 140-154.

Sokolowski, K., Schmalt, H. D., Langens, T. A., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: The Multi-Motive Grid (MMG). *Journal of Personality Assessment*, *74*(1), 126-145.

Stacy, A., & Wiers, R. (2010). Implicit cognition and addiction: a tool for explaining paradoxical behavior. *Annual Review Of Clinical Psychology*, *6*551-575.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24-54.

Thrash, T. M., Cassidy, S. E., Maruskin, L. A., & Elliot, A. J. (2010). Factors that influence the relation between implicit and explicit motives: A general implicit-explicit congruence framework. *Implicit motives*, 308-346.

Thrash, T. M., Elliot, A. J., & Schultheiss, O. C. (2007). Methodological and dispositional predictors of congruence between implicit and explicit need for achievement. *Personality and Social Psychology Bulletin*, *33*(7), 961-974.

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 173-180). Association for Computational Linguistics.

Toutanova, K., & Manning, C. D. (2000). Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13* (pp. 63-70). Association for Computational Linguistics.

Triandis, H. C. (1980). Values, attitudes, and interpersonal behavior. In H. E. Howe, Jr. & M. M. Page (Eds.), *Nebraska Symposium on Motivation: Beliefs, altitudes, and values, 1979* (Vol. 27, pp. 195- 259). Lincoln: University of Nebraska Press.

Williamson, J. Computer scoring of n-Affiliation. Unpublished manuscript, 1965.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological review*, *107*(1), 101.

Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text.

Winter, D. G. (1994). Manual for scoring motive imagery in running text (Version 4.2). Department of Psychology, University of Michigan. *Ann Arbor: Unpublished Manuscript*.

Winter, D. G. (1998). The contributions of David McClelland to personality assessment. *Journal of Personality Assessment*, *71*(2), 129-145.

Woike, B., & Polo, M. (2001). Motive-Related Memories: Content, Structure, and Affect. *Journal of Personality*, *69*(3), 391-415.

ace#n#3	They#CL#
achieve#v#1	be#v#1
challenge#v#4	high_school#n#1
best#r#1	boy#n#1
	play#v#1
	a#CL#
	basketball#n#1
	game#n#1
	.#IT#
	The#CL#
	man#n#10
	in#IT#
	white#n#11
	be#v#1
	on#IT#
	defense#n#6
	and#IT#
	the#CL#
	man#n#10
	in#IT#
	black#n#6
	be#v#1
	on#IT#
	offense#n#3
	.#IT#

Figure 1. Input spreadsheet for the MSSR procedure. Synsets of interest are in the first column and disambiguated sentences are input (one word at a time) in the second column with three spaces between them. This format is similar to the output file created by WordNet::SenseRelate::AllWords.

ace#n#3 be#v#1
ace#n#3 high_school#n#1
ace#n#3 boy#n#1
ace#n#3 play#v#1
ace#n#3 basketball#n#1
ace#n#3 game#n#1
achieve#v#1 be#v#1
achieve#v#1 high_school#n#1
achieve#v#1 boy#n#1
achieve#v#1 play#v#1
achieve#v#1 basketball#n#1
achieve#v#1 game#n#1
challenge#v#4 be#v#1
challenge#v#4 high_school#n#1
challenge#v#4 boy#n#1
challenge#v#4 play#v#1
challenge#v#4 basketball#n#1
challenge#v#4 game#n#1
best#r#1 be#v#1
best#r#1 high_school#n#1
best#r#1 boy#n#1
best#r#1 play#v#1
best#r#1 basketball#n#1
best#r#1 game#n#1
ace#n#3 man#n#10
ace#n#3 white#n#11
ace#n#3 be#v#1
ace#n#3 defense#n#6
ace#n#3 man#n#10
ace#n#3 black#n#6
ace#n#3 be#v#1
ace#n#3 offense#n#3
achieve#v#1 man#n#10
achieve#v#1 white#n#11
achieve#v#1 be#v#1
achieve#v#1 defense#n#6
achieve#v#1 man#n#10
achieve#v#1 black#n#6
achieve#v#1 be#v#1
achieve#v#1 offense#n#3
challenge#v#4 man#n#10
challenge#v#4 white#n#11
challenge#v#4 be#v#1
challenge#v#4 defense#n#6
challenge#v#4 man#n#10
challenge#v#4 black#n#6

challenge#v#4 be#v#1
challenge#v#4 offense#n#3
best#r#1 man#n#10
best#r#1 white#n#11
best#r#1 be#v#1
best#r#1 defense#n#6
best#r#1 man#n#10
best#r#1 black#n#6
best#r#1 be#v#1
best#r#1 offense#n#3

Figure 2. The first of two output files from the first step in the MSSR procedure. This is a space delimited file that can be directly input into WordNet::Similarity.

1	1	ace#n#3	be#v#1
1	1	ace#n#3	high_school#n#1
1	1	ace#n#3	boy#n#1
1	1	ace#n#3	play#v#1
1	1	ace#n#3	basketball#n#1
1	1	ace#n#3	game#n#1
1	2	achieve#v#1	be#v#1
1	2	achieve#v#1	high_school#n#1
1	2	achieve#v#1	boy#n#1
1	2	achieve#v#1	play#v#1
1	2	achieve#v#1	basketball#n#1
1	2	achieve#v#1	game#n#1
1	3	challenge#v#4	be#v#1
1	3	challenge#v#4	high_school#n#1
1	3	challenge#v#4	boy#n#1
1	3	challenge#v#4	play#v#1
1	3	challenge#v#4	basketball#n#1
1	3	challenge#v#4	game#n#1
1	4	best#r#1	be#v#1
1	4	best#r#1	high_school#n#1
1	4	best#r#1	boy#n#1
1	4	best#r#1	play#v#1
1	4	best#r#1	basketball#n#1
1	4	best#r#1	game#n#1
2	1	ace#n#3	man#n#10
2	1	ace#n#3	white#n#11
2	1	ace#n#3	be#v#1
2	1	ace#n#3	defense#n#6
2	1	ace#n#3	man#n#10
2	1	ace#n#3	black#n#6
2	1	ace#n#3	be#v#1
2	1	ace#n#3	offense#n#3
2	2	achieve#v#1	man#n#10
2	2	achieve#v#1	white#n#11
2	2	achieve#v#1	be#v#1
2	2	achieve#v#1	defense#n#6
2	2	achieve#v#1	man#n#10
2	2	achieve#v#1	black#n#6
2	2	achieve#v#1	be#v#1
2	2	achieve#v#1	offense#n#3
2	3	challenge#v#4	man#n#10
2	3	challenge#v#4	white#n#11
2	3	challenge#v#4	be#v#1
2	3	challenge#v#4	defense#n#6

2	3	challenge#v#4	man#n#10
2	3	challenge#v#4	black#n#6
2	3	challenge#v#4	be#v#1
2	3	challenge#v#4	offense#n#3
2	4	best#r#1	man#n#10
2	4	best#r#1	white#n#11
2	4	best#r#1	be#v#1
2	4	best#r#1	defense#n#6
2	4	best#r#1	man#n#10
2	4	best#r#1	black#n#6
2	4	best#r#1	be#v#1
2	4	best#r#1	offense#n#3

Figure 3. The second of two output files from the first step in the MSSR procedure. This is a comma delimited file that keeps track of the sentence number in the first column, and the comparison word number in the second column.

1	1	ace#n#3	be#v#1	0.188027672
1	1	ace#n#3	high_school#n#1	0.037982347
1	1	ace#n#3	boy#n#1	0.062156237
1	1	ace#n#3	play#v#1	0.15979256
1	1	ace#n#3	basketball#n#1	0.08157315
1	1	ace#n#3	game#n#1	0.085343656
1	2	achieve#v#1	be#v#1	0.362210458
1	2	achieve#v#1	high_school#n#1	0.115177434
1	2	achieve#v#1	boy#n#1	0.102956552
1	2	achieve#v#1	play#v#1	0.198779241
1	2	achieve#v#1	basketball#n#1	0.123596746
1	2	achieve#v#1	game#n#1	0.135041013
1	3	challenge#v#4	be#v#1	0.263412064
1	3	challenge#v#4	high_school#n#1	0.038786575
1	3	challenge#v#4	boy#n#1	0.080604498
1	3	challenge#v#4	play#v#1	0.107215782
1	3	challenge#v#4	basketball#n#1	0.067418087
1	3	challenge#v#4	game#n#1	0.095271577
1	4	best#r#1	be#v#1	0.106998165
1	4	best#r#1	high_school#n#1	0.019167419
1	4	best#r#1	boy#n#1	0.038987214
1	4	best#r#1	play#v#1	0.062977284
1	4	best#r#1	basketball#n#1	0.035990019
1	4	best#r#1	game#n#1	0.040652305
2	1	ace#n#3	man#n#10	0.06523304
2	1	ace#n#3	white#n#11	0.057923173
2	1	ace#n#3	be#v#1	0.188027672
2	1	ace#n#3	defense#n#6	0.098701509
2	1	ace#n#3	man#n#10	0.06523304
2	1	ace#n#3	black#n#6	0.057311064
2	1	ace#n#3	be#v#1	0.188027672
2	1	ace#n#3	offense#n#3	0.083950348
2	2	achieve#v#1	man#n#10	0.123341399
2	2	achieve#v#1	white#n#11	0.115750311
2	2	achieve#v#1	be#v#1	0.362210458
2	2	achieve#v#1	defense#n#6	0.178845803
2	2	achieve#v#1	man#n#10	0.123341399
2	2	achieve#v#1	black#n#6	0.112317047
2	2	achieve#v#1	be#v#1	0.362210458
2	2	achieve#v#1	offense#n#3	0.165897802
2	3	challenge#v#4	man#n#10	0.077144546
2	3	challenge#v#4	white#n#11	0.07041964
2	3	challenge#v#4	be#v#1	0.263412064
2	3	challenge#v#4	defense#n#6	0.123147159

2	3	challenge#v#4	man#n#10	0.077144546
2	3	challenge#v#4	black#n#6	0.069002439
2	3	challenge#v#4	be#v#1	0.263412064
2	3	challenge#v#4	offense#n#3	0.142510087
2	4	best#r#1	man#n#10	0.030774122
2	4	best#r#1	white#n#11	0.028059208
2	4	best#r#1	be#v#1	0.106998165
2	4	best#r#1	defense#n#6	0.047500514
2	4	best#r#1	man#n#10	0.030774122
2	4	best#r#1	black#n#6	0.027029551
2	4	best#r#1	be#v#1	0.106998165
2	4	best#r#1	offense#n#3	0.041370496

Figure 4. This is a comma delimited spreadsheet that combine the spreadsheet in Figure 3 with the similarity scores produced by WordNet::Similarity.

Sentence	Word1	Word2	Word3	Word4
1	0.188027672	0.362210458	0.263412064	0.106998165
2	0.188027672	0.362210458	0.263412064	0.106998165

Figure 5. The final output from the MSSR procedure. The numbers in the first column indicate a given sentence. The subsequent columns indicate the maximum similarity of all words in a sentence with a particular synset of interest.

Table 1

The 10 most influential variables in the MSSR neural network model

Synset	Gloss Definition	Instance of Perfect Match (score=1.00)	Instance of Related Word (score<1 and >0.50)
valet#n#1	a manservant who acts as a personal attendant to his employer	The captain is doubtful in his own beliefs but admires the fortitude of the gentleman who wants to stow away (no NAch).	None
flawlessly#r#1	in an adroit manner	They perform their act flawlessly.	They have been performing together for years and are very comfortable with each other which makes their performances flawless.
winner#n#1	the contestant who wins the contest	Both undefeated he is nervous for who will come out the victor.	As they begin to box he easily and confidently wins the match.
endowment#n#1	natural abilities or qualities	His father knew that Joe could use and exercise his talents in more productive ways.	None
breakthrough#n #2	making an important discovery	Judy has a breakthrough and discover a cure for a disease.	Curie is an extremely particular lady and likes everything to be very exact she is watching everything with much attention.
vitally#r#1	to a vital degree	The novel aspect of the fighting had just become vitally important to winning the war which everyone onboard was beginning to realize.	She lived for the adulation as well as for technical mastery of her craft.
highly#r#3	in a high position or level or rank	She is unconvinced but enjoys the attention knowing he is oblivious to her highly infectious fatal disease (No Ach).	This architect will decide that although the plant will provide work for his fellow townsmen it is far too dangerous to complete the plans and put his family in danger (No Ach).
correctly#r#1	in an accurate manner	Next she will mix them and perform the experiment correctly.	The reporter wants good answers so his status as a reporter with truth and accuracy can continue.
brainy#a#1	having or marked by unusual and impressive intelligence	Martha was always brilliant.	In this scene two women scientists who are both highly intelligent and educated in their fields are in some type of conflict.
precisely#r#2	in a precise manner	Here she was flying 50 feet above the big top floor attached to nothing soaring through the air but confident to know that her partner would be there at exactly the right second to grab her and take her back to terra firma.	Not even simple measurements will come out right.

Table 2

Bivariate Correlations among Variables

	1	2	3	4	5	6	7	8	9	10	11
1. Part1MSSR											
2. Part1LIWC	.71**										
3. Part2MSSR	.45**	.34**									
4. Part2LIWC	.55**	.36**	.76**								
5. NAch	.52**	.57**	0.04	0.14							
6. Fear Success	-0.08	-0.05	-.23	-0.07	0.07						
7. NPow (Winter)	.35**	.31**	-0.04	0.09	.32**	.19*					
8. NPow (Veroff)	-0.12	-0.18	0.03	-0.05	-0.18	-0.06	-.37**				
9. NAff	0.09	0.16	.30**	0.17	-0.07	-.21*	0.02	0.03			
10. DAQ	-0.14	-0.16	-0.05	-0.05	-0.06	0.04	0.02	-.26*			
11. SEI GEN	.00	-.02	-.08	-.11	.06	-.13	-.01	-.06	.12	.00	
12. SEI Academic	-.05	-.07	-.03	-.04	-.11	.12	-.08	.12	-.06	.22*	.03

Note: Part1MSSR = predictions from the MSSR-based neural network from part 1, Part1LIWC = predictions from the LIWC-based neural network from part 1, Part2MSSR = predictions from the MSSR-based neural network from part 2, Part2LIWC = predictions from the LIWC-based neural network from part 2, NAch = human coded achievement motive, Fear Success = human coded fear of success, NPow (Winter) = the power motive coded using the coding system developed by Winter, NPow (Veroff) = the earlier power motive coded using the coding system developed by Veroff, NAff = the affiliation motive, DAQ = Debilitating anxiety questionnaire, SEI GEN = Self Esteem Inventory, General subscale, SEI Academic = Self Esteem Inventory, School Academic subscale

Table 3

Partial Correlations among Variables Controlling for Number of Sentences

	1	2	3	4	5	6	7	8	9	10	11
1. Part1MSSR											
2. Part1LIWC	.51**										
3. Part2MSSR	-.17	-.23*									
4. Part2LIWC	-.41**	-.25*	.16								
5. NAch	.53**	.56**	-.21	-.35**							
6. Fear Success	-.04	-.03	.20	-.10	-.01						
7. Npow (Winter)	.33**	.25*	-.06	-.26*	.27*	.14					
8. Npow (Veroff)	-.03	-.14	.13	.09	-0.18	-.05	-.35**				
9. NAff	-.08	.02	-.12	.10	-0.08	-.14	.03	0.02			
10. DAQ	-.07	-.09	.09	.13	.02	.02	.03	-.26*	.00		
11. SEI GEN	.07	-.04	-.09	-.10	.09	-.20	-.05	-.04	.15	-.04	
12. SEI Academic	-.01	-.07	.18	.11	-.14	.06	-.08	.08	-.107	.21	-.01

Note: Variable abbreviations are the same as those used in table 2.

Appendix I: Initial Words of Interest

a great deal	create	feeling	lawyer	perform	talent
a lot	culminated	find	leading	performance	talented
abilities	cure	fine	legendary	person	task
ability	defeat	first	lifetime	plan	time
accomplish	degree	flawlessly	long	potential	tirelessly
accomplished	design	formula	longest	practice	top of her class
accomplishment	desires	fruition	lose	praise	train
achieve	determined	get it right	lost	prize	trained
adulation	develop	give up	make a difference	productive	training
advances	difficult	goal	make millions	professional	tremendous
amazing	difficulty	gold	make no mistakes	progress	trophies
antidote	diligently	good	mankind	promise	try
arise	disappointed	grade	mastery	promote	ultimate
as yet unknown	discover	graduate	match	proud	unique
athlete	discovery	graduate_school	medals	publish	unsuccessful
authority	do	great	medical_school	race	up-and-coming
award-winning	do well	gymnast	messed up	renowned	vaccine
awe	doctor	hard	mission	result	very
become	done right	hard work	mistake	right	victorious
best	easy	Harvard	months	scientist	wants
better	educated	highest	more	seeking	well
big	effective	his best	most	sharpen	well known
biggest	efficient	history	most in history	show	whiz
boxer	effort	hope	most on earth	skills	will
breakthrough	effortless	hours	move up	slow	win
brilliant	entrepreneur	how well	much	smart	without mistakes
career	ever before	important	nation	smartest	wonderful
careful	ever known	improve	natural gift	smooth	work
challenge	ever seen	improvement	neck and neck	solution	work hard
champion	excel	improvement	needs	sportsman	world
college	excellent	innovator	new	star	year
commendable	experiment	inoculate	Nobel	star performer	years
competent	expert	intelligent	opportunity	strive	youngest
competition	fail	intense	passionately	strong	
confident	failure	invent	perfect	study	
confidently	famous	invention	perfected	success	
correct	fastest	job	perfecting	superb	
correctly	feat	knowledge	perfectly	system	

Appendix II: Initial MSSR Variables

able#a#2	pull_off#v#3	adulation#n#1	perform#v#1	discover#v#1	advance#v#8	cogitation#n#2	diligence#n#2
accomplishment#n#1	race#n#1	mastery#n#1	perform#v#4	find#v#9	ahead#a#1	competent#a#2	disappoint#v#1
acme#n#1	remedy#n#2	mastery#n#2	performance#n#1	develop#v#1	ahead#r#5	competition#n#3	discover#v#3
acquire#v#5	science#n#1	vaccine#n#1	work#v#1	new#a#1	ahead#r#7	competitive#a#1	discover#v#7
advancement#n#3	scoop#v#2	inoculate#v#3	study#n#6	first#a#1	alone#a#4	competitive#a#3	discovery#n#3
ambition#n#2	singular#a#6	develop#v#1	result#n#1	graduate#n#1	ambition#n#1	confidence#n#5	disease#n#1
arduous#a#1	skill#n#1	improve#v#1	well#r#2	graduate#v#1	ambition#v#1	confident#a#1	do#v#3
bang-up#a#1	solution#n#5	improvement#n#1	well#r#3	degree#n#1	ambitious#a#1	confident#a#3	dream#n#1
best#n#1	stunt#n#1	improvement#n#2	well#r#4	college#n#1	antidote#n#1	consequence#n#1	drill#v#3
breakthrough#n#2	succeed#v#1	experiment#n#1	well#r#9	college#n#2	assurance#n#1	contest#n#1	earth#n#1
brilliant#a#1	success#n#3	opportunity#n#1	flawlessly#r#1	Harvard#n#1	attention#n#6	contribution#n#5	ease#n#2
competent#a#1	successful#a#1	effort#n#1	right#a#2	become#v#1	authoritative#a#1	convinced#a#1	edison#n#1
diligently#r#1	superb#a#2	degree#n#1	passionately#r#1	doctor#n#1	authority#n#1	correct#a#1	effective#a#1
excellent#a#1	work#v#3	publish#v#1	correct#a#1	lawyer#n#1	authority#n#3	correct#a#3	effective#a#2
first#a#5	right#a#2	publish#v#2	smooth#a#1	professional#n#1	authority#n#7	correctly#r#1	efficient#a#1
flair#n#1	perfectly#r#1	publish#v#3	wonderful#a#1	work#v#1	avail#n#1	create#v#2	einstein#n#1
forwarding#n#2	ultimate#a#1	important#a#1	lifetime#n#1	practice#v#1	beat#v#2	create#v#3	elegance#n#2
fresh#a#4	commendable#a#1	mistake#n#1	tremendous#a#2	practice#v#2	beautifully#r#1	create#v#4	eminent#a#1
gallant#a#3	knowledge#n#1	medical_school#n#1	world#n#1	practice#v#3	belt#n#2	create#v#5	employment#n#2
good#a#1	authority#n#1	graduate_school#n#1	world#n#2	work#n#1	best#a#1	creative#a#1	endeavor#v#1
hard#a#7	confident#a#1	promote#v#2	world#n#3	training#n#1	best#n#2	creative#a#2	endowment#n#1
hard#r#9	confident#a#2	award-winning#a#1	boxer#n#1	train#v#1	best#r#1	crown#v#2	energetic#a#2
intelligent#a#1	confidently#r#1	failure#n#1	boxer#n#2	long#a#1	brainy#a#1	crucial#a#1	entrepreneur#n#1
learn#v#4	star#n#1	failure#n#2	scientist#n#1	hard#a#1	brilliance#n#3	crucial#a#2	epitome#n#2
loser#n#1	star#n#4	failure#n#3	athlete#n#1	years#n#2	bring_around#v#2	crusade#v#1	err#v#1
lost#a#4	ability#n#2	graduate#a#1	sportsman#n#1	much#a#1	bring_around#v#2	culminate#v#1	error#n#6
originally#r#1	promise#n#2	years#n#2	gymnast#n#1	well#r#11	calendar_month#n#1	culminate#v#2	exceed#v#1
outdo#v#2	whiz#n#1	difficulty#n#1	nation#n#1	tremendous#a#2	cancer#n#1	culminate#v#4	exceed#v#2
outplay#v#1	professional#a#1	difficulty#n#2	renowned#a#1	intense#a#1	capable#a#1	decoration#n#2	excel#v#1
outstanding#a#2	entrepreneur#n#1	difficulty#n#3	famous#a#1	ability#n#1	careful#a#3	dedication#n#4	excellently#r#1
perfectionist#n#1	innovator#n#1	difficult#a#1	great#a#1	about#r#7	carefully#r#1	defeated#a#2	execute#v#2
perfectly#r#2	careful#a#1	do#v#1	great#a#2	accompaniment#n#2	cautiously#r#1	degree#n#1	experience#n#1
performance#n#2	careful#a#2	do#v#2	great#a#3	accurate#a#2	celebrated#a#1	deliberate#a#1	expert#n#1
performer#n#1	careful#a#3	do#v#3	accomplished#a#1	ace#n#3	chairwoman#n#1	demote#v#1	expertly#r#1
praise#v#1	great#a#3	do#v#4	first#a#1	achieve#v#1	challenge#n#3	dependent#a#2	expertness#n#1
precisely#r#1	up-and-coming#a#1	do#v#5	right#a#2	achiever#n#1	challenge#v#4	determine#v#1	fabled#a#1
pride#v#1	awe#n#1	do#v#6	formula#n#1	acknowledge#v#6	choice#a#1	develop#v#1	fail#v#7
progress#n#2	great#a#1	do#v#7	formula#n#2	adequate_to#a#1	class#n#2	devote#v#3	failure#n#4
promote#v#1	great#a#2	do#v#8	formula#n#3	advance#a#2	clear#v#9	difficult#a#1	fall_upon#v#1
proud#a#1	great#a#3	do#v#9	formula#n#4	advance#v#1	clock_time#n#	difficulty#n#	fast#a#3

				1	2	
faultlessly#r#1	help_oneself#v#1	long#a#9	perfective#n#1	promote#v#2	smooth#a#3	train#v#2
feat#n#1	high#a#2	long#r#1	perform#v#1	promote#v#5	smoothly#r#2	triumphant#a#2
field#n#3	highly#r#1	lose#v#2	perform#v#3	promotion#n#1	solution#n#2	trophy#n#1
finely#r#3	highly#r#3	lose#v#3	performance#n#3	promotion#n#2	solution#n#3	trophy#n#2
finish#n#4	history#n#2	lose#v#7	performance#n#4	promotion#n#3	solve#v#1	try#v#1
finish#v#4	important#a#1	loss#n#5	perplex#v#1	proper#a#4	stanford_university#n#1	unique#a#4
first#a#1	important#a#5	major#a#5	pioneer#n#1	properly#r#1	star#n#4	uniquely#r#1
first#n#1	impressive#v#3	mania#n#1	place#v#15	qualified#a#1	star#n#6	unsuccessful#a#1
first#n#5	improvement#n#1	mark#n#1	place#v#6	qualify#v#1	star#v#1	unwrap#v#2
flawless#a#1	improvement#n#2	master#n#9	populace#n#1	qualify#v#2	start#n#3	up#r#1
fly#v#13	improvement#n#3	master#v#4	position#n#6	qualify#v#4	strive#v#2	up#r#5
focus#v#2	improving#a#1	mastering#n#2	potent#a#3	quantify#v#2	study#n#2	valet#n#1
gain#v#5	in_the_first_place#r#1	medicine#n#3	potential#a#1	race#n#2	study#n#4	victor#n#1
galileo#n#1	inaugural#a#2	methodically#r#1	potential#n#1	race#v#2	study#n#9	victorious#a#1
gamey#a#2	increase#n#2	mighty#r#1	potion#n#1	race#v#3	study#v#2	victoriously#r#1
genius#n#1	indefatigably#r#1	mistake#n#1	practice#n#1	race#v#4	study#v#3	victory#n#1
genius#n#4	independent#a#3	month#n#2	practice#n#4	rate#v#1	study#v#5	virus#n#3
get_the_better_of#v#1	industriously#r#1	more#a#1	precise#a#1	reach#v#1	study#v#6	vitally#r#1
global#a#1	influence#v#1	more#r#1	precisely#r#2	reach#v#7	success#n#1	well#n#1
goal#n#1	intelligence#n#2	most#a#2	preciseness#n#2	realization#n#6	success#n#2	well#r#11
goal#n#4	invent#v#1	most#r#1	prepare#v#5	recognition#n#5	successfully#r#1	well#r#5
graciously#r#1	inventor#n#1	move#v#14	prepare#v#7	rehearse#v#1	sufficiently#r#1	whiz#n#2
grade_point_average#n#1	job#n#2	new#a#1	pride#n#1	resolve#v#6	surpass#v#2	win#n#1
grade#n#2	job#n#3	new#a#11	pride#n#2	respect#v#1	talent#n#2	win#v#1
grade#v#3	job#n#4	new#a#4	pride#n#3	result#n#3	talented#a#1	winner#n#1
grade#v#4	job#n#6	new#a#5	prize#n#1	result#v#1	target#v#1	winner#n#2
graduate#a#1	know#v#1	nice#a#3	problem#n#1	result#v#3	technical#a#1	winning#n#1
graduate#v#1	know#v#2	occupation#n#1	problem#n#2	retentive#a#1	technical#a#4	work#n#1
great#a#2	know#v#4	oeuvre#n#1	professional#a#2	right#a#7	technique#n#1	work#v#1
great#n#1	know#v#7	opportunity#n#1	professional#a#3	right#a#9	tense#a#1	work#v#5
greatness#n#1	knowing#a#4	original#a#1	professional#a#4	right#r#2	test#v#1	work#v#8
gutenberg#n#1	knowledgeable#a#2	original#a#3	professional#a#5	right#r#4	thoroughly#r#1	workplace#n#1
hanker#v#1	lead#n#15	original#a#4	professional#n#1	rival#n#1	thoroughly#r#2	world#n#8
hard#r#2	lead#v#8	outstanding#a#1	professional#n#2	salute#v#6	thoroughness#n#1	year#n#1
hard#r#4	leading#a#1	paragon#n#1	professionally#r#1	scientist#n#1	time#n#1	year#n#3
hard#r#7	legendary#a#1	perfect#a#1	proficiency#n#2	scoop#n#3	time#n#2	graduate#v#1
hardworking#a#1	life_sentence#n#1	perfect#a#3	progress#n#3	significant#a#1	time#n#3	sportsman#n#1
head#n#26	long_time#n#1	perfect#v#1	progress#v#1	skill#n#2	time#v#5	
headline#v#1	long#a#1	perfection#n#1	project#n#2	skilled#a#1	top#v#3	
headliner#n#1	long#a#5	perfection#n#3	promise#n#1	smart#n#1	top#v#6	

Appendix III: Stopword List

\d+	instead/RB	'd/VBD	beside/RB	moreover/RB	're/VBP	'll/MD	're/VBP
's/VBZ	it/PRP	though/IN	can/MD	my/PRP$	about/IN	keep/VB	thorough/JJ
's/VB	itself/PRP	thus/RB	not/RB	near/IN	n't/RB	latterly/RB	through/IN
s/VBZ	later/RB	toward/IN	come/VB	neither/DT	and/CC	's/POS	upon/IN
s/VB	lest/IN	under/IN	could/MD	nine/CD	are/VBP	meanwhile/RB	uses/NNS
allows/VBZ	likely/JJ	unto/NN	do/VBP	noone/JJ	n't/RB	most/RBS	'd/MD
amongst/IN	ltd/NN	vs/CC	done/RB	off/IN	becomes/VBZ	myself/PRP	what/WP
anyways/NNS	me/PRP	we/PRP	edu/VB	old/JJ	behind/IN	never/RB	's/VBZ
aside/RB	next/IN	've/VBP	elsewhere/NN	only/RB	besides/IN	no/DT	've/VBP
because/IN	none/NN	were/VBD	etc/NN	out/RP	both/DT	nor/CC	be/VB
below/IN	nothing/NN	n't/RB	example/NN	rd/NN	c/NN	now/RB	becoming/VBG
between/IN	of/IN	whence/VB	for/IN	said/VBD	's/POS	often/RB	being/VBG
by/IN	okay/JJ	whereas/IN	from/IN	says/VBZ	cant/JJ	on/IN	n't/RB
ca/MD	others/NNS	's/VBZ	had/VBD	seem/VB	comes/VBZ	onto/IN	'd/MD
n't/RB	ourselves/PRP	why/WRB	has/VBZ	some/DT	contain/VBP	ought/MD	is/VBZ
com/NN	own/VBP	within/IN	n't/RB	sometime/RB	n't/RB	que/FW	is/VB
corresponding/VBG	placed/VBN	would/MD	he/PRP	sub/NN	does/VBZ	re/FW	's/POS
definitely/RB	rather/RB	n't/RB	's/VBZ	take/NN	eg/FW	regards/VBZ	s/POS
different/JJ	regarding/VBG	you/PRP	here/RB	than/IN	five/CD	same/JJ	became/VBD
do/VBP	seeing/VBG	'll/MD	hereupon/JJ	that/DT	had/VBD	second/NN	been/VBN
n't/RB	seen/VBN	able/JJ	howbeit/FW	's/VBZ	n't/RB	seven/CD	seems/VBZ
et/CC	she/PRP	across/IN	i/FW	them/PRP	have/VB	should/MD	was/VBD
four/CD	so/RB	against/IN	've/VBP	there/EX	here/RB	n't/RB	're/VBP
gets/VBZ	something/NN	almost/RB	in/IN	's/POS	's/POS	sometimes/RB	'm/VBP
goes/VBZ	still/RB	although/IN	into/IN	theres/NNS	hers/NNS	such/JJ	's/VBZ
has/VBZ	t/NN	an/DT	it/PRP	they/PRP	his/PRP$	taken/VBN	
herein/RB	's/VBZ	anyhow/NN	'd/MD	'll/MD	however/RB	thats/NNS	
how/WRB	th/DT	anywhere/RB	just/RB	this/DT	ie/FW	themselves/PRP	
i/FW	that/IN	are/VBP	latter/JJ	three/CD	inasmuch/FW	thereafter/RB	
'm/VBP	there/RB	ask/VB	little/JJ	to/TO	inward/JJ	thereupon/VBP	
immediate/JJ	therein/RB	beforehand/RB	mainly/RB	towards/IN	it/PRP	they/PRP	

Appendix IV: Final MSSR Variables

acme#n#1	up-and-coming#a#1	college#n#1	crown#v#2	improvement#n#1	prize#n#1	valet#n#1
advancement#n#3	great#a#1	Harvard#n#1	crucial#a#1	improvement#n#2	problem#n#1	victoriously#r#1
arduous#a#1	great#a#2	become#v#1	culminate#v#1	increase#n#2	problem#n#2	vitally#r#1
bang-up#a#1	vaccine#n#1	lawyer#n#1	defeated#a#2	indefatigably#r#1	progress#n#3	well#n#1
breakthrough#n#2	improve#v#1	training#n#1	degree#n#1	know#v#1	progress#v#1	well#r#11
brilliant#a#1	improvement#n#1	train#v#1	devote#v#3	leading#a#1	promotion#n#1	winner#n#1
competent#a#1	improvement#n#2	long#a#1	discovery#n#3	life_sentence#n#1	qualify#v#1	work#n#1
diligently#r#1	degree#n#1	intense#a#1	effective#a#1	long#a#1	race#n#2	year#n#1
excellent#a#1	publish#v#3	ace#n#3	effective#a#2	long#a#9	resolve#v#6	sportsman#n#1
forwarding#n#2	mistake#n#1	achieve#v#1	elegance#n#2	long#r#1	respect#v#1	
gallant#a#3	graduate_school#n#1	adequate_to#a#1	employment#n#2	lose#v#2	result#v#1	
hard#a#7	promote#v#2	ahead#r#5	endowment#n#1	lose#v#3	right#r#2	
hard#r#9	failure#n#1	ambition#n#1	exceed#v#2	master#v#4	rival#n#1	
intelligent#a#1	failure#n#2	ambitious#a#1	excellently#r#1	methodically#r#1	significant#a#1	
outplay#v#1	graduate#a#1	antidote#n#1	experience#n#1	mighty#r#1	skilled#a#1	
perfectionist#n#1	do#v#5	assurance#n#1	fabled#a#1	most#a#2	smoothly#r#2	
perfectly#r#2	flawlessly#r#1	best#a#1	faultlessly#r#1	new#a#1	solution#n#2	
performer#n#1	right#a#2	best#r#1	finely#r#3	new#a#11	solve#v#1	
praise#v#1	smooth#a#1	brainy#a#1	first#a#1	outstanding#a#1	star#n#4	
precisely#r#1	world#n#1	bring_around#v#2	flawless#a#1	perfect#a#3	star#n#6	
progress#n#2	boxer#n#2	cancer#n#1	fly#v#13	perfect#v#1	star#v#1	
proud#a#1	scientist#n#1	capable#a#1	graciously#r#1	perfection#n#1	strive#v#2	
remedy#n#2	nation#n#1	challenge#v#4	graduate#a#1	perplex#v#1	study#v#6	
science#n#1	famous#a#1	choice#a#1	graduate#v#1	potent#a#3	success#n#1	
success#n#3	great#a#1	clear#v#9	great#n#1	potential#a#1	success#n#2	
work#v#3	accomplished#a#1	contribution#n#5	hanker#v#1	potion#n#1	talented#a#1	
perfectly#r#1	formula#n#2	convinced#a#1	hardworking#a#1	precisely#r#2	tense#a#1	
confident#a#2	discover#v#1	correctly#r#1	head#n#26	prepare#v#5	time#n#1	
ability#n#2	graduate#n#1	create#v#5	highly#r#3	pride#n#1	top#v#3	
careful#a#1	degree#n#1	creative#a#1	important#a#5	pride#n#2	trophy#n#2	

Appendix V: Part 2 Anagrams

balm	save	nest	cork
bate	vail	stew	idea
cape	lamp	flow	oils
care	last	grin	south
deaf	snap	sign	spot
dale	aunt	sink	beak
dare	warp	loop	
dawn	swap	thing	
hare	blot	cafe	
lake	cone	drier	
wake	keen	relay	
leap	felt	rival	
late	hose	tacit	
peat	lure	acts	
east	note	baker	

VITA

Marc Halusic received his B.A. in Psychology from the UC Berkeley, where he was introduced to Piagetian theory by Joe Campos and Peter Gillette, and was introduced to motivation research by Ari Malka. After college, he worked as a paid research assistant UC Berkeley's Institute of Human Development, where he worked assessing the mathematical abilities of children involved in a study of the efficacy of a new form of math intervention designed for low SES preschoolers. He then worked on a Masters degree in Educational Psychology at the University of Iowa under the instruction of Johnmarshall Reeve, learning about the connection between Self Determination Theory and educational practice. Finally, he earned a second Masters degree (this time in Social and Personality Psychology) and Ph.D. at the University of Missouri-Columbia while working with his advisor, Kennon M. Sheldon, and other professors in the Department of Psychological Sciences, notably Laura King and Laura Scherer. In that time, he conducted research on a variety of subjects, including Self Determination Theory and self complexity, the role of fear in producing an intuitive cognitive style, and Motive Disposition Theory.